Studies in Complex Financial Instruments and their Valuation

Niklas Ekvall

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framlägges till offentlig granskning
måndagen den 15 februari, 1993
kl. 15.15 i sal Ruben
Handelshögskolan, Saltmätargatan 13–17

STOCKHOLM 1993

Studies in Complex Financial Instruments and their Valuation

EFI/THE ECONOMIC RESEARCH INSTITUTE

Stockholm School of Economics

Address: Sveavägen 65, Box 6501, S-113 83 Stockholm. Tel. +46 8-736 90 00

Basic Orientation

The Economic Research Institute, at the Stockholm School of Economics (EFI) is devoted to the scientific study of problems in management science and economics. It provides research facilities for scholars belonging to these disciplines, training advanced students in scientific research. The studies to be carried out by the Institute are chosen on the basis of their expected scientific value and their relevance to the research programs outlined by the different sections of the Institute.

Research Divisions:

A	Management and Organization Theory
B	Accounting and Managerial Finance
C	Managerial Economics
CFR	Center for Risk Research
CHE	Center for Health Economics
D	Marketing, Distribution and Industry Dynamics
ES	Economic Statistics
FI	Finance
F	Public Administration
IEG	International Economics and Geography
I	Information Management
P	Economic Psychology
S	Economics

Research Programs:

Man and Organization
The Foundation for Distribution Research
Policy Sciences

Additional information about research in progress and published reports is described in our project catalogue. The catalogue can be ordered directly from The Economic Research Institute, Box 6501, S-113 83 Stockholm, Sweden.

Studies in Complex Financial Instruments and their Valuation

Niklas Ekvall

STOCKHOLM SCHOOL OF ECONOMICS
THE ECONOMIC RESEARCH INSTITUTE

A Dissertation for the
Doctor's Degree in Philosophy
Stockholm School of Economics 1993

© EFI and the author
ISBN 91-7258-358-4

Keywords:
Corporate finance
Financial innovation
Complex securities
Asset valuation
Contingent claims analysis
Numerical methods

Distributed by:
EFI, Stockholm School of Economics,
Box 6501, S-113 83 Stockholm, Sweden.
Graphic Systems AB, Stockholm 1993

Preface

After two years of intensive work it is time to bring this project to a halt, and present what I have accomplished. This presentation is in the form of this Ph.D. dissertation. To take all credit for the dissertation myself would be unfair. Some of the credit must, of course, go to friends and colleagues, without whom this project had not been possible.

First and most, I want to thank two persons. One of them is, of course, my wife Eva. Despite many lonely evenings and weekends, she has encouraged me at every stage of the writing of the dissertation. The other person is my advisor Peter Jennergren (Stockholm School of Economics). He took me on as doctoral student, and has guided me through the writing process. Furthermore, in spite of a heavy workload, he has always managed to find the time to read my drafts and to discuss my research problems.

In addition to Peter Jennergren, Bertil Näslund and Bo Sellstedt have been members of my Ph.D. committee at the Stockholm School of Economics. I am indebted to both of them for improving the quality of my work by providing valuable comments and suggestions.

The Section for Managerial Economics at the Stockholm School of Economics has provided a stimulating environment for research. I thank Tomas Sörensson for good friendship and support. I am also very grateful to Harriet Lundh. She has helped me with many practical details, and always found solutions to practical problems.

Paul Kleindorfer and Stavros Zenios made it possible for me to visit the Wharton School of the University of Pennsylvania during 6 months in 1992. Stavros Zenios also provided me with excellent computer facilities at the HERMES Laboratory for Financial Modeling and Simulation during my stay at Wharton.

Some of the calculations in the dissertation are carried out on a supercomputer of the type Connection Machine Model CM-2000 (CM-2000). I am grateful to the PDC Center for Parallel Computers at the Royal Institute of Technology for allowing me to use their supercomputer. Fredrik Hedman and especially Erik Wallin instructed me in programming and using the CM-2000.

The following individuals have contributed helpful comments and suggestions: Olle Axling (Linköping Institute of Technology), Tommy Elving (Linköping Institute of Technology), Peter Högfeldt (Stockholm School of Economics), Jesper Oppelstrup (Royal Institute of Technology), and Staffan Viotti (Stockholm School of Economics).

Marc Platten (in particular), Lucy Loercer and Jennifer Forsberg of the Professional Communication Skills at the Stockholm School of Economics have helped me with my English. I also want to thank Anders Utling for helping me with the graphical work.

I am indebted to Rune Castenäs, the director of the Economic Research Institute at the Stockholm School of Economics, for solving many different kinds of practical problems. Finally, financial support from Bankforskningsinstitutet is gratefully acknowledged.

Stockholm in December, 1992

Niklas Ekvall

Contents

Chapter 1

An overview and summary of the dissertation

This dissertation consists of four different papers, placed in separate chapters. More precisely, the dissertation has the following structure:

Chapter 2: Paper A: Financial Innovation and Instruments

Chapter 3: Paper B: How "Errors" in Boundary Conditions Affect Solutions when the Implicit Finite Difference Method is Used

Chapter 4: Paper C: Two Finite Difference Schemes for Evaluation of Contingent Claims with Three Underlying State Variables

Chapter 5: Paper D: A Lattice Approach for Pricing of Multivariate Contingent Claims

Papers A to D can be read independently of each other. To be more concrete, each of the papers stands by itself in the sense that each of them can be read and understood even if nothing else in the dissertation is read. Nevertheless, papers A to D are related. In short, the relations between the papers can be described as follows:

Papers B to D have a common theme, which is valuation of complex financial instruments with the help of Contingent Claims Analysis (CCA)[1] and numerical methods. Paper A can be viewed as an independent and non-traditional prelude to papers B to D.

Before proceeding with this overview/summary, the following practical matters regarding the layout of the dissertation will be mentioned:

- One appendix is attached to paper A, and three appendices to paper B. The appendix/appendices belonging to a specific paper is/are placed last in the chapter that contains that particular paper. The appendix attached to paper A is called appendix A, and the three appendices belonging to paper B are called appendix BI, BII, and BIII respectively.
- Even though papers A to D are independent and separate entities, a single bibliography, common to all papers, is given last in the dissertation. The reason for having a single bibliography is that all papers have many sources in common.
- In the bibliography, all sources are numbered according to the alphabetical order of the names of the sources' authors. Whenever a source is referred to in the text, the number of the source in the bibliography is given.
- As previously mentioned, papers B to D have a common theme. Furthermore, papers B to D can be read completely independently of each other, since each of them contain and explain all theory needed to understand them. A consequence of these facts is that some repetitions can be observed between papers B to D.

In the remainder of this chapter, short summaries of the four papers which constitute the dissertation are given.

Paper A: Financial Innovation and Instruments

The pace of financial innovation has been very fast during the last two decades. In particular, there has been an explosion in the number of financial instruments that are traded

[1]In short, CCA is a very flexible technique for determining the value of an asset whose payoffs depend upon the evolution of one or more underlying state variables.

on the financial markets. There are, furthermore, many reasons to believe that there will be a rapid rate of financial innovation also in the future. Paper A provides a discussion of the process of financial innovation. A lengthy appendix (appendix A) is attached to paper A. In appendix A, more than 100 more or less complex financial instruments are described briefly.

Paper A is an independent and non-traditional prelude to papers B to D. It is independent in the sense that it can (hopefully) be interesting to read independently of the other papers in the dissertation. It is non-traditional since it does actually not provide a basic theoretical synopsis intending to simplify the reading of the rest of the dissertation.

Rather, the purpose of paper A is to arouse an interest in, and to motivate, papers B to D (which are of a rather technical nature) also for individuals other than financial economists. This is done by briefly describing a large number of often fairly complex financial instruments (appendix A), and by placing CCA as well as the valuation methods discussed in papers B to D, and also the financial instruments in appendix A, in a broader context.

That broader context is provided by discussing CCA's role in the process of financial innovation. The process of financial innovation is, however, very complex. A discussion that only focuses on CCA's role in the process of financial innovation would therefore be very incomplete. For this reason, paper A also addresses a number of other topics. These topics are: three frameworks for financial innovation suggested in the literature, driving forces behind financial innovation, the social value of financial innovation, and the pace of financial innovation in the future.

The main conclusion in paper A is that the importance of CCA, in the context of financial innovation, is twofold. Firstly, CCA provides a "production mechanism" that can be used by financial intermediaries to construct new financial instruments. Secondly, at the same time, the importance of CCA successively increases as the number of complex financial instruments that can be found on the financial markets increases.

To be more precise, CCA and dynamic portfolio-replication provide financial intermediaries with production technologies that allow them to manufacture new financial instruments. When the "successes" of the new instruments have attracted sufficiently large volume, they are likely to migrate from intermediaries to markets. An increasing number of disparate traded securities supplies financial intermediaries with more flexible and wide-ranging production technologies. That is, when new securities become traded on the financial markets, financial intermediaries can use them in their CCA/dynamic portfolio-replicating production process. This enables financial intermediaries to synthesize cash flow patterns which could previously not be synthesized. In turn, this makes it possible for financial intermediaries to further customize new financial instruments to meet the needs of their customers. This process continues onward, moving toward the theoretically limiting case of dynamically complete markets.

The dynamic development described above, however, also makes CCA increasingly more important independently of its role in the process of financial innovation. This is due to the fact that along the path toward the limiting case of dynamically complete markets, an increasing number of complex financial instruments are traded on the financial markets, and many complex financial instruments have attributes that necessitate the use of CCA to value them in a theoretically correct fashion.

One problem that often arises when CCA is used is that it is not possible to find a closed-form solution for the value. Numerical methods must therefore often be relied on. Furthermore, in many cases and especially in cases where there is more than one underlying state variable, numerical methods become computationally laborious. Due to this last fact, it is unfortunate that many complex financial instruments require CCA with several underlying state variables, and associated numerical methods, for accurate valuation.

Naturally, being able to value the increasing number of traded complex securities has an academic interest. Such valuation methods are, however, certainly not only of interest to academics. For the agents on the financial markets, it is of crucial importance to be able to assess the value of complex financial instruments. This is equally true for financial services firms that construct and promote the products, borrowers that sell the products for the financing of their activities, and investors that buy the products.

Paper A and its attached appendix should make it clear that research on and development of efficient numerical methods that can be used in the CCA context is an important research area. Furthermore, this research area is likely to become increasingly more important in the future.

Paper B: How "Errors" in Boundary Conditions Affect Solutions when the Implicit Finite Difference Method is Used

Many financial instruments, as well as other assets, have features that make CCA superior to other valuation methods. When CCA is used, the value of the financial instrument is often obtained through the solution of a partial differential equation. To solve a partial differential equation uniquely, side conditions have to be imposed on the equation. The partial differential equation looks almost the same for all financial instruments. The features that distinguish one financial instrument from another are incorporated into the side conditions.

It is impossible to find a closed-form solution to the partial differential equation subject to the relevant side conditions, for most financial instruments. For the instruments that lack closed-form solutions, numerical methods have to be used when establishing their values.

The implicit finite difference method is one of the most commonly used among the numerical methods suggested in the CCA context with only one underlying state variable. The implicit finite difference method is flexible, and can be used for pricing financial instruments with very different characteristics.

The side conditions consist of initial condition and boundary conditions. The initial condition is often fairly easy to model. Modelling the boundary conditions is, however, often troublesome. "Errors"[2] in the boundary conditions are thus not unusual.

[2] There is usually not a single true way to model the boundary conditions for a given financial instrument. By using the term "error", both pure errors and deviations between more or less reasonable modelling choices are referred to.

The research task of paper B is to investigate how errors in the boundary conditions affect the solution when the implicit finite difference method is used. To the knowledge of the author, there has not been any work regarding this research task. Furthermore, the research task has an unclear status in the academic literature.

Some researchers seem to have the apprehension that modelling errors in the boundary conditions of the magnitude likely to occur do not have any noticeable effects on the results. This apprehension manifests itself by the fact that the boundary conditions are not specified or incompletely specified in many of the articles in which the implicit finite difference method is used.

On the other hand, some researchers seem to have the feeling that different ways of modelling the boundary conditions have major effects on the results when the implicit finite difference method is used. This feeling manifests itself through the arguments for methods where the modelling of boundary conditions can be avoided, and by the arguments against the implicit finite difference method.

How errors in the boundary conditions affect the solutions is, however, not only of academic interest. It also has clear practical implications. If valuation by means of CCA and the implicit finite difference method is to be used for commercial purposes, it is important to know how robust the method is with respect to "errors" in the boundary conditions.

The results from the investigation in paper B can be summarized as follows:

The boundary conditions consist of a lower boundary condition and an upper boundary condition in the case of a CCA problem with one underlying state variable. The reduction of the effect from errors in the lower boundary condition increases very rapidly when moving away from the lower boundary in all evaluations performed. In fact, the reduction of the effect from errors in the lower boundary condition becomes amazingly large only a small distance away from the lower boundary. Almost any value can thus be placed at the lower boundary without affecting the solution.

The reduction of the effect from errors in the upper boundary condition is much slower than the reduction of the effect from errors in the lower boundary condition. The reduction of the effect from errors in the upper boundary condition is, however, very large for reasonable values of the current value of the underlying state variable.

To summarize, the results from the investigation in paper B imply that errors in the boundary conditions of the magnitude likely to occur in any real application will not lead to solution errors of any practical significance.

From the above facts, it can be concluded that the argument against the implicit finite difference method based on the method requiring the specification of boundary conditions is weak. Furthermore, the results from the study in paper B may justify the lack of discussion of how the boundary conditions are modelled in many articles in which the implicit finite difference method is used. The exact modelling is simply not interesting, since differences in the modelling choice very rarely give any differences in the result.

The above results also imply that the implicit finite difference method seems to be robust enough with respect to errors in the boundary conditions for commercial purposes.

Paper C: Two Finite Difference Schemes for Evaluation of Contingent Claims with Three Underlying State Variables

There are many examples of valuations of financial instruments with the help of CCA and numerical methods in the literature. These examples are in an overwhelming majority valuations of financial instruments with only one underlying state variable. As should be clear from the summary of paper A, there has been an explosion of new exotic financial instruments during the last two decades. Many of the new instruments have features which make the assumption that the value of the instrument depends only on one underlying state variable an oversimplification. Indeed, to price accurately these exotic instruments it is often necessary to assume that the value of the financial instrument depends on two or even more underlying state variables.

In the summary of paper B, it was mentioned that the value of a financial instrument is often obtained through the solution of a partial differential equation when CCA is used. The complexity of the partial differential equation increases with every additional state variable. Consequently, as the number of state variables is increased, the likelihood of finding a closed-form solution to the partial differential equation subject to the relevant side conditions is decreased. Thus, the probability that numerical methods will have to be used for the valuation of financial instruments also increases as the number of state variables increases. However, as the number of underlying state variables increases so does the degree of complexity and the need for computing power in numerical work. Moreover, the degree of complexity, and computing power needed, increases substantially for every additional state variable.

For the case of financial instruments with two underlying state variables, the literature contains some examples of valuation by numerical methods. These examples are, however, considerably fewer than the examples of valuation of financial instruments with only one underlying state variable. For the case of financial instruments with three underlying state variables, there is, as far as the author has been able to ascertain, only one published example of a valuation by numerical methods. This example is a valuation by means of a lattice (or tree) approach.

Finite difference schemes have proved to be very flexible numerical methods for the pricing of financial instruments with one and two underlying state variables. The schemes are flexible both with regard to the features (puttable, callable, extendable etc.) of the instruments as well as the types of stochastic processes followed by the underlying state variables. The flexibility of the finite difference schemes and the steady stream of new complex financial instruments imply that finite difference schemes for the valuation of financial instruments with three underlying state variables can supposedly be very useful. In paper C, two such schemes are developed and tested.

Leaving the details to paper C, the first of the two schemes is a pure explicit finite difference scheme. This scheme is among the simplest conceivable of the finite difference ones that can be used for the valuation of financial instruments with three underlying state variables. The second scheme is called the generalized ADI-method, which is more advanced than the explicit finite difference scheme.

Before summarizing the results from the investigation in paper C, it should be mentioned that valuations of financial instruments with three underlying state variables by means of finite difference schemes require considerable computing power. In fact, the valuations require so much computing power that they can hardly be executed on an ordinary PC. For this reason, all "heavy" computations in paper C have been performed on a supercomputer of type Connection Machine Model CM-2000.

Because of the use of the supercomputer, it could be argued that the practical relevance of the research performed in paper C is low, since it is commonly believed that there are very few advanced architecture computers in the financial community. This is, however, incorrect. Several financial companies are currently using advanced architecture computing, and their number is likely to increase fast in the future.

The results from the investigation in paper C can be summarized as follows:

Valuation of financial instruments with three underlying state variables by means of the finite difference schemes is a story of mixed success. The story is, however, mainly a success story.

Let us first consider the negative aspects. Both the explicit finite difference scheme and the generalized ADI-method require a transformation of the partial differential equation in order to become usable. The transformation required is, however, a simple logarithmic one.

The positive aspects then remain. After the transformation, both the explicit finite difference scheme and the generalized ADI-method proved accurate in every valuation performed.

A comparison between the explicit finite difference scheme and the generalized ADI-method indicates that the generalized ADI-method has much better stability properties than the explicit finite difference scheme. This means that the explicit finite difference scheme requires many more time steps to achieve stability than the generalized ADI-method. This drawback for the explicit finite difference scheme is partly eliminated since it iterates faster through time than the generalized ADI-method. The explicit finite difference scheme is, however, far from fast enough to make up for the extra time steps required (due to its bad stability properties) when using the Connection Machine Model CM-2000.

The superior stability properties of the generalized ADI-method clearly speak for using that scheme. The explicit finite difference scheme is, however, easier to implement. Both schemes have thus advantages and drawbacks.

Numerical evaluation of financial instruments with several underlying state variables by means of finite difference methods on a massively parallel computer, like the CM-2000 machine, hence seems to be a usable approach. Moreover, that approach should be usable for academics as well as for practitioners.

Paper D: A Lattice Approach for Pricing of Multivariate Contingent Claims

The basic research task in paper D is essentially the same as that in paper C. The research task in paper D is thus to develop and test a numerical method which can be used when pricing financial instruments with several underlying state variables. The numerical method developed in paper D is a lattice (or tree) method.

As previously mentioned, one common approach for pricing financial instruments with the help of CCA is to derive a partial differential equation which the price of the contingent claim must satisfy. The price of the financial instrument is then given by the solution to that partial differential equation. If no analytical solution can be found, as is often the case, the partial differential equation can be solved numerically with finite difference methods. It is this approach that is used in paper C.

Another approach for solving numerically the price of a financial instrument in the context of CCA is to approximate the continuous processes of the underlying state variables with discrete versions. It is this last approach that is used in paper D, i.e., when the price of the financial instrument is calculated with the help of a lattice method.

Lattice methods are among the most popular of the numerical methods which have been suggested in the context of CCA. Reasons for this popularity are that most lattice approaches are intuitively simple, flexible and can easily handle early exercise conditions.

The method developed by *Cox, Ross and Rubinstein* (CRR) in 1979 (see [43]) is the most well-known, and presumably also the most commonly used, of the lattice methods. The CRR method has later been extended and modified in several articles. The CRR method and most of its extensions and modifications are, however, only applicable to pricing problems with only one underlying state variable.

In the search for a lattice method that can easily be extended to handle more than one underlying state variable, a method developed by *Boyle, Evnine and Gibbs* (henceforth called the BEG approach) (see [16]) is among the most promising. The BEG approach is a multivariate extension of the CRR method. BEG approximate the system of underlying processes (all assumed to follow geometric Brownian motions) with a multivariate binomial tree. BEG first set the jump amplitudes and then determine the jump probabilities in a fashion that assures convergence of the discrete approximation to the system of continuous processes, as the step size in the time dimension approaches zero.

The BEG approach has, however, some problems. It has a rather slow convergence, and hence is not efficient. Furthermore, the BEG approach is neither general with respect to the processes of the underlying state variables, nor with respect to the number of underlying state variables. The last problem follows from the fact that the BEG approach does not guarantee that all jump probabilities will be non-negative when the number of underlying state variables is larger than two. Moreover, the problem with possible negative jump probabilities increases when the number of underlying state variables increases.

The lattice approach developed in paper D (henceforth called the NEK approach) can be viewed as an improvement of the BEG approach. In short, the NEK approach can be described as follows:

The NEK approach starts with transformations, in two steps, of the processes of the underlying state variables (if needed). The first step transforms the processes of the underlying state variables into processes with constant instantaneous drifts, standard deviations and correlation coefficients. The second step transforms the processes from the first step into a system of uncorrelated processes. The system of uncorrelated processes is then approximated with a multivariate binomial tree, as in the BEG approach. However, unlike the BEG approach, all jump probabilities are specified first (all probabilities equal), and then the jump sizes are determined in a fashion that assures convergence between the discrete approximation and the system of uncorrelated processes as the step size in the time dimension approaches zero.

There are several advantages in setting all jump probabilities equal[3], as compared to first setting the jump sizes and then deriving the jump probabilities. For example, a more efficient method (faster convergence) and easier implementation is achieved. Furthermore, the problem with possible negative jump probabilities is avoided.

Other advantages of the NEK approach are:

- It can easily be extended to cases with an arbitrary number of underlying state variables (the computational burden will, of course, be large for cases with many state variables).
- It can handle a certain variety of stochastic processes (through the transformation procedure).
- It is simple and easy to understand.
- It gives *extremely* simple formulas for the jump sizes and the jump probabilities. This makes the NEK approach easy to implement.
- It makes the number of branches emanating from each node in the tree increase fairly slowly as the number of state variables increases.

A large number of valuations of options with three and four underlying state variables have been performed, both with the NEK approach and with the BEG approach. These valuations show that the NEK approach has a very fast convergence. The computed values are accurate enough for most needs with as few as 10 time steps, and the computed values are astonishingly accurate with 20 time steps.

If the accuracy of the NEK approach is compared to that of the BEG approach, it can be concluded that the NEK approach is superior to the BEG approach, even if the possibility of negative jump probabilities in the BEG approach is disregarded. This is

[3]The strategy to set all jump probabilities equal cannot be used in the BEG approach. It is the transformations which make it possible to use that strategy in the NEK approach.

due to the superior rate of convergence of the NEK approach. A fast rate of convergence is especially important when valuing financial instruments with several underlying state variables, since the computational burden increases very fast as more time steps are used.

In paper D, valuations of American options with the help of the NEK approach are also performed. The NEK approach can thus easily handle financial instruments with early exercise conditions.

In papers C and D, numerical methods which can be used to price financial instruments with several underlying state variables are thus developed and tested. The methods in paper C are finite difference schemes, and the method in paper D is a lattice method (the NEK approach). This chapter is concluded with a brief comparison between the finite difference schemes in paper C and the NEK approach.

Compared to the finite difference schemes in paper C, the NEK approach has several advantages. Among those, the following can be mentioned:

- The accuracy of the NEK approach is better.
- Finite difference schemes require more computing power.
- The NEK approach is easier to implement.
- The NEK approach is less complex, and is easier to understand.
- The NEK approach is more elegant.

From all of the advantages of the NEK approach listed above, it is easy to get the impression that the NEK approach always is superior to the finite difference schemes in paper C. This is, however, not the case. Finite difference schemes are more flexible than the NEK approach. This is true both with regard to applications (e.g., valuation of linked projects or valuation of financial instruments embedded in a capital structure with several other instruments) as well as the processes of the underlying state variables. There are hence cases which can be handled with finite difference methods, but not with the NEK approach. In cases where it is possible to do so, it is, however, preferable to use the NEK approach rather than any of the finite difference schemes in paper C.

Chapter 2

Paper A: Financial Innovation and Instruments

2.1 Introduction

During the last two decades financial innovation has been fast-paced. As a consequence of this, there has also been an explosion in the number of financial instruments traded on the financial markets. This paper provides a discussion of the process of financial innovation. In appendix A, a selection of over 100 more or less complex financial instruments is briefly described.

This paper is one of four in a Ph.D. dissertation. More precisely, the dissertation consists of this and three papers of a more technical nature. The common theme of the technical papers is valuation of complex financial instruments with the help of Contingent Claims Analysis (CCA) and numerical methods (a short description of CCA is given later in this introduction). It is in the three technical papers that the "proper" research in the dissertation is performed.

The purpose of this paper is to function as an independent prelude to the other three papers in the dissertation[1]. The paper is not, however, a traditional introduction that

[1]This paper is a prelude, but also independent in that it can hopefully be interesting to read independently of the other papers.

provides a basic theoretical synopsis intending to simplify the reading of the rest of the dissertation[2].

Instead, the aim is to arouse an interest in, and to motivate, the subsequent three papers of a more technical nature also for individuals other than financial economists. This is done by briefly describing more than 100 often fairly exotic financial instruments (appendix A), and by putting CCA, the three technical papers, and also the financial instruments in appendix A, into a broader context.

The broader context is provided by discussing the process of financial innovation and the role of CCA in this process. The connection, even if indirect, between this paper and the other papers in the dissertation is the interaction between CCA and the explosion of new financial instruments. More precisely, CCA provides a "production mechanism" that can be used by financial intermediaries to construct new financial instruments, while at the same time the importance of CCA successively increases as the number of complex financial instruments (which, of course, must be valued) that can be found on the financial markets increases. The use of CCA, both to construct new financial instruments and to value existing ones, often requires numerical methods. This fact implies that development of numerical methods that can be used in the CCA context is an important research topic. Moreover, the importance of this research topic will probably increase in the future.

As previously mentioned, the aim of this paper is to arouse an interest in, as well as to motivate, the papers of a more technical nature in the dissertation. This is partly done by discussing the role of CCA in the process of financial innovation. The process of financial innovation is, however, very complex . A discussion concerning the process of financial innovation that only describes the role of CCA would therefore be very incomplete. For this reason, the process of financial innovation is also approached from other angles in the paper. In addition to the role of CCA in the process of financial innovation, a number of other subjects is consequently discussed in section 2.2. These are: three frameworks for financial innovation suggested in the financial literature, driving forces behind financial

[2]A traditional introduction is not necessary. This follows from the fact that each of the technical papers contains and explains all the theory that is needed to understand them.

innovation, the social value of financial innovation, and the pace of financial innovation in the future. Some very brief remarks on the comparison of financial innovation to innovation of physical products/processes and characteristics of successful financial innovations are also presented in section 2.2.

To summarize, the fast stream of financial innovation during the last 20 years has, at times, given rise to a heated debate in politics, the mass media and among academics. It also appears that many people have an opinion concerning the social value/evil of the process of financial innovation. Since CCA is a part of the process, the hope is that this paper will motivate and arouse interest in the three technical papers in the dissertation also for those who are not interested in financial economics, CCA and numerical technicalities.

A lengthy appendix (appendix A) is attached to this paper. The appendix contains brief descriptions of more than 100 financial instruments. Many of these instruments are fairly complex. Appendix A provides a unified and concise description of more financial instruments than can be found in any other single source (as far as the author has been able to ascertain). Furthermore, appendix A is written so that it can be read without reading the rest of the paper. Although all financial instruments available on today's financial markets are not included in appendix A, it can, perhaps, nevertheless be used as a reference list for financial instruments.

Many new financial instruments defy classification. The instruments in appendix A have in any case been divided into the groups: basic derivative products and some of their extensions[3] (subsection A.1), floating-rate securities (sub-subsection A.2.1), zero coupon bonds and several other bond variants (sub-subsection A.2.2), asset-backed debt (sub-subsection A.2.3), commodity-linked, exchange rate-linked, and index-linked debt instruments (sub-subsection A.2.4), preferred stock innovations (sub-subsection A.3.1), convertible debt innovations (sub-subsection A.3.2), and common equity innovations (subsection A.4).

One main reason for adding an appendix like appendix A is that the mere existence of all these financial instruments unquestionably motivates the three technical papers in the

[3]Henceforth, the terms financial instrument and financial product will be used interchangeably.

dissertation. The variety and often considerable complexity of the financial instruments on today's financial markets imply that valuation is a relevant and important research topic. Methods for pricing complex financial instruments are, of course, not only of interest for academics. They are of great interest for institutions involved in the financial markets as well. This is equally true for financial services firms that construct and promote the instruments, borrowers that sell the instruments for the financing of their activities and investors that buy the instruments.

CCA has proved to be a very useful and flexible framework for pricing many of the new complex financial instruments. As a matter of fact, many new financial instruments have attributes that make CCA superior to other currently known valuation methods. This is also true for most of the instruments (even if not all) described in appendix A.

CCA is a technique for determining the price of an asset whose payoffs depend upon the evolution of one or more underlying state variables. One problem that often arises when this framework is used is that it is not possible to find a closed-form solution for the price. Numerical methods must therefore often be relied on. Furthermore, in many cases and especially in cases where there is more than one underlying state variable (using more than one underlying state variable is a natural modelling choice for many of the instruments mentioned in appendix A) numerical methods become computationally laborious. From this it can be concluded that research concerning and development of efficient numerical methods that can be used in the CCA context is an important research area. As previously mentioned, the three technical papers in the dissertation belong to this research area.

At this point, an example will be used for clarification.

Consider the case of convertible bonds (henceforth, abbreviated convertible). A convertible is a well-known and fairly simple financial instrument (at least compared to many of the instruments described in appendix A). In short, a convertible is a bond with a right (no obligation) to convert into common stock of the issuing company according to prespecified terms. Upon conversion, the issuing company issues new shares, resulting in dilution of the equity. In addition to this, most convertible issues have put and/or call features. (The call feature provides the issuing firm with an option to buy back the convertible according to prespecified terms, and the put feature provides the convertible bond holders with an option to sell back the convertible to the issuing firm according to prespecified terms.)

Convertibles have characteristics that imply that CCA must be used to value them in a theoretically correct fashion. There are, however, several modelling choices within CCA when valuing convertibles.

The simplest approach is to assume a constant interest rate and view the convertible as a package of a bond and a European call option. Using this approach, an analytical formula can be used for the valuation.

Since almost all convertible issues can be prematurely converted, and also have call/put features, the very simple approach above is naturally a great simplification. To take possible premature conversion and call/put features into account, numerical methods have to be used.

There are, however, more modelling choices to be made. Consider the choice of underlying state variables. The simplest choice is, of course, to use only one state variable. With one state variable, the modeller can choose to use either the stock price or the firm value. Without going into any details, the use of the stock price is the simpler approach, whereas the use of the firm value is theoretically more correct (the stock price is used as underlying state variable in the simple approach where the convertible is viewed as a package of a bond and a European call option).

The value of a bond is mainly dependent on the term structure of interest rates[4]. Since a convertible is partly a bond, the value of a convertible is naturally also dependent on the term structure of interest rates.

It is, however, difficult to model the stochastic behaviour of the term structure of interest rates. The simplest alternative is to use a one-stochastic-variable approach[5]. If

---

[4]In short, the term structure of interest rates relates the interest rate on a default-free discount bond to its time to maturity (see [67] p.257).

[5]In a one-stochastic-variable approach, the stochastic variable is usually the instantaneous riskfree

a one-stochastic-variable term structure model is included when valuing convertibles, the result is a CCA model with two underlying state variables (i.e., the firm value/stock price and the stochastic variable in the term structure model) which must be evaluated with the help of numerical methods.

To attain a more accurate model of the term structure of interest rates, a number of authors have suggested two-stochastic-variable models (e.g., see [24] and [26]). The inclusion of a term structure model like this creates a CCA model with three underlying state variables when valuing convertible bonds. To be more precise, the result is a CCA model with the firm value/stock price and the two stochastic variables in the term structure model as underlying state variables, which must be evaluated with numerical methods[6].

An argument concerning modelling choices in the CCA framework like the one in the context of convertibles above can be applied to most of the financial instruments in appendix A.

In this paragraph, the organization of the remainder of the paper is described. The discussion of financial innovation in general is presented in section 2.2. This discussion is rather superficial. A thorough and complete discussion of the process of financial innovation lies outside the aims of the paper. The discussion of the process of financial innovation is partly achieved by addressing some selected aspects. Finally, some concluding remarks are given in section 2.3.

2.2 Financial innovation

The pace of financial innovation has been fast during the last twenty years. This is true for financial products (e.g., swaps, asset-backed bonds) as well as for financial processes (e.g., shelf registration, automated teller machines).

As already mentioned, the discussion of the process of financial innovation in this

interest rate (e.g., see [38], [41] and [135]).

[6]It should also be noted that the modelling choices discussed above are only a selection of all possible choices when valuing convertibles with the help of CCA.

section is partly achieved by addressing some selected aspects. These aspects are: "The role of CCA in the process of financial innovation" (subsection 2.2.1), "Stimuli for financial innovation" (subsection 2.2.2), "Are financial innovations beneficial for society?" (subsection 2.2.3), and "The pace of financial innovation, a brief look into the future" (subsection 2.2.4).

The aspects discussed here are, of course, only a few selected dimensions of the very complicated process of financial innovation. Given the aims of this paper, these aspects are also discussed in a fairly superficial manner. Instead, the aim is to put CCA and all the more or less complex financial instruments described in appendix A into a slightly broader context. However, before attention is turned to the above mentioned aspects, the process of financial innovation will be discussed (briefly) from a more general point of view.

Some (to varying degrees) general frameworks for financial innovation are suggested in the financial literature. Of these, the frameworks developed by *Kane*, *Silber*, and *Ross* respectively will be described shortly. Before this is done, however, the *functional perspective* suggested by *Merton* (see [102], [103], [104], [105] and [106]) should be mentioned.

Merton's framework is much more than a framework for financial innovation. It provides a view of the entire financial system, and for example, gives directive rules for efficient regulation of the financial system. Merton's perspective focuses on the core functions of the financial system, and financial innovation is viewed as a continuous process that successively, like a "spiral", improves the financial system's ability to fulfil its core functions. Merton's perspective has been an important source of inspiration when writing subsections 2.2.1 and 2.2.3. More will consequently be said about the functional perspective later in the paper.

A very interesting model is the *regulatory dialectic* framework, developed by *Kane* (see [78], [79], [80] and [81]). The regulatory dialectic framework is a dynamic model, in which financial innovation is an outcome of a continuing struggle between opposing economic and political forces. More precisely, Kane's model treats political processes of regulations and economic processes of regulatee avoidance (possibly by financial innovation) as opposing forces that continually adapt to each other. This continuous interactive adaptation evolves as a series of responses with regulators and regulatees trying to maximize their objectives, conditional upon how they believe their counterparty will react.

Kane's framework can be interpreted as an eternal process in 3 stages. The stages are regulation, avoidance and re-regulation, where the second or third stage can be interpreted as the first stage in a new sequence.

In Kane's framework, features like technological change and change in the economic environment (i.e., some of the things that are called stimuli for financial innovation in subsection 2.2.2) are viewed as disturbing forces that can "kick off" a sequence of regulation, avoidance and re-regulation. This is due to the fact that change in the economic environment can force bureaucrats and politicians to introduce new regulation or to change existing regulation. Furthermore, change in the environment can force regulatee avoidance by increasing the opportunity cost burdens associated with pre-existing regulations or reducing the marginal costs of avoiding regulatory burdens (reduced marginal costs of avoiding regulatory burdens is a likely outcome of technological change). All of these facts can start a sequence of regulation, avoidance and re-regulation.

The main theme in Kane's framework is that most new financial products/processes are invented for the purpose of circumventing regulatory constraints. The framework suggested by *Silber* can be viewed as a generalization of Kane's theme. Thus, Silber's framework suggests that new financial products/processes are innovated to lessen the financial constraints (not only regulatory) imposed on firms (see [123]).

Silber uses a simple linear programming model to illustrate his framework. In this model, a firm maximizes its objective function subject to a set of constraints. The constraints in the set are both of internal (imposed by the firm itself) and external (e.g., government regulations) nature.

During normal action, the firm will sell securities or accept deposits and invest the proceeds in a way that maximizes its objective function, within the bounds of existing constraints. New financial products/processes are invented when exogenous changes in the firm's environment make it too costly for the firm to be restricted by the set of constraints with its current stock of financial tools.

Silber's constraint-induced innovation framework suggests that search and development costs of new financial products/processes are continuously balanced against the shadow prices of the set of constraints. Inertia and search/development costs imply that only an increase sustained over time in shadow prices will stimulate financial innovation.

In Silber's framework, the stimuli for financial innovation that will be discussed in subsection 2.2.2 can be viewed as external forces that affect a firm's set of constraints and/or increase the cost of adhering to existing constraints.

Finally, Silber claims that empirical results support his framework. That is, empirical results support the claim that rising costs of adhering to constraints stimulate financial innovation (see [8]).

Let us now briefly consider *Ross's* framework (see [118]). This framework combines agency theory and marketing in a way that makes financial innovation a natural part of the model. Ross views financial markets as institutional markets, at which players' actions are constrained by agency theoretic considerations. These considerations are explicitly expressed by contracts and regulations.

Financial innovation arises as a consequence of the interaction between the demand and supply of market participants which are constrained by agency theoretic considerations. In Ross's framework, new financial products/processes are invented and introduced until the marginal benefit from further financial innovation equals the marginal cost for marketing further innovation.

In short, the process of financial innovation in Ross's framework can be described as: The process starts with some institutions being moved out of their "acceptable sets" by an unanticipated event. Such an event can be that one of the stimuli for financial innovation that will be discussed in subsection 2.2.2 alters the boundaries of institutions' acceptable sets. New financial products/processes are now invented in an effort to bring the institutions back into their acceptable sets, at the lowest possible cost. In Ross's framework, the institutional agency structure permits a new financial product/process to remain long after the incentive that first stimulated its development has disappeared.

Above, three frameworks for financial innovation that have appeared in the financial literature have been described. The process of innovation has, however, inspired an extensive amount of research in a large number of fields other than financial economics. For example, in an excellent description of the history of diffusion[7] research, *Rogers* mentions nine different research traditions (see [117] pp.38-86). These traditions are: anthropology, early sociology, rural sociology, education, public health and medical sociology, communication, marketing, geography, and general sociology. Rogers also gives narrative descriptions of, and important references to, each of the nine research traditions.

There are, of course, also important research fields other than those discussed by Rogers, in which much research concerning the innovation process has been carried out. Among these, the theory of organizations and economics can be mentioned. A good survey, with many references, of innovation research in the first of these two fields is given by *Kimberly* (see [84]). Three good textbooks concerning innovation research in the field of economics are [37], [77] and [128]. A large part of the innovation research in all research traditions/fields mentioned here aims to give insights into the process of physical product/process innovation.

It is beyond the scope of this paper to integrate the above mentioned research traditions/fields into the context of financial innovation. It is, however, important to point out that extensive work with an aim of providing a better understanding of the process of innovation has been carried out in many research traditions/fields other than financial economics. Researchers interested in financial innovation certainly have much to learn from the research in those traditions/fields.

As a matter of fact, some attempts to use the knowledge about the innovation process for physical products/processes in order to gain knowledge about the innovation process for financial products/processes have been made. Let us briefly compare two aspects of financial innovation to innovation of physical products/processes. These aspects are science push versus demand pull and the advantage of being first.

[7] According to *Rogers* (see [117] p.5), "*diffusion* is the process by which an innovation is communicated through certain channels over time among the members of a social system".

The distinction between demand pull and science push is important in physical product/process innovation literature. When a new product/process is developed as a response to demand from customers, it is called demand pull. Science push is when scientific results make it possible to develop a new product/process that customers do not yet want (because the customers do not have any knowledge about the product/process). Demand arises, however, when the product/process is presented to the market (e.g., see [77] pp.33-36 and [111]).

Both demand pull and science push play important roles for financial innovation. Clearly, exchanges' and financial intermediaries' contacts with customers provide the ideas for many financial innovations (see [112]). Thus, a large number of new financial products/processes are developed by demand pull.

A large number of substantial breakthroughs in financial theory have been made during the last 30 years. These theoretical breakthroughs have stimulated the development of many financial products/processes; cf. *Bernstein* (see [9]). Thus, much financial innovation is also stimulated by science push (more about academic work's stimulus for financial innovation in subsection 2.2.2).

For physical products, a theoretical model by *Schmalensee* (see [121]) demonstrates that pioneering brands are more likely to succeed than brands subsequently introduced in a market characterized by product differentiation and uncertainty over product quality. Empirical results that somewhat support Schmalensee's model are given in [6]. That it is important to be first to introduce a new product is also the inference from a model by *Glazer* (see [58]).

Being first seems also to be very important for financial innovations. *Black and Silber* (see [11]) show that the level of success of pioneering futures contracts is significantly higher than the level of success of later "me-too" product designs. According to Black and Silber's taxonomy, a pioneering futures contract is a new contract that provides significantly improved risk reduction possibilities and/or different cash flow stream as compared to existing contracts. A me-too product is a duplicate of a contract that already exists on another exchange.

One reason for the importance of being first introducing a new financial product is urgency in obtaining volume, since volume provides liquidity and reduces transaction costs (see [11]). Being first simplifies the efforts of achieving volume. An explanation of this last fact is: As trade starts in the pioneering product, volume and liquidity increase. A reputation of liquidity in the pioneering product arises. After this, it is hard for other exchanges to attract customers to start to trade in their me-too products with uncertain liquidity (see [111] and [113] p.152).

Black (see [10]) shows that each exchange can be viewed as trying to maximize the volume of trading in the contracts it introduces. Thus, Black and Silber (see [11]) conclude that product success in the futures industry can be measured by the volume of trading in the particular contract.

Empirical results presented in [11] also clearly indicate that exchanges are aware of the advantage of being first. The results in [11] show that more than 70% of the new financial futures contracts that were introduced between 1975 and 1982 can be classified as pioneering products. Thus, despite the fact that innovation and introduction of pioneering products are much more costly than introduction of me-too products, a large majority of new financial futures contracts were of the pioneering type. This reflects, of course, the fact that the high cost strategy is the most successful one.

Another reason for the importance of being first is that the innovator of a new financial product/process becomes known as the expert. Hence, it can be hard for followers to attract customers to their me-too products/processes (see [111]).

Being first is therefore very important for a successful introduction of many financial innovations. Two other characteristics of a successful financial innovation mentioned in the literature will also be mentioned very briefly.

Firstly, when examining successes and failures of international financial instruments during the postwar era, *Dufey and Giddy* conclude: ".... in order to succeed, the innovation must involve a genuinely new combination of standard financial elements and must bridge a gap arising from some change in the regulatory or economic environment" (see [45] p.46).

Secondly, most successful financial innovations are profitable[8], e.g., by reducing transaction costs or taxes. The profit need not be divided equally between issuer and investor. None of the parties should, however, be worse off compared to before the introduction of the innovation (see [111])[9].

After this discussion, we are now ready to go back to the aspects of financial innovation that were mentioned in the outset of section 2.2.

2.2.1 The role of CCA in the process of financial innovation

The range of applications of CCA is very broad. This is true in particular when it comes to the pricing of financial instruments. An enormous amount has been written, both by academics and practitioners, on the economics and mathematics of CCA and dynamic portfolio-replication. A fairly recent book by *Merton* provides excellent and up-to-date reference material (see [103]).

Due to its flexibility, CCA is also important in the theory of financial intermediation[10]. In more detail, financial intermediaries rely on the power of modern computer technology and highly-skilled personnel, trained in advanced contingent-claims pricing[11]. In this

[8]The term "successful" means an innovation that becomes commonly used. Profitability can, of course, be regarded as a success in its own right.

[9]There is also considerable evidence that profitability greatly enhances the diffusion of physical product/process innovations (see [84]).

[10]In short, financial intermediation is the process of transforming financial assets from one form to another. Financial intermediaries such as banks, insurance companies, and mutual funds perform this transformation by buying financial assets and issuing other financial assets which are contractual obligations or liabilities of the intermediaries (see [106] p.1).

[11]According to *Merton*, a financial intermediary can produce a new financial instrument by underwriting both sides of the transaction (agent), by synthesizing it through a dynamic trading strategy (principal), or by using a combination of both (see [106] p.21). In this subsection, however, financial intermediaries are assumed to use the dynamic trading strategy, i.e., act as principals.

With this assumption, *Merton* describes financial intermediaries' main task as follows: ".... their primary function is to act as principals and provide financial instruments and products that cannot be efficiently supported by trading in organized secondary markets. .... A stereotypical financial intermediary context CCA provides the "blueprints" or production technologies that financial intermediaries use to manufacture financial instruments.

In short, with the help of CCA and dynamic portfolio-replication, a financial intermediary can construct "simple" securities[12] in the least costly way[13]. These simple securities can then be used as "building blocks" when a financial intermediary constructs and prices its financial products. To liken the simple securities to building blocks is appropriate, because most financial services customers do not demand any of the simple securities individually. Nevertheless, once a financial intermediary has constructed and determined the production costs for a "complete" set of simple securities, the simple securities can be put together to construct and price other financial instruments demanded.

Another, more direct way to view the production process is as follows[14]:

Customer demand/capital markets (demand pull) or financial engineers at the intermediary (science push) determine the specifications of the instrument to be constructed. If the intermediary agrees to issue (and is able to sell) the instrument according to the specifications, a customer liability is created.

To determine what price to charge for the instrument, and also to hedge the liability, experts at the financial intermediary use CCA to design a dynamic trading strategy to synthesize the payoff structure of the customer obligation in the least costly way. The investment required to perform the dynamic trading strategy is the production cost to the intermediary for the instrument.

---

purchases and issues nontraded financial instruments to its customers, and uses the financial markets to raise capital and to hedge its positions." (see [102] p.242).

[12]Ideally, the simple securities would be Arrow-Debreu securities. An Arrow-Debreu security is a security that pays its holder one unit if a particular state of the world obtains at a particular point in time, otherwise it pays nothing (see [103] p.441).

In reality, many of the simple securities are basic financial products that are traded on financial markets, i.e., options, futures etc. (see subsection A.1, for descriptions).

[13]Since financial products cannot be patented, the least-cost producer is likely to have an important advantage (see [106]).

[14]It should, however, be noted that both views are essentially identical.

The production process used by financial intermediaries can be likened to the assembly-line production process in the manufacturing sector (see [102] and [106]). On this "assembly-line", the dynamic trading strategies to synthesize financial instruments provide the blueprints for production, and the simple securities used in the dynamic portfolio-replication are the "raw input". Thus, the dynamic trading strategies prescribe the combinations over time of simple securities that are needed to manufacture an output that matches the cash flow liabilities on the financial instruments.

Merton summarizes the analogy between financial intermediaries' production of financial instruments and the manufacturing sector's production of physical products as follows: ".... by analogy with numerically controlled machines on a physical assembly-line, the intermediary need only change the "dials" (mixing rules) to have the same line produce a different output. This approach thus offers the opportunity to create custom-tailored financial products at a (assembly-line) standard-product level of cost." (see [106] p.29).

In the discussion above, the production technologies provided by CCA and dynamic portfolio-replication appear to be somewhat static. The flexibility and usefulness of the assembly-line improves, however, in a dynamic development. The dynamic development can be described in the following manner:

The production technologies provided by CCA and dynamic portfolio-replication allow financial intermediaries to manufacture new financial instruments. When the successes of the new instruments have attracted sufficiently large volume, they are likely to migrate from intermediaries to markets. An increasing number of disparate traded securities supplies financial intermediaries with a more flexible and wide-ranging assembly-line. That is, when new securities are traded on the financial markets, financial intermediaries can use them in their CCA/dynamic portfolio-replicating production process. This enables financial intermediaries to synthesize cash flow patterns previously not possible to synthesize. This in turn makes it possible for financial intermediaries to further customize new financial instruments to meet the needs of their customers[15].

[15]The above described dynamic development is, of course, a long run (and simplified) vision. In the short run, financial intermediaries can have incentives not to create markets for their new products. This

Merton has called the aforementioned dynamic development a *financial innovation "spiral"* and describes it as follows: ".... as products such as futures, options and securitized loans become standardized and move from intermediaries to markets, the proliferation of new trading markets in those instruments makes feasible the creation of new custom-designed financial products that improve "market completeness"; to hedge their exposures on those products, the producers, financial intermediaries, trade in these new markets and volume expands; increased volume reduces marginal transaction costs and thereby makes possible further implementation of more new products and trading strategies by intermediaries, which in turn leads to still more volume. Success of these trading markets and custom products encourages investment in creating additional markets and products, and so on it goes, spiralling toward the theoretically limiting case of zero marginal transaction costs and dynamically-complete markets." (see [106] pp.19-20).

The aforementioned discussion concerned the important role of CCA as a tool used by financial intermediaries within the dynamic development of financial innovations. The spiral, however, also makes CCA increasingly more important, "independently" of its role in the process of financial innovation. This is due to the fact that, along the path of the financial innovation spiral, an increasing number of complex financial instruments are traded on the financial markets, and many complex financial instruments have attributes that make CCA necessary to use, to value them in a theoretically correct manner. Furthermore, many of the instruments require CCA with several underlying state variables and

is due to the fact that it can be very profitable for them to construct and sell the innovation directly to their customers in bilateral agreements.

If the innovation is successful, however, the volume increases and an organized market is likely to emerge. *Merton* expresses this as: "Financial markets, as we know, tend to be efficient institutional alternatives to intermediaries when the products have standardized terms, can serve a large number of customers, and are well-enough "understood" that transactors are comfortable in assessing their prices. Intermediaries are better suited for low-volume products." (see [106] p.18), and he writes further: ".... financial markets and intermediaries are *competing* institutions when viewed from the perspective of a particnlar product activity. When viewed, however, from the perspective of the evolving financial system, the two are *complementary* institutions, each reinforcing and improving the other in the performance of their functions." (see [106] p.20).

the use of numerical methods to value with accuracy[16]. Naturally, being able to value the increasing number of complex securities is important to all agents on the financial markets, i.e., not only to financial intermediaries.

To summarize, CCA is an important tool that financial intermediaries use in the dynamic development of financial innovation. At the same time, however, the increasing number of complex financial instruments also makes CCA increasingly more important, independently of its role as a production technology for financial intermediaries. This follows from the fact that CCA is needed for the pricing of complex securities, and pricing is important for all agents on the financial markets.

Changes in the economic environment continuously give "new fuel" to the above described financial innovation spiral. In this respect, changes in the economic environment can be viewed as stimuli for financial innovation, which is the subject of the next subsection.

2.2.2 Stimuli for financial innovation

The basic driving forces for financial innovation are to make the financial markets more efficient and/or expand the reach of the markets, and by so doing realize a profit and/or reduce risk. These basic driving forces are, however, not new. Thus, they cannot by themselves explain the explosion of financial innovations during the last two decades.

Changes in the economic environment are the fertilizers that make the soil fertile, in which financial innovation thrives. All opportunities for profitable and/or risk reducing financial innovations would be exhausted in a steady-state market equilibrium. Indeed, changes in the economic environment are the creators of exploitable opportunities, and thus the stimuli for financial innovation[17].

[16]Remember the discussion concerning modelling choices within CCA for convertible bonds, which are fairly simple instruments (at least compared to many of the instruments in appendix A), in section 2.1.

[17]Financial innovations affect, of course, also the environment. There is therefore an interaction between financial innovation and changes in the economic environment (cf. Kane's regulatory dialectic, previously described in section 2.2).

A number of factors (or stimuli), by themselves and in combination, have made the economic environment very favourable for financial innovation during the last two decades. The following factors have been mentioned as important in the literature (e.g., see [9], [45],[51], [52], [62], [78], [79], [80], [81], [107], [108], [111], [113], [123], [133] and [136]):

- intensified competition among financial institutions
- regulatory and tax changes
- technological advances
- increased volatility of interest rates and inflation
- the move to floating exchange rates
- changes in the level of economic activity
- world economic growth
- academic work

Many of the above factors, of course, overlap. In particular, intensified competition among financial institutions overlaps with regulatory and tax changes, increased volatility of interest rates and inflation overlap with the move to floating exchange rates, and changes in the level of economic activity overlaps with world economic growth. For the purpose of a more manageable discussion, the factors are treated separately. Furthermore, the above factors are probably not the only important ones. Next, the factors will be discussed in some more detail.

Intensified competition among financial institutions:

During the 1980's, a deregulation of the financial services industry took place. One result of the deregulation is that boundaries that previously separated different areas of the industry were lowered or disappeared altogether. This means that the distinctions between different financial institutions have become blurred. Market participants have entered niches which the earlier regulation prevented them from entering. Thus, competition has increased.

The intensified competition has forced financial intermediaries to find more efficient production technologies and to reduce production costs. Furthermore, in the new environment of fierce competition, an ability to invent new financial products and processes has become necessary in order to stay ahead of competitors. A reputation for innovation is seen as a key competitive advantage in acquiring market shares.

In this context, it is appropriate to mention that it is not clear what type of market structure is most likely to stimulate a high rate of innovation. According to *Schumpeter*, there is a positive relationship between innovation and monopoly power. Two reasons for believing this are: Firstly, a firm with monopoly power can prevent imitation and thereby can capture more profit from an innovation. Secondly, a firm with monopoly profits is better able to finance research and development (see [77] pp.22-47).

Little empirical support has been found, however, for Schumpeter's hypothesis. A new hypothesis has instead emerged suggesting that a market structure intermediate between monopoly and perfect competition would promote the highest rate of inventive activity (see [77] pp.49-104).

So, perhaps, the market structure of many financial markets that has evolved during the last two decades can be characterized as something between monopoly and perfect competition, rather than close to perfect competition.

Regulatory and tax changes:

Miller considers changes in taxes and regulations as the major stimulus for financial innovation[18] (see [107]). He formulates this role of changes in taxes and regulations in the following way: "There are cases, particularly in the politically sensitive housing area, where the U.S. government has been the major pioneer of new financial instruments. For the most part, however, the role of the government in producing the pearls of financial innovation over the past twenty years has been essentially that of the grain of sand in the

[18] Miller is not, however, so categorical about this in a more recent article (see [108]).

oyster." (see [107] p.461). The grain of sand in Miller's metaphor is, of course, taxes and regulations.

With each change in the tax system or in the framework that regulates the financial services industry, a flurry of activity occurs. Financial innovations are invented to exploit the opportunities for profit and/or risk reduction that are created by the change. It seems, moreover, that every change in the tax system or in the regulatory framework creates an almost infinite number of profit and/or risk reduction opportunities, independently of the competence of the government.

Most governments also seem always to have an inexhaustible amount of reasons when it comes to making changes in the tax system or the regulatory framework. One obvious reason is, of course, to close up the unintentional loopholes in the previous change that gave exploitable opportunities, and therefore led to "undesirable" financial innovations. Another, more obscure reason, is to make the changes in the tax system and in the regulatory framework which strong interest groups want, in an effort to be re-elected (cf. Kane's regulatory dialectic, described earlier in section 2.2).

Many new financial products and processes disappear, of course, when the loopholes in the tax system or regulatory framework that initiated their creation are closed up. Some new products and processes, however, not only survive, but also continue to grow after the loophole has been removed. Any such new device which survives must have made some fundamental improvement to the function of the financial system. Miller calls these new financial products or processes truly significant (see [107]). An example (taken from [107] p.462) of one truly significant innovation (with Miller's terminology) is the following:

A strange regulation in the U.S., called Regulation Q, initiated the start of the Eurodollar market. One of the things Regulation Q accomplished was to place a ceiling on the rate of interest that U.S. commercial banks could offer on their time deposits. The rate ceiling was above, or at least not drastically below, the market clearing rate during the 1950's and most of the 1960's. This changed in the late 1960's and early 1970's, because of the rise in U.S. and world interest rates. In this new economic environment, U.S. money-center commercial banks were forced to search for loopholes in Regulation Q. They found what they were looking for. Regulation Q did not apply to the dollar denominated time deposits in their overseas, and particularly Western European branches. The U.S. banks and their foreign rivals could and did bid competitively for short-term dollar denominated accounts. Moreover, today they still do so to a large extent even though Regulation Q was removed a long time ago.

Technological advances:

Without fast development within the fields of computer and information technology, the explosion of financial innovations could not have occurred. This is equally true of financial products and financial processes.

From the discussion in subsection 2.2.1, it should be clear that financial intermediaries' CCA/dynamic portfolio-replicating production technologies require computers, sometimes with considerable computing power. Thus, many of the new financial instruments that have been introduced during the last decade would never have been possible to manufacture, and especially not to trade in, without the extraordinary improvement and increase in accessibility of powerful desk-top and other types of computers.

The rapid development within computer and information technology has also improved the way the financial system works. Innovations like automated teller machines, electronic security trading etc., have lowered transaction costs and increased the number of possible applications[19]. Electronic providers of price information, such as Reuters, Telerate, Quotron etc., have allowed borrowers and investors to quickly become aware of prices in the market. This has contributed to an explosion in trading volume around the world, and also given markets breadth and liquidity.

Increased volatility of interest rates and inflation:

An economic environment with high interest rates and price volatility creates demand for

[19]From subsection 2.2.1 and the financial innovation spiral, it is also known that lower transaction costs reduce financial intermediaries' production costs, which in turn makes it feasible to manufacture new financial instruments.

hedging against these risks. The financial services industry has responded to this demand with financial innovations.

Thus, a large variety of new financial products, like options, futures, swaps, floating rate notes etc., have been introduced as vehicles for reallocating interest and price risk from issuers and borrowers to others better able to bear them. Moreover, the new more volatile environment has, partly, initiated the trend toward increased negotiability of financial products.

The move to floating exchange rates:

One of the initiating stimuli of the last two decades' explosion of financial innovations was the abandonment (collapse) of the Bretton Woods system of fixed exchange rates. In the regimes after Bretton Woods, the possibilities for changing exchange rates have added a new major uncertainty to all transactions between countries that do not have a fixed exchange rate between their currencies.

The hedging demand that followed from the more volatile exchange rates after Bretton Woods was one important driving force behind the introduction of exchange traded foreign-exchange futures contracts at the Chicago Mercantile Exchange[20]. Later, many other financial products, like swaps, dual currency bonds etc., for handling the exchange rate risk, have also been introduced.

Changes in the level of economic activity:

Financial innovation is stimulated by changes in the economic activity. In their endless

[20]A very vivid narrative of the impulses for the introduction of exchange traded financial futures at the Chicago Mercantile Exchange is given by *Miller* in [107]. In [107], Miller also gives the award for the most significant and successful financial innovation to (exchange traded) financial futures. This choice is, partly, motivated as follows: "Instituting futures trading of foreign exchange , in sum, reduced transaction costs and provided thereby all the classical gains from trade. That alone, however, would not have earned it my award: nomination, or honorable mention perhaps, but not the big prize. What earns it that award is its having served as a model and exemplar for trading in so many instruments in addition to foreign exchange."(see [107] p.466).

striving after profits and expansion, financial institutions are eager to try new financial products and processes in times of economic prosperity. In times of low economic activity, financial institutions stress things like liquidity and risk reduction. Changes in the level of economic activity tend thus to alter not only the magnitude, but also the type of new financial devices.

An illustration of this, which is taken from *Van Horne* (see [133] p.624), is an example from international banking.

Many of the financial innovations during the 1970's, in the international banking area, were aimed at growth. Currency option loans, parallel loans, special swap arrangements etc., were introduced in the pursuit of expanding loan volumes. In the 1980's, with the increasing number of problem loans to developing countries, financial innovations from international banks became more defensive. The emphasis shifted from expansion to risk reduction.

World economic growth:

The possibility that *Miller* finds most persuasive (see [108]) is that the explosion of financial innovations that started in the early 1970's was only a delayed return to the long run growth path of financial improvement. Miller expresses this as: "The burst seems striking only in contrast to the dearth of major innovations during the long period of economic stagnation that began in the early 1930's and that for most of the world continued well into the 1950's." (see [108] p.6).

Even if much of the U.S. regulatory framework, which became the cause of many financial innovations in the 1970's, was put in place in the 1930's and 1940's, there were few financial innovations in the period between 1930 and the late 1960's. During the depressed 1930's, during the war and the slow recovery after the war, there were more urgent matters for creative people to handle. As trade and wealth increased, however, the regulatory framework was becoming more and more of a hurdle for further economic expansion. At last, in the late 1960's, the regulatory framework was becoming so onerous that, among other things, attempts to circumvent it triggered the explosion of financial innovations.

Academic work:

Option Pricing Theory originated from the works of *Black, Scholes* and *Merton* in the early 1970's. The theoretical concepts found in Option Pricing Theory have proved to be very useful and flexible. The concepts also provide the foundation for a new branch of financial economics, namely Contingent Claims Analysis (CCA). CCA is very wide-ranging. Its applications range from the pricing of complex financial instruments to corporate capital-budgeting and strategic decisions.

Simply by referring to the discussions in section 2.1 and subsection 2.2.1 it should be obvious that CCA plays an important role in the context of financial innovation.

Naturally, theoretical work other than Option Pricing Theory and CCA have also provided important stimuli to the process of financial innovation. In the 1950's, 1960's and 1970's, several seminal ideas in the fields of financial and economic theory were presented in academic literature. Among those ideas, the following can be mentioned:

- Calls for the creation of, and analyses of the economic benefits from, a foreign exchange futures market by *Friedman* and others in the 1950's and 1960's.
- The Mean-Variance Portfolio selection model developed by *Markowitz* in the 1950's.
- *Lintner, Mossin* and *Sharpe's* Capital Asset Pricing Model derived with the Mean-Variance Portfolio selection model as foundation in the 1960's.
- Arbitrage Pricing Theory with *Ross* as originator in the mid 1970's.

The impetus from the aforementioned ideas, as well as other academic ideas, to the explosion of financial innovations during the last two decades should not be undervalued; cf. *Bernstein* (see [9]).

It is also likely that academic work has given great impetus to the explosion of financial innovations through the increasing number of graduating students trained in modern finance theory. As these students enter the industry, they prompt financial innovations influenced by the theory they were taught at the business schools.

2.2.3 Are financial innovations beneficial for society?

Before the discussion regarding the social value of financial innovation begins it should be mentioned that the view given in this subsection is predominantly optimistic[21]. It is optimistic in the sense that the process of financial innovation is viewed as successively improving the financial system, and by doing this, the process of financial innovation is believed to also improve the performance of the "real economy".

Merton is a prominent advocate of the optimistic view (see [104], [105] and [106]). He is, however, far from alone in having a predominantly optimistic view of the process of financial innovation. Among other advocates of the optimistic view, *Herring and Santomero* (see [64]), *Miller* (see [107] and [108]) and *Van Horne* (see [133]) can be mentioned.

In opposition to the optimistic view is a pessimistic view. Before the discussion concerning the social value of financial innovation from an optimistic point of view is begun, financial innovation viewed from a pessimistic standpoint will be addressed briefly.

The optimistic view has its point of departure in the assumption of economically rational agents that act on well-functioning financial markets. As a consequence of this market model, financial markets are believed to be (more or less) efficient (e.g., see [36] p.330 ff.). Advocates of the pessimistic view are sceptical about the theory of efficient financial markets. They believe that mass psychology and gambling behaviour are important when prices are determined on the financial markets. With a view like this, agents on financial markets speculate in how other agents are likely to do, rather than using fundamental economic information[22]. From the pessimistic point of view, the speculative character of financial markets can lead to instability which, in turn, can negatively affect the real economy. Furthermore, in this view the speculative character of financial markets is increased by the process of financial innovation. Two theories on this theme will be mentioned.

The first of these theories is *the financial instability hypothesis*, of which *Minsky* is a

[21] Furthermore, the optimistic view influences the whole paper.

[22] One classic reference to this view is *Keyne's* "General Theory" (see [134] p.104).

prominent advocate (e.g., see [109])[23]. Very superficially, the financial instability hypothesis can be described in the following way (for more details, e.g., see [109]):

The mechanisms built into a modern market economy with a complex financial system lead to business cycles. From time to time, the cycles lead to runaway inflation with subsequent debt deflation and deep depression. In this vision, financial innovation can abet excessive expansion by increasing the financing available. Financial innovation enables expansion to lead to inflation. Inflation and the attempts by the monetary authorities to slow inflation lead to increases in interest rates. High interest rates make it hard for borrowers to fulfil their debt commitments. The combination of some borrowers defaulting and the complex liability structures built up by creative financing can give consequences for the financial system that are hard to foresee (in severe cases a complete breakdown of the financial system can occur). The resulting instability in the financial system can have severe influence on the real economy.

The other theory is that of *"bubbles"*. To be more precise, there is no unified theory of bubbles. Rather, there is a set of different theories/models that can be called bubble theories/models (see [85] and [132])[24]. A bubble can, in short, be described in the following fashion:

For some reason, prices start to rise just because anticipations of further price rises. These types of price movements become self-fulfilling prophecies. As this continues, investors suffer "speculative fever", and the prices move further and further away from the values that can be motivated by fundamental valuation principles. This cannot go on forever. Eventually, some investors that are "insiders" or have an eye on the fundamentals start to withdraw. When this first wave of investors starts to withdraw, others begin to realize

[23]According to Minsky (see [109] p.40), the financial instability hypothesis has deep roots in economic theory. The theory, Minsky asserts, can be traced back to economists such as John Stuart Mill, Alfred Marshall, Knut Wicksell, Irving Fisher and John Maynard Keynes. Minsky also believes that some of it can be found in the writings of later Marxists, e.g., Rudolf Hilferding.

[24]Perhaps, Minsky's financial instability hypothesis also belongs to the class of bubble theories.

the absurdity of the price levels and a panic to get out develops. The bubble bursts. If much of the speculation is financed by borrowing, the rapid value decreases that occur when a bubble bursts lead to defaults on debt-services and bankruptcies. In severe cases, a total collapse of the financial system with devastating effects on the real economy can be the result.

Furthermore, in the pessimistic view, there are two arguments against financial innovation that are often mentioned. Firstly, the financial services industry with its high salaries and prospects of great profits attracts a large number of very talented people. Clearly, this leads to a cost for society if these people could be placed for better use in the society (see [134]).

Secondly, if corporations are too concerned about how the stock market reacts, their behaviour will become too short-sighted, some argue. With this short-sightness corporations will invest in short financial investments, rather than in long-lived risky real investments. The result is a slower increasing GNP than would otherwise occur (see [134]).

After this brief discussion of a pessimistic view regarding financial innovation, it is time to return to the optimistic view.

An excellent and thorough analysis of the functions of the financial system is given by *Merton* in [106] (also see [104] and [105]). Merton makes the following claim (see [106]):

- The financial system's principal task is to facilitate the allocation and deployment, both intertemporal and spatial, of economic resources in an uncertain world.

  Savings are distributed, via the financial system, to firms which use them for capital investments. The resources are then returned to the households via the financial system (through security repurchases, dividends and interest payments) for consumption or further recycling. That is, the basic cash-flow cycle is made possible by the financial system.

Merton further claims that, from the principal task, it is possible to distinguish the following six core functions performed by the financial system (see [106]):

1. A payment system for the exchange of goods and services is provided by the financial system.

2. The financial system provides a mechanism for the pooling of funds to undertake large-scale indivisible enterprises.

   That is, the financial system facilitates the creation of resources that can be invested in large enterprises. This also means, however, that the financial system allows individual households to take part in large-scale indivisible investments.

3. The financial system provides a way to transfer economic resources through time and across geographic regions and industries.

   That is, a well-developed financial system facilitates efficient life-cycle allocation of household consumption and efficient allocation of resources to its most productive use.

4. The financial system provides means for controlling and redistributing risks among agents. Thus, risk-sharing and risk-pooling opportunities are supplied by the financial system.

5. The information provided by the financial system plays a crucial role in coordinating decentralized decision-making in a market economy. Households use interest rates and security prices when deciding how much to save and how to allocate their savings among competing investment alternatives. The same price system guides managers of firms when choosing investment projects and their financing.

6. A well-functioning financial system offers means to resolve asymmetric information problems that occur in some financial transactions.

From the aforementioned functions, it is obvious that a well-running financial system is imperative for a prosperous real economy. In the optimistic view, (true) financial innovations improve the financial system[25]. Thus, (true) financial innovations are beneficial for society.

[25] *Van Horne* uses the classification that a new financial product/process must be able to improve the

Merton captures the importance of financial innovation in the following metaphor: "financial innovation is viewed as the "engine" driving the financial system toward its goal of improving the performance of what economists call the "real economy"." (see [105] p.12).

Until now it has only been claimed that financial innovations can improve the financial system, without discussing how the improvements can be accomplished. Now it is time to discuss this. Thus, financial innovations can improve the financial system in (at least) the following three ways (see [105], [107] and [133]):

- Expanding the reach of the market (i.e., making the market more complete) with the help of new financial products. This increases the possibilities for agents to reduce and redistribute risks (through hedging, risk-sharing and risk-pooling), and to find more desirable patterns of payoff streams (i.e., improve the financial system's ability to transfer funds both through time and space).
- Reducing transaction costs or increasing liquidity.
- Lowering agency costs. This can be done through reducing/eliminating information asymmetries, monitoring costs, moral hazard problems and adverse selection problems.

Many of the advanced products/processes that can be observed on today's financial markets represent skilful financial engineering to improve the financial system through one or more of the three ways described above. That is, financial engineers in the financial services industry tailor new financial products/processes to the needs of their customers. The new products/processes provide previously nonexisting tools to accomplish more desirable cash flow patterns, tools for risk management, tools for decreasing agency costs,

---

financial system to be viable as a (true) financial innovation (see [133]). In this paper, a new financial product/process is called financial innovation, even though it is not completely sure that it is a (true) financial innovation. When it is, however, necessary to point out that only new financial devices that also improve the financial system are referred to, (true) financial innovations are explicitly written.

tools for reducing transaction costs etc., all in an effort to reduce the overall cost of capital for borrowers.

At this stage, an example that shows how a financial innovation can improve the performance of the financial system can be helpful for clarification. The example is taken from *Merton* (see [105] pp.17-18), and considers the case of exchange-traded futures and options on stock indexes.

Well-functioning markets for futures and options can be likened to a gigantic insurance company (see [108]), in that options and futures exchanges have greatly reduced the costs for an agent to find counterparties to whom to transfer (for a fee) whole or a part of the agent's risk exposure. This risk management function is also the main role of options and futures on stock indexes.

In general, the risk management function can be accomplished by three basic methods (see [105]). Futures and options on stock indexes provide means for handling risk by all three basic methods.

The first basic method is to move from risky assets to a riskless asset, that is *reducing risk by selling the source of it.* A portfolio manager can use this risk management method to a much lower cost using futures and options on a stock index than directly transacting in the spot cash market. This follows from the fact that "selling" the source of the risk with the help of futures and options on a stock index induces lower costs than simultaneously selling all the individual stocks.

The second basic method for risk management is *reducing the risk by diversification.* This is truly one of the fundamentals of financial economics. Through deploying the resources across a large number of securities, an agent can reduce the risk exposure of his portfolio. It is costly, however, to change the risk exposure of the portfolio by trading in small quantities in each security separately. Use of derivative securities on a stock index can accomplish the same change in risk exposure as trading in individual stocks, but with less cost and increased speed. Moreover, since there are bounds on the subdivision of the units of individual securities, the number of securities that can be held are limited for a given level of wealth. Thus, futures and options on a index give the possibility of a broader diversification by permitting arbitrary small amounts of the index's individual components.

The third basic factor for managing risk is *reducing risk by buying insurance against losses.* Insurance can be bought for a premium, and has the property of retaining the advantages of owning an asset while removing the possibility of unfavourable outcomes. This is exactly what a put option on the asset accomplishes {identical insurance can be created with a call and a bond, through put-call parity (e.g., see [67])}. Thus, a put (or a call) option is effectively an insurance contract. Therefore, options on a stock index can be used to insure a portfolio against broad stock market declines. Besides the differences in transaction costs, the insurance that options on a portfolio provide is also genuinely different from a portfolio of options on the individual securities in the portfolio (see [56] and [57]).

From the above discussion, it can clearly be concluded that futures and options on indexes offer investors risk management tools that are not available without these instruments. Options and futures on indexes also considerably decrease the transaction costs of making adjustments in a large portfolio. Thus, derivative securities on an index improve the financial system in two of the three ways earlier mentioned. Index futures and options have, however, the potential of improving the financial system in the third way as well. The reasoning behind this is given in the next paragraph.

There are many reasons for an investor to trade a stock. The investor may want to change his risk exposure or to revise his positions according to changed beliefs about the stock market in general. But the investor can also trade in an individual stock because of private information concerning that particular stock. The possibility of uninformed potential counterparties being "tricked" by informed traders tends to increase uninformed traders' reluctance to trade. The reluctance to trade leads to higher bid-ask spreads, which is a dead-weight loss to the uninformed traders. The possibility to trade in market aggregates, however, resolves this "market impact" problem. This follows from the fact that the opportunity to trade in market aggregates (e.g., futures and options on stock indexes) allows uninformed traders to adjust (broadly) their portfolios without having to trade in the individual stocks.

In the optimistic view, improvement of the financial system by financial innovation is a dynamic process in which new financial products and processes continuously make the financial system better and better. Consider the following simplified illustration:

Assume there is a financial innovation that reduces transaction costs. The innovation that reduces transaction costs certainly improves the financial system. The lower transaction costs, however, also make it feasible to construct and introduce further financial innovations. These new financial innovations improve the financial system, which in turn makes it feasible to construct further financial innovations. It continues like this, moving the financial system toward the theoretically limiting case of zero marginal transaction costs and dynamically complete markets. *Merton* has named this process *a financial innovation "spiral"*, which has already been discussed in subsection 2.2.1.

According to advocates of the optimistic view, the case of futures and options on indexes, and many, many others should be enough to convince most people that the continuous stream of new financial products and processes are good for society in the long run. Still, advocates of the optimistic view also admit that not all of the new financial products and processes are (true) financial innovations (i.e., improve the financial system). This problem, from an optimistic starting-point, is the next to be discussed[26].

Today's rapidly changing environment has made people receptive to new products

[26] At the outset of this subsection, a pessimistic view of financial innovation was discussed. The discussion in the remainder of this subsection addresses the negative effects of new financial products/processes that are not (true) financial innovations, from an optimistic point of view. Even though some parts of this discussion can appear to be similar to the things said in the discussion of the pessimistic view, there is a difference in the magnitude of the problems.

In the pessimistic view, financial innovation has negative effects on the real economy. Moreover, in the pessimistic view, financial innovation can in severe cases contribute to crises in both the financial system and the real economy. In the optimistic view most new financial products/processes have positive effects on the real economy. There are, however, also negative effects from some new financial products/processes in the optimistic view. It is these negative effects that will be discussed next.

and ideas. The complexity of the changing environment has also increased the need and intensified the search for financial innovations by investors and borrowers. In this environment, there are bound to emerge new financial products and processes which are good as well as bad.

The thirst for financial innovations by financial-service consumers has allowed investment bankers and other financial institutions to introduce devices disguised as (true) financial innovations. After a while, when the disguise has deteriorated, there is little or nothing to be found that can improve the financial system. Furthermore, the (excessive) zeal for new financial products and processes has created a situation where the benefits to financial-service consumers from (true) financial innovations are in some cases either eliminated or substantially reduced by promotion fees charged by financial intermediaries.

The fact that it is possible to introduce "untrue financial innovations" disguised as (true) financial innovations and that financial intermediaries sometimes can charge excessive promoting fees, can in part be explained by financial-service customers acting to some extent like a "shoal". As an example, if one corporation issues a new exotic product then other corporations will often follow. They do not want to miss any possible titbit. Moreover, they also are willing to pay for it.

Van Horne has called the outcome of the shoal behaviour a "balloon" effect (see [133]). The metaphor of a balloon is, of course, inspired by bubble theories (bubble theories are discussed earlier in this subsection). A balloon is briefly described in the next paragraph.

Sometimes when a new financial device is introduced expectations and promotion fees are blown up, but not to the extent that the balloon bursts. Later, when the first excitement has disappeared, a slow decline of the excessive fees and demand/supply is seen. The air is let out of the balloon. The devastating effects that occur when a bubble bursts, however, are not seen. The following example, taken from *Van Horne* (see [133] p.628), is an illustration of the balloon effect:

In the early 1980's there was an 85 percent tax exemption of dividend income for corporations in the U.S.. During the period 1982 to 1984, an explosion of new financial instruments, e.g., adjustable rate preferred stocks (ARPS) and convertible adjustable preferred stocks (CAPS)[27], were developed for the purpose of taking advantage of this loophole in the tax legislation. The eagerness to exploit the loophole created an excessive demand that allowed financial intermediaries to take very high promotion fees for these products. This was particularly evident for ARPS. At the peak in February 1983, the adjustable return on an ARPS security was nearly 500 basis points below the yield on the appropriate benchmark treasury security. At this point, the air started to rush out of the balloon, and the negative differential declined. The balloon was further tapped by the 1984 Tax Act in the U.S..

What then are the costs of balloons and new financial products and processes that are not (true) financial innovations but disguised as such?

Firstly, there are the obvious costs from the sometimes excessive promotion fees. These out-of-pocket costs reduce, and in some cases eliminate the benefits from (true) financial innovations to financial-service customers.

Secondly, there are the potentially much larger costs of distortions of resource allocations. To the extent that new financial devices do not lead to improved economic performance, there is a waste of human resources. The human talent that is needed to develop these non-fruitful devices could be put to better use in society. Moreover, the sometimes unrealistic expectations among investors and issuers concerning the likelihood that new financial devices will give higher expected return with less risk, can lead to misallocations of investment capital. That is, capital resources can be invested in a suboptimal way which can impair the long run economic performance.

The costs discussed above, although not large compared to the potential benefits from (true) financial innovations (at least, from an optimistic standpoint), should not be ignored. Legislation is, however, supposedly not the best way to correct the negative sides of the process of financial progress[28]. Instead, it is more likely that legislation (in any

[27]ARPS and CAPS are described in sub-subsection A.3.1.

[28]This is not to say that all legislation is bad. A stable and thoroughly considered general regulatory framework is important to give long-run rules for players at financial markets. A very important public policy issue especially with respect to innovation, is to handle the conflict between the financial

case, the type of short-sighted and populistic legislation that is often observed) prohibits the development of many financial innovations with potential to significantly improve the financial system, and also stimulates development of new financial products/processes that are not (true) financial innovations.

In the optimistic view, much of the corrections must come from the market participants themselves. *Van Horne* expresses this as follows: "No, the ultimate discipline must come from the market, but hopefully an increasingly better informed and suspicious market — a market able to respond rationally and effectively to the cascade of financial ideas thrust upon it." (see [133] pp.629-630).

2.2.4 The pace of financial innovation, a brief look into the future

Many financial economists share the opinion that the stream of new financial products and processes will continue. Many financial economists also believe that the stream will continue to be fast. Much depends, of course, on what happens with the factors that stimulate the development of financial innovations (see subsection 2.2.2). It is, however, not only important what happens to each factor separately. It is also important how the factors develop relative to each other.

As previously mentioned, a changing economic environment constitutes the basis for financial innovation. For example, if a country were to leave its regulatory framework unchanged (ceteris paribus), eventually a point might be reached where marginal benefits

"infrastructure" {i.e., the institutional interfaces between intermediaries and financial markets, regulatory practices, organization of trading and clearing facilities, and management information systems (see [106] p.49)} and some of the new financial products/processes. The conflict is due to the fact that many new financial products/processes can be implemented rather quickly, while changes in the infrastructure usually take longer to implement. The difference in implementation time can cause problems, or as *Merton* puts it: "It is indeed possible that at times, the imbalance between these two elements could become large enough to jeopardize the very functioning of the system. Hence, the need for policy to protect against such breakdown, without unnecessarily inhibiting innovation." (see [106] p.50).

Potential market failures at financial markets also exist. These might justify regulation (see [87]).

from and marginal costs of further financial innovation balance each other exactly. Thus, an equilibrium might emerge where all incentives to exploit loopholes in the regulatory framework were exploited. This will, of course, never happen. Politicians never seem to lack reasons, and the temptations often appear to be overwhelming to change the regulatory framework (cf. Kane's regulatory dialectic, previously described in section 2.2)[29]. This is likely to be equally, or even more true in the future. This is due to the scattered political landscapes and the many strong interest groups that can be observed in most industrial countries.

During the last decade, competition among corporations in the financial services industry has increased. There are several reasons for this, of which two will be mentioned. Firstly, deregulation has eliminated niches which were previously safe, and diminished the difference between different types of financial intermediaries. The areas of business overlap are expanding. Secondly, advances in information technology have increased the global competition in the financial services industry.

The process of increasingly tougher competition is likely to continue. It will be very important for financial firms to develop new financial products/processes in this environment with stiff competition[30].

Consider the volatility of financial prices. Volatility of financial prices is one very important stimulus for financial innovation, so the forecast of the evolution of the volatility is important for the forecast of the future rate of financial innovation. Some factors point towards greater stability. There is increased cooperation among the central banks

[29]One recent article in "The Economist" (see [48]) claims that financial professionals are "panting behind" in their attempts to price and assess the risk of derivative products. The article also claims that the less than full understanding of derivative products by financial services firms is a potential risk for the financial system. According to the article, this fact worries financial regulators, and tighter control of derivative products seems inevitable.

[30]In this context, it is important to recall the discussion about market structure and innovation in subsection 2.2.2. In subsection 2.2.2, it was suggested that it is not sure that increased competition leads to an increased rate of innovation. It is the author's belief that most financial markets, however, have not yet reached the level of competition (if there is one) which gives the fastest rate of innovation.

of the Group of Seven (G7) countries on exchange rate stabilization policies (see [46]). There is also the long term movement toward a tighter exchange rate mechanism within the European Monetary System (EMS), with the possibility of eventual currency union. These things, however, can change, and as a matter of fact, there has already been some change. During the late summer and fall of 1992 an increased tension among the G7 countries, as well as severe problems for the EMS were observed. The problems for the EMS led to realignments of the exchange rates within the EMS, and certain currencies (sterling and lira) were disconnected from the EMS. Today, a currency union (at least in the form previously conceived) does not seem to be possible to install until in the far future, if ever.

Furthermore, there are many other factors which point towards continued high volatility of financial prices. Three of these factors will be mentioned in the remainder of this paragraph. Firstly, there are still large divergencies between many countries' growth rates, inflation rates and balance of payment flows. Oscillating interest and exchange rates are likely outcomes from these divergencies. Secondly, tension exists between the U.S. dollar as reserve currency and the persistent current account deficit of the U.S., which forces the country to attract capital inflows from abroad (see [46]). Thirdly, there is a potentially destabilizing impact on many industrialized countries from the integration of the former Soviet republics and the East European countries with the economies in the West[31]. In fact, these three factors are important causes of both the recent tensions within G7, and the problems for the EMS discussed above. In summary, it is reasonable to expect financial prices to also be volatile in the future.

Progress within the areas of computer and information technology and telecommunication is faster than ever. This progress will help to create better working financial markets, not the least of which will be by reducing transaction costs and improving electronic providers of economic information. The progress within computer and information technology, however, will also improve financial intermediaries' production technologies, as

[31]The integration of Eastern and Western Germany has already led to tensions within the EMS and among the G7 countries.

discussed in subsection 2.2.1. The improved production technologies will, ceteris paribus, make it possible to manufacture financial instruments not previously feasible.

Finally, the research within the field of financial economics is intense, and is likely to continue to also be so in the future. This research will, presumedly, improve financial intermediaries' production technologies and thereby, exactly as the progress within computer and information technology, make it possible to, ceteris paribus, manufacture financial instruments not previously feasible. It is also likely that the research will affect the demand side. That is, as theoretical results reach financial-services customers/capital markets new financial products suggested by the theory will be demanded.

From the discussion in this subsection, it can be concluded that there is every reason to believe that a rapid rate of financial innovation will also be observed in the future.

2.3 Concluding remarks

During the last two decades there has been a fast pace of financial innovation. As a consequence of this, there has also been an explosion in the number of financial instruments that are traded on the financial markets. There are, furthermore, many reasons to believe that a rapid rate of financial innovation will also be observed in the future.

In this paper, it is shown that the importance of CCA, in the context of financial innovation, is twofold. This is due to the fact that CCA provides a "production mechanism" that can be used by financial intermediaries to construct new financial instruments, while at the same time, the importance of CCA successively increases as the number of complex financial instruments that can be found on the financial markets increases.

To be more precise, CCA and dynamic portfolio-replication provide financial intermediaries with production technologies that allow them to manufacture new financial instruments. When the successes of the new instruments have attracted sufficiently large volume, they are likely to migrate from intermediaries to markets. An increasing number of disparate traded securities supplies financial intermediaries with more flexible and wide-ranging production technologies. That is, when new securities become traded at the financial markets, financial intermediaries can use them in their CCA/dynamic portfolio-replicating production process. This enables financial intermediaries to synthesize cash flow patterns previously not possible to synthesize. In turn, this makes it possible for financial intermediaries to further customize new financial instruments to meet the needs of their customers. It continues onward, moving toward the theoretically limiting case of dynamically-complete markets.

The dynamic development described above, however, also makes CCA increasingly more important independently of its role in the process of financial innovation. This is due to the fact that along the path toward the limiting case of dynamically-complete markets, an increasing number of complex financial instruments are traded on the financial markets[32], and many complex financial instruments have attributes that make CCA necessary to use to value them, in a theoretically correct fashion.

One problem that often arises when CCA is used, is that it is not possible to find a closed-form solution for the value. Numerical methods must therefore often be relied on. Furthermore, in many cases, and especially in cases where there are more than one underlying state variable, numerical methods become computationally laborious. Due to this last fact, it is unfortunate that many complex financial instruments require CCA with several underlying state variables and the use of numerical methods to value with accuracy[33].

Naturally, being able to value the increasing number of traded complex securities has an academic interest. Such valuation methods are, however, certainly not only of interest to academics. For the agents on the financial markets, it is of crucial importance to be able to assess the value of complex financial instruments. This is equally true for financial-services firms that construct and promote the products, borrowers that sell the products for the financing of their activities and investors that buy the products.

[32]In appendix A, more than 100 more or less complex financial instruments that exist (or have existed) on the financial markets are described.

[33]Remember the discussion concerning modelling choices within CCA for convertible bonds, which are fairly simple instruments (at least compared to many of the instruments in appendix A), in section 2.1.

Let us conclude with the following inference:

From the above discussion it should be clear that research concerning, and development of, efficient numerical methods that can be used in the CCA context is an important research area. Furthermore, this research area is likely to become even more important in the future.

Appendix:

A New financial products

During the last two decades there has been an explosion in the number of financial instruments that are traded on the financial markets. This appendix provides descriptions (brief) of more than 100 more or less exotic financial instruments.

Although many financial instruments are mentioned in this appendix, the list is far from exhaustive. Since the list of financial instruments is not complete, some kind of selection has been performed. In short, the selection procedure which was used can be described as follows:

The first step was to carry out a literature search. The work on this appendix was started during my stay at *The Wharton School of the University of Pennsylvania* in the spring of 1992. A natural starting point for the literature search was to use the University of Pennsylvania's comprehensive literature data base. The search in the data base was performed by searching on key terms, e.g., financial innovation, financial product/instrument, derivative product/instrument etc.. Beside the search in the University of Pennsylvania's literature data base, all issues from the mid 1980's of certain relevant journals were scanned. The journals were: *The Continental Bank: Journal of Applied Corporate Finance*, *The Journal of Portfolio Management* and *Financial Management.* After this, the rest of the literature search was made by picking important references from the reference lists of the literature found by using the data base and scanning the journals[34].

The next step was to select financial instruments from the literature found in the literature search. This instrument selection was carried out by the simple, and not particularly scientific, principles: on the basis of what the author judged as important, on the basis of what the author found interesting, and according to available information.

[34]Among the sources that were found in the literature search, in particular two articles by *Finnerty* (see [51] and [52]) and a book by *Wamsley* (see [136]) have been of great help when writing this appendix.

Of the financial instruments described in the appendix, a fairly large number were first introduced in Europe or Japan. The main part of the financial instruments in a survey like this, however, are of course developed in the U.S.. This is due to the fact that the U.S.'s large and well-developed financial markets have been the birthplace of a majority of all financial innovations during the last twenty years.

A rather large number of more or less complex financial products will be described in subsection A.1 through subsection A.4. Many of these products are variants of the "traditional" instruments, i.e., equities, preferred stock, convertible bonds and warrants, and there are also often basic derivative products, i.e., forwards, futures, options and swaps (see subsection A.1 for a discussion of basic derivative products), added to these variants. To put it simply, as a first step, most exotic financial products can be viewed as packages of traditional instruments and basic derivative products.

Viewing complex/customized financial instruments as packages of traditional instruments and basic derivative products is called the "building block" approach (or the LEGO approach) (e.g., see [125])[35]. The LEGO approach provides an excellent framework for structuring the thinking of both when complex/customized financial instruments should be valued and when they should be constructed[36].

From the above discussion it should be clear that, to understand the new exotic financial products, it is essential to first gain a knowledge about the traditional instruments and the basic derivative products. Thus, for the purpose of making it easier to follow the rest of the appendix the traditional instruments will now be described[37] superficially, even

[35]Actually, as will be seen later, the traditional instrument convertible bond can be viewed as a package of an ordinary bond and a warrant, from the perspective of the LEGO approach. Furthermore, if the convertible bond also has put and/or call features (see sub-subsection A.2.2), as convertible bonds usually have, it is "only" a matter of adding put and/or call options to the package of an ordinary bond and a warrant.

[36]It should be noted that the LEGO approach can be viewed as a simplified (or applied) version of the more theoretical discussion concerning financial intermediaries' production technologies in subsection 2.2.1.

[37]The descriptions of the traditional instruments are based mainly on [136].

though they are well-known (basic derivative products are discussed in subsection A.1).

- **Equities:** The certificate of incorporation settles the common stock of a company. This defines the shares, that is the number and type of identical units in which the stock is split. Shareholders are part-owners in the company, and thereby also part-owners in the company's assets and possible profits/losses. Shares have no fixed maturity. Furthermore, they are claims on the residual of the assets of the firm after all other creditors have received their claims.

  Shareholders have the right to participate in the long-run management of the company. They do this through appointing the board of directors. The shareholders also have the right to vote directly on major changes in the company.

  A dividend is paid to shareholders, if one is declared and approved at the meeting of shareholders. The value of a share is the present value of the future stream of dividends.

- **Preferred stock:** In the same manner as common stock, preferred stock is a share in a company's assets and profits/losses. Preferred stock differs, however, from common stock in several respects. These differences make preferred stock hybrid securities. They are debt-like equity.

  What then is the difference between preferred stock and common stock?

  Firstly, the dividends that should be paid to preferred stockholders are predetermined. Preferred stockholders also have the right to receive their dividends in full before common stockholders receive any dividend at all. Furthermore, traditional preferred stock usually has cumulative attributes. That is, if preferred stockholders are not paid the full dividend, the part that is not paid out accrues to the benefit of the preferred stockholders. Common stockholders do not receive any dividends until preferred stockholders have been paid all accumulated dividends that the company has previously failed to pay.

  Secondly, in almost all cases preferred stock has higher priority than common stock in a liquidation of the company.

Thirdly, in general, the corporation that issued the preferred stock has the right to buy back the preferred stock at a predetermined price (the company has a call option on the preferred stock).

- **Bonds:** Since traditional bonds are highly well-known, only a very brief description will be given. This should not, however, mislead one to believe that there is not much to be said about bonds. On the contrary, there have been numerous books and articles written about this type of instrument.

  If the risk of possible default is disregarded, traditional bonds are instruments with little uncertainty. This is due to the fact that a traditional bond has a fixed rate of interest and has also a predetermined maturity date. Thus, the bondholder knows beforehand with certainty (default risk is still disregarded) both size and date for all cash flows during the life of the bond. Yet, one remaining uncertainty is the price that the bondholder can receive if he/she sells the bond before maturity.

  As will be evident from subsection A.2, there have been many extensions and variants of the traditional bond.

- **Convertible bonds:** Like preferred stock, convertible bonds are hybrid securities. Preferred stock is debt-like equity whereas convertibles are equity-like debt.

  The traditional convertible is a fixed-rate bond with a right (no obligation) to convert the convertible into common stock of the issuer according to specified terms added to it (usually that the convertible can be exchanged for shares at a predetermined price). At conversion the issuing company issues new shares, thus creating a dilution of the equity.

  From the above description, it is clear that the value of a convertible consists of two parts. Firstly, the value of the fixed-rate bond. Secondly, the value of the right to take part in sufficiently large increases in the value of the common stock.

- **Warrants:** A traditional warrant gives the holder the right (but not the obligation) to buy a specified number of shares of the issuing company's common stock at a predetermined price. Furthermore, the holder (normally) has the right to make the purchase at any time before or on a specified expiration date. Thus, warrants have a very close relationship with American call options on a stock (see subsection A.1). The major difference is that the option is written on already existing shares while warrants are written on yet un-issued shares of the issuer of the warrants. Thus, if warrants are exercised, the total number shares in the company will increase and there will be dilution of the equity of the issuer.

Usually, warrants are issued as a "sweetener" attached to a bond or a preferred stock issue. Moreover, warrants can normally be detached from the parent issue and, thus, be traded as a separate instrument.

From the previous description of convertible bonds, it is evident that a bond issue with warrants attached to it is similar to a convertible bond issue. One difference is, as just described, that warrants (normally) can be detached and traded separately. Another difference is that when warrants are exercised, new capital is usually paid for the shares, while debt capital is simply shifted into equity capital when a convertible is converted (convertible bondholders have to give up the bond to exercise the warrant).

A.1 Basic derivative products and some extensions

The basic derivative products are forwards, futures, options and swaps. The characteristics of these products are well-known, and are discussed in many textbooks (e.g., see [46], [67], [92] and [136]). This subsection will in any event start with a very brief description of forwards, futures, options and swaps.

- A forward transaction is one in which a party agrees to buy (or sell) an asset at a specific price at a specific time in the future. A forward contract is a customized agreement between two parties.

- London International Financial Futures Exchange (LIFFE) defines a financial futures contract as: "an agreement to buy or sell, on an organized exchange, a standard quantity of a specific financial instrument or a foreign currency at a future date and at a price agreed between two parties." (see [46] p.71). From this definition, it can be seen that futures contracts and forward contracts are similar on many points. There are, however, four main differences between a forward agreement and a futures contract. These differences are (see [46] and [92]):

- Futures contracts are openly traded on recognized exchanges. Each party clears the trade through the exchange. That is, each party has credit risk only with the exchange. Moreover, each counterparty to the initial trade can clear the trade with any number of other counterparties.
- Futures contracts are based on standardized amounts.
- Futures contracts have standardized delivery rules and delivery dates.
- Futures contracts use margining arrangements to help avoid credit risk problems. Positions are valued daily and additional margin is required when position values decline.

- The basic characteristic of an option is that ownership carries a right, but not an obligation. A *call option* gives the owner the right, but not the obligation, to buy the underlying asset by a certain date (the expiration date) for a certain price (the exercise price). A *put option* carries the right, but not the obligation, to sell the underlying asset by the expiration date for the exercise price. Options are further divided into *American options*, which can be exercised at any time up to the expiration date, and *European options*, which can only be exercised on the expiration date itself.

- A swap is a derivative product that involves the swap of one type of cash flow (say, floating or fixed payment in U.S. dollars based on an interest rate index of some maturity) with another type of cash flow (say, fixed or floating payment in sterling based on an interest rate index of some maturity). The cash flows are calculated according to a formula that depends on the values of one or more underlying variables (say, an exchange rate and/or an interest rate index of some maturity).

The risk management function is the main role of derivative products in the financial system. The development of well-functioning markets for options, futures, forwards and swaps has made it possible for corporations, and others, to hedge against a large variety of risks that were impossible or far more costly to hedge against before. This improvement of the financial system is due to the fact that derivative products and the markets where they are traded have made markets more complete as well as reduced transaction costs.

The "financial revolution", during which we have seen an explosion of new financial products, has only lasted for approximately two decades. Exchange-traded options[38], forwards and futures contracts, however, have a much longer history[39]. In the 17th century, the Dojima rice market in Osaka was a forward market. This market developed into a fully organized futures market by the 18th century (see [105]). Options and contracts similar to futures accounted for most of the transactions on the Amsterdam stock exchange in the 17th century (see [105] and [108]). In Frankfurt in 1867 and in London in 1877, organized futures markets were created (see [105]). In the 1920's, options on commodity futures (under the name of "privileges") were traded at the Chicago Board of Trade (CBT)[40]. In one of its drives against speculation, the U.S. Congress prohibited trade in these options in the 1930's. During the 1920's, common stocks were also traded at CBT in ways that had some points in common with futures style market arrangements.

Although important in their times, the impact of these early markets for forwards, futures and options cannot be compared to the impact of the regeneration of organized trading in derivative products in the 1970's. After a long struggle with regulators, the International Money Market (IMM) [an offshoot to the Chicago Mercantile Exchange

---

[38] Options are indeed, a very old invention. According to Aristotle, Thales offered an option on olives in the 6th century BC (see [48]).

[39] Agreements similar to swaps also have a somewhat longer history. In [108], *Miller* writes that British overseas travelers invented vacation-home swapping. These were agreements in which a couple of free weeks of occupancy of a flat or house in Britain could compensate for a corresponding stay abroad. One important factor behind the invention of these agreements was that capital controls severely limited the amount of currency British citizens could take abroad.

[40] CBT was, however, founded as early as 1848 (see [105] and [107]).

(CME)] was allowed to open futures trading in seven foreign currencies in May 1972. Soon thereafter (in 1973), and after an even harder (at least longer) struggle with regulators, the CBT was allowed to open the Chicago Board of Option Exchange (CBOE) and the trading in options on 16 different stocks. In many ways, the introduction of these financial futures and options at IMM and CBOE were the events that started the last two decades' rapid stream of new financial products.

Since 1973, there has been an dramatic increase in exchanges (all around the world) where futures and options can be traded. There has also been an explosion in the number of different types of underlying variables that options and futures are written on. Nowadays, it is possible to trade in options and futures on a very wide range of commodities, stocks, debt instruments (interest rates), currencies (exchange rates) and indexes. There are also options on other derivative products available, like options on futures and options on swaps (swaptions).

As already mentioned, forward and futures contracts have many similarities. Forwards are also bought and sold for many of the same reasons as futures are traded. As with futures, forwards are common on several different underlying variables. Forwards on interest-bearing instruments and currencies are the dominating types, and these types of forwards are bought and sold in very large volumes in many countries all over the world. Moreover, since forwards are customized products, they can to some extent be tailored to suit the exact needs of the counterparties. Partly due to this flexibility, forwards are not normally traded on an exchange. There are, however, organized markets for many types of forward agreements although they are not exchanges.

From the very brief description of a swap arrangement given earlier in this subsection, it can be understood that there is no limit to the number of different kind of swap arrangements that can be invented. Indeed, many types of swaps are also used. Without going into any details (the interested reader can instead refer to [46], [67], [92] or [136]) plain vanilla swaps, amortizing swaps, step-up swaps, step-down swaps, forward swaps, extendable swaps and cancellable swaps can be mentioned.

As for the variables underlying swap arrangements, interest rates and currencies are those which totally dominate. Currency swaps are the older of the two, while interest-rate swaps are transacted in larger volume nowadays.

The technique of using currency swaps was introduced in Britain in the 1960's. Due to exchange controls, U.K. companies were prevented from directly borrowing overseas. The introduction of currency swaps enabled companies to circumvent the exchange controls and thereby raise long-term funds abroad. The next important step for the volume of currency swaps, and also for the legitimacy of swaps in general, came in 1981. This step was the start of the World Bank's extensive currency swap program. Salomon Brothers estimated that currency swaps linked to new issues grew from $1 billion in 1981 to $35 billion 1986 (see [136]). The currency swap market also continued to grow very quickly after 1986.

Interest-rate swaps emerged in 1981 and early 1982 (see [136]). Since 1982, the interest-rate swap market has undergone tremendous growth. International Swap Dealers Association (ISDA) estimated that the total volume of interest-rate swaps transacted in 1988 was $568 billion notional par amount (see [92]). Since 1988, the volume has increased further. An estimate of the volume outstanding 1991 is approximately $2.8 trillion (see [47]).

During the last few years, basic swaps have changed from being customized transactions to becoming standardized products traded in organized markets (see [105]). Nowadays, notional amounts of currency and interest-rate swaps outstanding at the over-the-counter (OTC) markets are in the trillions of dollars (see [47]) and rising rapidly.

From the previous discussion, it is obvious that the development of well-functioning markets for forwards, futures, options and swaps was extremely important in itself. The importance of well-organized markets for these basic derivative products becomes, however, even more pronounced if their role is considered in the explosion of new financial products during the last two decades.

As previously mentioned in this appendix, many of the new exotic financial products are basically traditional instruments (like bonds, stocks etc.) with features that closely correspond to the features of the basic derivative products added to them. Development of well-organized markets for forwards, futures, options and swaps was absolutely essential for the emergence of these new exotic financial products. This is due to the fact that these markets enable financial-services firms to hedge their own risk exposures associated with the provision of the new complex financial products[41]. The well-organized markets for basic derivative products have thus made possible and stimulated the growth of different kinds of financial products (of which many are custom-designed). In subsection 2.2.1, financial intermediaries' use of traded securities when constructing new customized financial instruments is discussed in some detail.

Many of the most exciting of the recent extensions of the basic derivative products are option-based. This subsection will be concluded with brief descriptions of some of these new option-based products. These products have been developed in the fixed income, foreign currency and equity markets, and they are representative for the new portfolio management products that have revolutionized the financial markets (see [92]).

- **Interest-rate caps:** Interest-rate caps are derivative securities that provide protection for floating-rate borrowers against increases in interest rates, while allowing them to profit from decreases in interest rates. Below, *instantaneous interest-rate caps* {there also are types called *average interest-rate caps* and *hybrid caps* (see [67])} will be described briefly.

  To specify a cap structure a number of terms have to be defined. These are: the length of the cap protection, the interest rate level at which protection will be provided (the strike rate) and the frequency of protection (usually 3 or 6 months, these periods are called reset periods). The function of an interest-rate cap now is as follows:

[41]The increased use, by financial-services firms, of derivative products for hedging purposes is partly confirmed by the following figures:
The Bank for International Settlements (BIS) estimates banks' outstanding volume of exchange-traded derivatives to $3.5 trillion at the end of 1991. This can be compared to $583 billion at the end of 1986. BIS also estimates that banks' outstanding volume of over-the-counter derivatives was $4.1 trillion in June 1991. This volume was "only" $500 billion at the end of 1986 (see [48]).

At the beginning of each reset period, during the length of the cap protection, the prevailing interest rate is compared to the strike rate. If the prevailing interest rate is higher than the strike rate, the owner of the cap should receive a payment from the seller of the cap. This payment is calculated by multiplying the difference between the prevailing interest rate and the strike rate with both the length of a reset period (expressed in years, if the interest rates are annual rates) and the notional amount of the cap. If the prevailing interest rate is lower than the strike rate the cap owner should not receive any payment. There is one more important aspect. If any amount should be paid, it is paid at the end of the reset period (even though the amount, if any, is determined at the beginning of the reset period).

If the payoffs from an interest-cap are viewed more closely, it can be seen that the cap can be viewed as a series of European call options on the interest rate with the payoffs from the options occurring the length of one reset period in arrears (see [67]).

- **Interest-rate floors:** The principle of an interest-rate floor is very similar to the principle of an interest-rate cap. An interest-rate floor, however, provides protection against decreases in interest-rates. Interest-rate floor contracts are equal to interest-cap contracts except for the fundamental difference that the holder of a floor receives payments only if the interest rate is lower than the strike rate (at the resetting dates). Moreover, if a cap can be viewed as a series of European call options then a floor can be viewed as a series of European put options on the interest rate with payoffs occurring the length of one reset period in arrears (see [46]).

- **Compound options:** A compound option is an option on an option. For example, a European call (or put) option on an option gives the holder the right, but not the obligation, to buy (or sell) an option for a prearranged price at the expiration date.

  When a hedger buys a compound option, a premium must be paid. This initial premium is called a front fee. If the hedger exercises the compound option and actually (assume that it is a call option on an option) buys the underlying option, the hedger must pay the prearranged price. This prearranged price is called a back fee. The feature of compound options that (still assume a call option on a option) a hedger has to pay two premiums at two different times if the hedger actually buys the underlying option has led to compound options often being referred to as *split-fee options* (see [92]).

Compound options are a fast developing product. Call options on call options or put options can be traded. Put options on call options or put options can also be purchased (see [92]). Compound options can be written on gold options, FX (foreign exchange) options, bond options, swaptions, caps (then called captions), collars[42] and floors (see [92]).

- **Spread options:** Spread options are an interesting new risk management instrument. The option type has the value of a relationship between two different assets as underlying variable.

  Options on spread relationships between non-dollar interest rates and those of the U.S. can be traded. These options are also called *spreadtions* . A spreadtion gives limited risk and unlimited profit potential that cannot be duplicated with any combination of bond options (see [92]).

- **Options on average prices:** Options on average prices are also referred to as *Asian options.* The value at expiration of an Asian option is determined by comparing the (arithmetic) average of the underlying asset's prices, say the average of daily closing prices, during the life of the option to the exercise price. An Asian call option has a value at expiration equal to the average price minus the exercise price if the average price exceeds the exercise price, otherwise it is worthless. The value of an Asian put option at expiration is determined in a similar way.

  The most common use of Asian options are to hedge foreign exchange risk, and if so, primarily by multinational corporations that receive payments of foreign currency spread out over time (see [92]).

[42] A collar is a one-for-one combination of a cap sale (purchase) with a floor purchase (sale) (see [92]).

- **Lookback options:** A lookback option is an option with a stochastic exercise price. At expiration, the exercise price is set to be equal to the most favourable price of the underlying asset (for the option holder) that actually occurred during the life of the option (this explains the name lookback option). That is, for a lookback call option at expiration, the exercise price is set to the lowest price of the underlying asset during the life of the option, and similarly, the exercise price is set to the highest price for a lookback put option.

  Few lookback options have actually been traded. One reason for this, beside the newness of the instrument, is that lookback options do not hedge away common risk positions (see [92]).

- **Up-and-out and down-and-out options:** A down-and-out option is identical to an ordinary option except that it disappears if the price of the underlying asset reaches a specified level that is below the exercise price. If the price of the underlying asset has once reached the specified level, the option will not be "reborn" if the price of the underlying asset should start to trade above the specified price again. An up-and-out option is similar to a down-and-out option, except that this option disappears if the price of the underlying asset reaches a specified level that is above the exercise price.

  The most common of the up-and-out/down-and-out options are down-and-out call options and up-and-out put options, although up-and-out call options and down-and-out put options can also be traded. Up-and-out/down-and-out options are, however, rather new products and have to date not been traded to any great extent (see [92]).

- **Variable-Quantity options:** An ordinary call (or put) option that is in-the-money has a payout at expiration that increases (or decreases) one-for-one with the price of the underlying asset. A variable-quantity option has a payout structure at expiration that differs from the payout structure of ordinary options.

A variable-quantity option is an option for which the amount of the underlying asset to be delivered, if exercised, changes at specified levels of the price of the underlying asset at expiration. More precisely, in a variable-quantity option contract a quantity and a stepsize are specified. For every step level the option is in-the-money at expiration, the amount to be delivered is reduced with the quantity[43]. This continues until *[(# step levels in-the-money) × quantity = max. amount of the underlying asset to be delivered (the max. amount to be delivered is specified in the contract, and is the amount to be delivered at the lowest step level in-the-money)]*, after that there is no amount to be delivered and the option does not give the holder any privileges (see [92] for a more thorough description of variable-quantity options).

- **Basket options:** A basket option is an option on the sum of the assets that belong to the basket. That is, a call option on a basket is in-the-money if the sum of the values of the assets in the basket is higher than the exercise price. Thus, the option can be in-the-money even though some assets in the basket have had a fall in value as long as values of other assets have risen enough to compensate for this.

  One very common type of basket options are options on stock indexes[44]. Options are available on indexes for whole stock markets as well as on particular sectors {e.g., computer technology, oil and gas, transportation, telephone (see [67])}.

  Option-based products on baskets and indexes can become fairly complex. Below is an example (taken from [92] p.87) of a quite complex warrant.

  On the behalf of one of its customers, the Bankers Trust issued a zero-coupon Eurobond in February 1990. Attached to this bond was a warrant on a basket of five stock indexes. The indexes were Germany's FAZ, France's CAC 40, the Netherland's EOE, Italy's BCI and Switzerland's SMI. One further complicating

[43]The exercise price is specified as a price per unit of the underlying asset.

[44]Options on stock indexes have already been discussed in subsection 2.2.3. It was concluded there that options on stock indexes (and other indexes as well) had the potential of being beneficial for the society. Moreover, it was also concluded that index options offer risk management opportunities that are not available without them.

feature of this warrant was that the average of different indexes in the basket was composed only of the appreciation of the real indexes. If one index depreciated during the option period, zero was used to calculate the average.

Another common type of basket option are options on a basket of currencies. This product is favourable for companies with a continuous portfolio with currency risk (see [92]).

A.2 Selected debt innovations

During the two last decades, a tremendous number of new debt instruments has been introduced. A selection of more than 60 of these debt innovations is briefly described in this subsection.

Although many new debt instruments defy classification, they are divided into four classes in this subsection. Thus, in sub-subsection A.2.1 a selection of floating-rate debt instruments are described. Sub-subsection A.2.2 treats zero coupon bonds and developments/extensions of the zero coupon concept. In sub-subsection A.2.2 several instruments that are not of zero coupon type, but do not fit in any other sub-subsection, are also described. Asset-backed debt products are discussed in sub-subsection A.2.3. Finally, sub-subsection A.2.4 handles debt instruments linked to commodities, exchange rates and indexes.

A.2.1 Floating-rate securities

The use of floating-rate securities increased substantially during the 1980's. One reason for this increased use was the volatile and high level of interest rates. Another reason was, of course, that investors became more and more familiar with these rather novel instruments. In this sub-subsection a selection of floating rate instruments will be briefly described. This selection is listed in table 2.1.

The oldest type of floating-rate security is the *Floating-Rate Note (FRN)*, or the *floater*. The first issue of FRNs was in the Euromarket in 1970 (see [136]).

SELECTED DEBT INNOVATIONS
Floating-Rate Securities
• Accelerated-Coupon FRN • Capped FRN • Collateralized FRN • Deferred-Coupon FRN • Delayed Cap FRN • Drop-Lock FRN • Flip-Flop • Floating-Rate Note (FRN) • FRN Convertible into Fixed-Rate Bond • FRN Convertible into FRN in another Currency • FRN with Warrant into Bull FRN • FRN with Warrant into Equity • FRN with Warrant into Fixed-Rate Bond • FRN with Warrant into a Second Currency • Mini-Max (Collared) FRN • Mismatch FRN • Partly Paid FRN • Perpetual FRN • Puttable Perpetual FRN • Rating Sensitive FRN • Step-Down FRN • Step-Up FRN • Variable Spread FRN • Yield Curve Anticipation Note (YCAN)

Table 2.1: Selected Debt Innovations: Floating-rate securities.

FRNs appear in very many variants. The basic characteristic of FRNs is that the coupon rate floats with an interest rate index. Some of the interest rate indexes that have been used are Libor (London interbank offered rate), Libid (London interbank bid rate), Limean (average of Libor and Libid), Treasury bill rate and rate fixed every 49 days by Dutch auction (see [136]). The floating coupon rate leads to a reduced principal risk for the lender by transferring interest risk to the borrower.

In addition to the many interest rate risk indexes used for FRNs, there have also been many different rate-setting methods. Libor FRNs originally reset every 3 months if paying 3-month Libor, every 6 months if paying 6-months Libor etc.. This is no longer always true.

If the resetting period differs from the length of the interest rate, it is called a *"mismatch"* formula, e.g., a floater that is priced at a spread over 6-month Libor with a monthly resetting of the Libor.

Another rate-setting method is the *variable-spread* formula. The formula is best illustrated by an example. The example is the first issue of a variable-spread FRN, and is taken from [136] p.191.

In December 1984, the Nordic Investment Bank (NIB) issued a floater with the rate-setting formula: The rate was to be fixed quarterly at a spread over the T-bill rate. The spread itself was fixed by taking 55 % of the differential between Libor and the T-bill rate with a minimum of 35 basis points. This formula was chosen because NIB wanted to diversify its funding sources by borrowing at a rate linked to the Treasury bill rate, without risking the FRN to weaken much in price relative to other (Libor based) FRNs if the T-bill rate should fall relative to Libor.

There have been numerous modifications/variants of the aforementioned "classical" FRNs. Some of these will be briefly described. In February 1985, the *"mini-max"* was introduced, with an issue for the Kingdom of Denmark (see [136]). A mini-max FRN is a FRN paying a minimum and a maximum coupon. That is, the coupons float with the underlying interest rate index. There is, however, an upper boundary and a lower boundary for each coupon payment. When this type of instrument linked to the 3-month T-bill rate was introduced in the U.S. domestic market in June 1985, it was instead called a *"collared"* FRN (see [136]).

Soon after the first mini-max FRN issue, *"capped"* FRNs were introduced. A capped FRN is a FRN with an upper boundary for the coupon rate. That is, even though the interest rate index to which the FRN is linked rises above the upper boundary the coupon rate payable never exceeds the upper boundary. After the capped FRN the next step was to introduce the *"delayed cap"* FRN. A delayed cap FRN works as an ordinary FRN during the first period after the issuing date, usually 2 or 3 years. The cap on the FRN does not become effective until after this first period.

A *perpetual* FRN is, exactly as the name suggests, a FRN without any maturity date. The first issue of a perpetual FRN was in 1984 by the National Westminster Bank (see [136]). In 1980, however, Citicorp had already introduced the *"puttable perpetual"* FRN (see [136]). This instrument is a perpetual FRN with the extension that it can be redeemed at the option of the noteholder on interest fixing dates after an initial time period. Soon after the issue of the first perpetual FRN, the *"flip-flop"* was introduced. The flip-flop is an instrument that starts as a perpetual floater, but allows the investor to "flip" out of the perpetual into a shorter-term note and then "flop" back again. The first issue of a flip-flop was for the Kingdom of Sweden in 1984 (see [136]). The majority of issuers of perpetual FRNs and its modifications have been banks. This is due to the fact that many countries have allowed banks to classify perpetuals as primary capital.

There have been a number of issues of convertible floaters and floaters with warrants attached. Some of these option-related FRNs are *FRN convertible into fixed-rate bond, FRN convertible into FRN in another currency, FRN with warrant into bull FRN* (bull FRNs are described below), *FRN with warrant into equity, FRN with warrant into fixed-rate bond, FRN with warrant into a second currency* and *drop-lock FRN.* All of these instruments, except drop-lock FRNs, have self-explanatory names. A drop-lock FRN converts automatically into a fixed-rate bond, once a certain lower level of the underlying interest rate index is reached. Drop-lock FRNs never became popular. Only two issues were made, the first one in 1979 (see [136]).

The source for many and often also exotic modifications of the classical FRN are variations of the coupon structure. Many of these modified coupon structures represent skilful financial engineering to meet specific, not seldom Japanese, investor requirements.

Yield Curve Anticipation Note (YCAN), later also called *"bull"* or *"reverse"* floater, was one of the first of these floaters with modified coupon structure. The first issue of an YCAN was in 1986 by the Student Loan Marketing Association (Sallie Mae) (see [136]). An YCAN bears an interest rate which equals a specified rate minus Libor (this explains the name reverse floater). It can be noticed that it is possible to combine an YCAN and a classical floater so they add up to a fixed-rate bond (see [136]).

A *deferred-coupon* FRN pays no coupons or has low coupon rates (relative to the underlying interest rate index) in early years of the floater and high coupon rates (relative to the underlying interest rate index) closer to maturity. A typical deferred-coupon FRN is given in the following example (see [136] p.195):

In June 1986, Banque Nationale de Paris (BNP) issued a 5-year floater that did not pay any interest the first two years and then paid Libor + 450 basis points the last three years.

As opposed to deferred-coupon FRNs, *accelerated-coupon* FRNs have also been issued. Not surprising, an accelerated-coupon FRN has high coupon rates (relative to the underlying interest rate index) the first years and low coupon rates (relative to the underlying interest rate index) closer to maturity.

Coupon structures that are related to deferred/accelerated-coupon FRNs, and are also more commmon, are *step-up* and *step-down* FRNs (another name for these instruments is *split spread* FRNs). Step-up/step-down FRNs are notes on which the interest margin over the specified benchmark (e.g., Libor) steps up/down to a larger/smaller margin on specified dates during the life of the instrument. One rationale for split spread FRNs is to sweeten issues that have maturities that are longer than normal. An example (taken from [136] pp.197-198) is the following:

In June 1986, the British merchant bank Hill Samuel issued a 30-year note. A 30-year note is normally considered long for a British merchant bank. To sweeten this issue, it paid Libor + $\frac{1}{2}$% for the first 5 years, Libor + $\frac{3}{8}$% during the next 5 years, and Libor + $\frac{1}{4}$% for the last 20 years. The issue is also callable at par during the last 20 years.

Another innovative FRN structure is the *partly paid* FRN. This FRN structure is best described by an example (taken from [136] p.201).

The first partly paid FRN was issued in June 1985 for BNP. The notes were offered through a conventional FRN syndication to an authorized list of 250 participating banks.

The notes are registered and tradable, but only between the 250 banks authorized to underwrite the issue. This partly paid FRN was a $600 million issue, but only $100 million was initially drawn. The underwriting banks are, however, committed, if they hold any of the original $100 million issue, to subscribe to any amount of the remaining $500 million that the borrower may issue during the 10-year life of the deal. Since the FRNs are registered, BNP can identify the banks that are holding them. The structure has created a committed back-up issue facility for BNP that is, at the same time, tradable. So the underwriting commitment is liquid.

Collateralized FRNs are rather new instruments on the FRN market. The issuer of a collateralized FRN must maintain a fixed amount of collateral on deposit with the issue's trustee. The percentage (of the FRN's nominal value) collateral depends on the type of collateral, how often the collateral is marked to the market, and how long the issuer is given to bring collateral back to acceptable levels if it depreciates. Many of the collateralized FRNs have been issued by U.S. thrift institutions (see [136]).

The last floating rate instrument that will be briefly described is *rating sensitive* FRN. As an example, this type of security was issued by Manufacturer Hanover in 1988 (see [34]). In that design, the issuing corporation agreed to pay investors a spread above LIBOR that increases with decreases in the rating of the issuing company's senior debt and the other way around. At first glance, it may look as if this instrument compensates the holders for increases in the risk. This is, however, only partial compensation since the instrument actually increases the probability of default for the issuing company (instead of reducing it). This is due to the fact that rating sensitive FRNs increase the issuing company's debt-service payments when the company can least afford it and vice versa (see [34]).

A.2.2 Zero coupon bonds and several other bond variants

In this sub-subsection, a selection of financial products with zero coupon features are described. A number of debt products that are not of zero coupon type are also discussed. The debt instruments that are handled in this sub-subsection are listed in table 2.2.

A small number of *zero coupon bonds* were issued as early as in the 1960's. One

SELECTED DEBT INNOVATIONS
Zero coupon bonds and several other bond variants
• Certificates of Accrual on Treasury Securities (CATS)
• Extendable Bond
• Growth and Income Securities (GAINS)
• Liquid Yield Option Note (LYON)
• M-CATS
• Medium-Term Note (MTN)
• M-TIGR
• Puttable Bond
• Retractable Bond
• Separate Trading of Registered Interest and Principal of Securities (STRIPS)
• Tap Bond
• Treasury Investment Growth Receipt (TIGR)
• Zero-Coupon Bond

Table 2.2: Selected Debt Innovations: Zero coupon bonds and several other bond variants.

example is the issue in the Euromarket for BP Tanker-Eriksberg in June 1966 (see [136]). In reality, however, the zero coupon market did not begin until the early 1980's. The first public issue was for J.C. Penney in April 1981 (see [136]).

Zero coupon bonds are non-interest-bearing instruments. A zero coupon bond does not make interim payments. It has only one cash-flow, which is at maturity. The structure of zero coupon bonds eliminates reinvestment risk. Interest is effectively reinvested and compounded over the life of the debt issue at the yield to maturity at which the investor purchased the bond. This elimination of reinvestment risk is one reason for issuing zero coupon bonds. Another reason, prior to 1982, was that zero coupon bonds allowed issuers and investors in the U.S. to benefit from tax-arbitrage. After a change in the U.S. tax code, which eliminated tax-arbitrage possibilities using zero coupon bonds, the number of zero coupon bond issues has decreased substantially (see [136]).

The introduction of receipt zero coupon products was the next step in the development of the zero coupon market. Among the most renowned of these products are *Treasury Investment Growth Receipt (TIGR)* and *Certificates of Accrual on Treasury Securities (CATS)* (see [136]). The idea behind these products is explained in the remainder of this paragraph. The investment bank involved purchases an amount of Treasury bonds. Next, the investment bank sells receipts that give evidence of ownership of the different interest payments and principal payment respectively. In this way, the ownership of principal and interest can be separated. Consider this example (from [136] pp.214-215).

Suppose an investment bank purchases $100 million worth of 10-year Treasury bonds, with a coupon rate of 10% ($5 million payable semi-annually). Thereafter, the investment bank sells 20 sets of certificates on $5 million each, maturing after 6 months, 1 year, 18 months, and so on (i.e., one set of certificates for each coupon payment). $100 million of receipts that provide evidence for the ownership of the principal also are sold.

CATS and TIGRs were followed by the introduction of *Separate Trading of Registered Interest and Principal Securities (STRIPS)* in February 1985. The principles for STRIPS are that selected Treasury securities are held in the book entry system operated by the Federal Reserve in a way that allows separate trading and ownership of the interest and principal payments.

The opportunity to separate interest and principal payments for municipal bonds in the U.S. was created by the 1986 Tax Reform Act (see [136]). Soon thereafter a number of products, e.g., *M-CATS* and *M-TIGRs*, were introduced in order to achieve this separation of interest and principal for municipal bonds. Most of those products have the original Treasury receipt issues, e.g., CATS, TIGR and STRIPS, as prototypes.

Growth And INcome Securities (GAINS) is an instrument that is a zero coupon bond during the first part of its life. After this first part, a GAINS converts to a conventional bond (it also remains a conventional bond during the rest of its life). GAINS was introduced by Utah Municipal Power Systems in March 1985 (see [136]).

A further development of the zero coupon concept came in April 1985 with Waste Management Inc.'s issue of *Liquid Yield Option Notes (LYONs).* A LYON is a zero coupon bond convertible into common equity. The LYON is both callable and puttable. The instrument is made even more complex by the fact that the prices at which the issuer may call the bond, as well as the prices at which the investor may redeem the bond, escalate through time. As an extra flavour, the LYON contains call protection. This is accomplished by the fact that the bond may not be called for a specified initial time period unless the underlying stock price rises above a prespecified level (see [95]).

This sub-subsection will be concluded with short descriptions of a number of debt variants that are not of zero coupon type. These debt variants are described in this sub-subsection, more due to the fact that they do not fit in any other sub-subsection rather than that they are related to zero coupon debt instruments.

Medium-Term Notes (MTNs) are (often) classical fixed-rate corporate bonds. The innovative aspects of MTNs are their rather short maturities, and that they are offered in continuous medium-term note programs. With the help of an MTN program, a corporation can sell individual notes in varying amounts (often relatively small amounts) and varying maturities (from 9 months to 10 years). MTN programs have a great deal in common with commercial paper (CP) programs. One difference is MTNs' longer maturities. Another difference is that MTNs (in the U.S.) are registered as to principal and interest while CPs are sold in bearer form. Besides being issued with fixed-rate coupon, MTNs have been issued with floating-rate coupon (see [136]).

A *puttable* bond is a bond that the investor can sell back to the issuer on prespecified dates at prespecified prices. A puttable bond will only be exercised if the interest rates rise or if the issuer's credit standing deteriorates. Thus, in a sense a puttable bond gives the holder an option on interest rates as well as an option on the issuer's credit spread (see [34]). Puttable bonds are fairly common. Furthermore, put option features are added to many different bond types nowadays. An example of a puttable bond is given below (taken from [136] p.224).

In 1972, the National Bank of Hungary issued $50 million at a coupon rate of $8\frac{1}{2}\%$. The first put date was in November 1979, and the put price was 100%. The bond could also be redeemed at 100% in 1980, 1981 and 1982. After 1982, the redemption prices changed year by year. More precisely, the redemption prices were 101.1% in 1983, 100.2% in 1984, 100.3% in 1985 and 93% in 1986. By 1983, 93% of this issue had been redeemed.

Retractable and *extendable* bonds have been used in Canada for a long time. A retractable bond gives the investor the option to redeem the bond at a prespecified date. An extendable bond gives the investor the option to extend the bond to a bond with longer maturity on a prespecified date. A 10-year bond extendable to 30 years is, however, for practical purposes, indistinguishable from a 30-year bond retractable after 10 years (e.g., see [3],[21] or [136]). That is, both retractable and extendable bonds can be included in the broad class of puttable bonds.

A *tap bond* issue is an issue for a specified total amount. Only a part, e.g., 50%, of the total amount is, however, initially issued. The remaining part is issued on demand. The "tap" is turned off and on. Tap bonds have some advantages (see [136]). One advantage is to save commission for the borrower. This is due to the fact that commission is usually only paid on the initial tranche. A second advantage is that the tap issue allows the borrower to quickly respond to increased demand. Another advantage is that subsequent issues increase the liquidity of the original issue.

A.2.3 Asset-backed debt

The aim of securitization is to transform non-liquid assets into tradable and liquid securities. The volume of securitized debt is a large part of the total debt volume. For example, Henry Kaufman of Salomon Brothers estimated in 1987 that the volume of securitized debt in the U.S. was $4.7 trillion out of a total debt of $7.9 trillion (see [136]).

One important form of securitization is to issue securities backed by assets. In this sub-subsection, a selected number of asset-backed debt instruments will be described. These instruments are listed in table 2.3.

Before discussing different asset-backed securities, it can be of value to clarify briefly the distinction between pass-through and pay-through securities.

A pass-through certificate is a share of ownership in the underlying assets, and thus gives the right to a share of the cash flow generated by the underlying assets. The seller of the pass-through securities acts primarily as a servicer that "passes through" the collected principal and interest from the underlying assets to the security holders. Normally, the

SELECTED DEBT INNOVATIONS
Asset-Backed Debt
• Certificates for Automobile Receivables (CARS)
• Certificates for Amortizing Revolving Debt (CARD)
• Collateralized Medium-Term-Notes
• Collateralized Mortgage Obligation (CMO)
• Computer Lease-Backed Security
• Floating Rate CMO (FRCMO)
• Mortgage-Backed Bond
• Mortgage-Backed Pass-Through Security
• Purchased Accelerated Recovery Right (PARR)
• Senior/Subordinated Pass-Through
• Stripped Interest Certificate (STRIP)
• Super-Sinkers

Table 2.3: Selected Debt Innovations: Asset-backed Debt.

pass-through structure in the U.S. is such that the underlying assets are sold to a "grantor trust". Legal title to the receivables is held by a trustee. The trust should, however, be passive and only hold the trust property to protect it. The trustee has only limited managerial freedom over the trust's assets. The pass-through payments should match the incoming payments on the underlying assets (see [136]).

Similar to a pass-through security interest and principal payments on a pay-through security are met by cash flows generated by the underlying assets, "paid through" the issuer. In the U.S., the underlying assets are normally held by a limited purpose finance subsidiary. Now, pay-through bonds are debt issued by the limited purpose finance subsidiary and collateralized by the underlying assets. This means that investors in pay-throughs are owners of bonds secured by the underlying assets, and not (as for pass-throughs) direct owners of the underlying assets. The pay-through structure allows the issuing entity to manipulate, without tax consequences, the cash flows generated by the underlying assets to a larger extent than the pass-through structure allows (see [136]).

With the development of *Mortgage-Backed Securities (MBS)*, the volume of securitized debt reached very large amounts. In the U.S., mortgage securitization began to grow explosively in the 1970's. This led to mortgage-backed securities emerging into an important class of securities during the 1980's, which is confirmed by the following facts (see [70]):

In the U.S., nearly \$2.5 trillion of residential mortgages were outstanding in the second quarter of 1988. Of the outstanding residential mortgages "only" about 25% were securitized through MBS. MBS represent, however, one of the fastest growing segments of the debt markets. In 1979 \$3 billion of MBS was outstanding, this figure increased to more than \$900 billion by the end of 1988. The volume of MBS traded in the secondary market has also increased significantly. The trading volume in MBS after issuance was \$243 billion in 1981. This volume increased to \$1.2 trillion in 1985.

Mortgage-backed securities are instruments that are backed by a pool of mortgages. An investor that purchases a mortgage-backed security buys a pro-rated share of the pool's cash flows. The issuing institution acts as an intermediary, which transfers the cash flows from the mortgage pool to the MBS holders (and retains a servicing fee). MBS are complex instruments, and, thus, difficult to value. One of the complexities that affect the value of MBS is the homeowners' possibility to prepay principal on mortgages within the underlying pool.

Mortgage-backed pass-through securities have been issued in many different forms. They all have, however, many features in common (see [136]). Each pool of mortgages has a coupon or pass-through rate, an issue date, a maturity date, and payment delay. The pass-through rate is the rate paid to the holders of the mortgage-backed pass-through securities. The pass-through rate is often lower than the (various) interest rates on the mortgages in the pool. This rate difference is called the servicing fee, and makes up the issuing intermediary's reward for administering the pass through of cash flows. The pool issue date is the date of issue of the pass-through security, not the issuing date of the mortgages in the pool. The pool maturity date is the date of the latest maturing of the underlying mortgages. There is a time lag between the date on which homeowners are scheduled to make their payments, and the date that the servicing intermediary pays the holders of the mortgage-backed pass-through securities. This time lag is called the payment delay. The payment delay is an important technical feature of the mortgage-backed pass-through security. The yield at a mortgage-backed pass-through security is reduced by the payment delay. The longer the payment delay, the bigger the reduction.

The cash flows from a mortgage-backed pass-through security would be rather predictable if no prepayments were made by the homeowners. Homeowners, however, make prepayments. The uncertain rate of prepayments leads to mortgage-backed pass-through securities having uncertain true maturities. One obvious cause for prepayment is refinancing by homeowners. To make refinancing worthwhile, interest rates have to be low (often lower than the mortgage-backed pass-through security yield). It is the possibility for homeowners to prepay that gives mortgage-backed pass-through securities negative convexity (see [136]).

The next step after the development of the market for mortgage-backed pass-through securities was the creation of the *mortgage-backed bond* market. Mortgage-backed bonds are intermediate or long-term debt instruments. These instruments are, as the name suggests, collateralized by a pool of mortgage loans. A feature of many mortgage-backed bonds is "value maintenance". Value maintenance means that the issuer must pledge more mortgage collateral if the market value of the initial pool falls under specified levels (see [136]).

The debt-service on mortgage-backed bonds is not (in a direct way) closely related to the cash flows from the pool of mortgage loans. A mortgage-backed bond is a direct obligation of the issuer. The issuer has, however, put down a pool of mortgage loans for a trustee to administer. If the issuer does not manage to pay the debt-service, or if the market value of the pool falls below a prespecified level, the trustee is obliged to liquidate the collateral and repay the holders of the mortgage-backed bonds. The debt-service on mortgage-backed pass-through securities (and also on CMOs, which are described later in this sub-subsection) is directly dependent on the cash flows from the underlying pool of mortgage loans. Mortgage-backed bonds are thus much more dependent on the issuer than mortgage-backed pass-through securities (and also CMOs).

The volume of outstanding mortgage-backed bonds has never been very large. Furthermore, in the mid 1980's mortgage-backed bonds lost much volume to *collateralized medium-term notes.* Collateralized medium-term notes are, to put it simply, MTNs (see sub-subsection A.2.2) with mortgages as collateral.

An interesting form of mortgage revenue bond is the *"super-sinker"*. In short, the super-sinker is a particular maturity of a mortgage revenue bond issue. When prepayments of the underlying mortgage loans occur, the first prepayments are used to retire bonds of the super-sinker maturity. This feature reduces, somewhat, the uncertainty due to possible prepayments. This follows from the fact that since prepayments are first used to retire the super-sinker maturity, the cash flow uncertainty is reduced for all other maturities of the issue (at least for the early years of the lives of the bonds). The super-sinker maturity itself can also be attractive for some investors. In particular, those investors who can tolerate some cash flow uncertainty and believe that interest rates will fall while activity in the housing market will increase can find the super-sinker attractive. This is due to the fact that they will be quickly repaid, if their beliefs are fulfilled (see [136]).

Collateralized Mortgage Obligations (CMOs) were introduced by the Federal Home Loan Mortgage Corporation (FHLMC, also commonly referred to as "Freddie Mac") in the U.S. in 1982 (see [136]). CMOs are, in part, a development of the ideas behind super-sinkers. The pool of mortgage loans underlying the CMOs are repacked for the purpose of reducing the cash flow uncertainty. The structure is best illustrated by an example (see [136] pp.257-258) of a typical CMO.

Consider a $400 million pool of mortgage loans. Backed by the cash flows from this pool, a typical CMO issue is constructed as: Four classes of bonds will be issued. Three of these bond issues will be at $100 million each. The first $100 million issue will be the "fast pay" tranche (also called A bonds). The first $100 million of prepayment of the underlying mortgage loans are used to retire the fast pay tranche. The second $100 million issue will be the "medium pay" tranche (also called B bonds). This tranche is retired by the second $100 million of mortgage loan prepayments. The last $100 million issue is the "slow pay" tranche (or C bonds). Not surprising, the third $100 million of prepayment is used to retire the slow pay tranche. The fourth class of bonds are the Z bonds. These bonds have features of both zero-coupon bonds and mortgage-backed pass-through securities. The zero-coupon feature is that Z bonds do not receive any coupon payments until all other classes have been fully retired. After the A, B and C bonds are paid off, the cash flow from the remaining mortgage loans in the pool is used to pay the debt-service on the Z bonds. It can also be mentioned that the simple CMO structure in this example has been refined and developed in several ways.

One rather recent CMO innovation is the *Floating-Rate CMO (FRCMO)*. The first issue of FRCMO was made by Shearson Lehman and Centex Acceptance Corporation in September 1986 (see [136]).

The next step of development in the mortgage-backed market was stripped securities. Stripped mortgage-backed instruments represent developments of the stripping techniques used earlier in the zero-coupon market (see sub-subsection A.2.2). The stripping process in the mortgage-backed market was greatly facilitated by the Real Estate Mortgage Investment Conduit (REMIC) legislation, which is a part of the U.S. 1986 Tax Reform Act (see [136]).

The *STRIPped interest certificate (STRIP)*, first issued in 1986, divides the interest and principal payments into two separate streams of claims. The issuer can allocate any part of the principal, from 0 to 100%, and any part of the interest (that is not reserved for servicing or other expenses), from 0 to 100%, to a given STRIP (see [136]). In the most extreme form, a STRIP can pay interest exclusively, called *IOs* (interest only), or principal repayments only, called *POs* (principal only) (see [52]).

Senior/subordinated pass-throughs were introduced in 1986 (see [136]). A brief description of the construction of these instruments is given in the remainder of this paragraph. The pool of underlying mortgage loans is divided into a subordinate and a senior portion. The cash flows from the subordinate portion are used to satisfy possible losses on the collateral. The idea to use cash flows from a portion of the underlying mortgage loan pool to restore losses in the collateral, allows issuers without an investment-grade rating to obtain an investment-grade financing on the senior part of the financing (see [136]).

Another interesting form of securitization is to pool receivables (e.g., automobile loan receivables, credit card receivables, lease receivables) and then issue securities backed by these pools.

The first issue of securities backed by automobile loan receivables was a $10 million issue of *Certificates for Automobile ReceivableS (CARS)* for Marketing Assistance Corporation in January 1985. The volume of outstanding CARS[45] increased rapidly, and by October 1986 the volume was approximately $9 billion (see [136]).

Usually (but not always), CARS have the following features (see [136]):

- CARS generally have a fixed coupon, with principal and interest paid monthly.
- In CARS issues, a pass-through or a pay-through structure is normally used. ("Fixed-payment" structure[46] has, however, also been used.)
- CARS issues usually have credit protection, which covers losses up to a fixed percentage of the pool balance.
- The weighted average life of CARS issues is normally between 1 to 3 years and the final maturities 3 to 5 years (CARS are usually prematurely redeemed, when the underlying loans are paid off ahead of schedule.).
- The short-term nature of CARS mitigates effects from changes in the prepayment level. Moreover, prepayment rates vary little with interest rates. This means that pass-through and pay-through CARS do not have the negative convexity problems of mortgage-backed securities.
- When the loans underlying CARS age, the prepayment rates increase. The major causes for prepayment are sales and trade-ins. Refinancing is rather rare.

[45]Henceforth, in this appendix, all securities backed by a pool of automobile receivables will be named CARS. CARS is the Salomon Brothers abbreviation for auto-receivable-backed securities.

[46]With a fixed-payment structure, the payments on CARS are separated from the prepayments on the underlying loan pools. An investment contract with an insurance company is used to guarantee that the fixed debt-service can be fulfilled irrespective of prepayments (see [136]).

After the introduction of CARS, it was a small step towards introducing a security backed by a pool of credit card receivables. In March 1986, a private placement of $50 million for Bank One in Ohio was made of *Certificates for Amortizing Revolving Debts (CARDs)* (see [136]). CARDs represent participation in a fixed pool of credit card accounts. Furthermore, CARDs have more or less the same structure and features as CARS.

There have also been issues of other types of asset-backed securities than those mentioned above. Of these, *computer lease-backed securities* and *Purchased Accelerated Recovery Rights (PARRs)* can be mentioned.

Computer lease-backed securities are illustrated by an example (see [136] p.244).

A computer leasing company from Illinois in the U.S., Com-L Corporation, issued $25 million in the form of a private placement of non-recourse bonds. The bonds were payable solely from the proceeds of its collateral. The collateral consisted of security interests in the lease receivables. An security in the leased equipment, backed with a letter of credit issued by Barclays Bank, was also added to the collateral.

PARRs are something as strange as securitization of loan losses. They represent rights to take part of future recoveries on charged-off loans. As an illustration, consider the following example (see [136] pp.244-245):

In January 1986, First City Bancorporation of Texas sold PARRs to Signal Capital Corporation. Signal Capital paid $20 million for the PARRs, which gave Signal Capital the right to the first $20 million + 13% interest rate that First City was able to recover from loans charged-off before 1986.

A.2.4 Commodity-, exchange rate-, and index-linked debt instruments

The use of debt instruments linked to economic variables (e.g., exchange rates, interest rates, commodity prices, indexes) increased significantly during the 1980's (see [34]). Many of these instruments provide efficient tools (not available with ordinary derivative products due to regulatory or other reasons) for corporations and others to handle wide ranges of financial and operating risks. By allowing issuers to increase the expected sta-

SELECTED DEBT INNOVATIONS
Commodity-, Exchange Rate-, and Index-linked Debt Instruments
• All-Ordinaries Share Price Riskless Index Note (ASPRIN)
• Bond Linked to the Value of a Bond of Longer Maturity
• Commodity-Linked Bond
• Dual Currency Bond
• Debt Linked to Trading Volume
• Inflation-Index-Linked Debt Instruments
• Mini-Max
• Principal Exchange Rate Linked Security (PERL)
• Reverse Dual Currency Bond
• Reverse Mini-Max
• Reverse PERL
• Standard & Poor's Indexed Note (SPIN)
• Stock Performance Exchange Linked (SPEL) bond

Table 2.4: Selected Debt Innovations: Commodity-, Exchange Rate-, and Index-linked Debt Instruments.

bility of their cash flows, these instruments may reduce issuers' overall cost of capital (see [34]). A list of the financial products that will be described in this sub-subsection is given in table 2.4.

Exchange rate-linked bonds allow investors and issuers to hedge against, and to speculate in, exchange rate movements. These instruments are especially attractive to market participants that cannot hedge/speculate by other means, for regulatory or other reasons.

Dual-currency bonds, which were introduced in the Euromarket in 1981, are among the most widely used of the exchange rate-linked debt instruments (see [136]). A dual-currency bond, in short, is a bond which has the issue price and interest payable in one currency while the principal is payable in another currency. In contrast to a dual-currency bond, a *reverse dual-currency bond* is a bond with issue price and principal payable in one currency while coupons are payable in another currency.

Instruments that are similar to dual-currency and reverse dual-currency bonds are *Principal Exchange Rate Linked securities (PERLs)* and *reverse PERLs*. A PERL is a U.S. dollar denominated bond. The principal repayment is, however, linked to a specified (foreign currency/U.S. dollars) exchange rate. The amount of principal repayment in U.S. dollars increases (decreases) as the specified exchange rate depreciates (appreciates). Reverse PERLs are identical to PERLs with the exception that the amount of principal repayment in U.S. dollars increases (decreases) as the exchange rate appreciates (depreciates) (see [51]).

The next step in the development of exchange rate-linked debt instruments was the creation of instruments for which changes in the exchange rate have a reduced effect on the bond value. Such an instrument is the *"mini-max"*, which is best illustrated by the following example (taken from [136] pp.286-288):

In April 1986, the Kingdom of Denmark issued a 5-year 10 billion yen bond issue with a coupon of $5\frac{3}{4}\%$. The issue price was 101.25%. The principal payment was to be in yen, and it was to be at par if the (yen/U.S. dollars) rate at maturity was between 90.01 and 263.55. If the exchange rate was lower than 90.01, the principal payable was at discount and calculated according to (spotrate/90.005)$\times$100, and if the exchange rate was higher than 263.55, the principal payable was calculated according to (spotrate/263.55797)$\times$100. The effect from the mini-max structure is to slacken value increases (in U.S. dollars) at maturity due to a stronger yen when the exchange is lower than 90.01, and to slacken value decreases (in U.S. dollars) at maturity due to a weaker yen when the exchange rate is higher than 263.55.

A peculiar exchange rate-linked debt instrument is the *"reverse mini-max"*. For this instrument, the final value (in U.S. dollars) of the bond falls sharply with increases in the exchange rate (yen/U.S. dollars) at maturity, if the exchange rate is lower than a fixed value. When the exchange rate is above this lower value but below another fixed upper value, the final value of the bond increases fast with increases in the exchange rate. At last, the final bond value decreases with increases in the exchange rate when the exchange rate is above the fixed upper value (see [136]).

It is interesting to note that a portfolio with equal amounts of appropriately designed mini-max and reverse mini-max bonds behaves similarly to a dual-currency bond for a large range of exchange rates at maturity (see [136]). This means that the characteristics of the dual-currency bond is divided between the mini-max and reverse mini-max.

"Bull" and *"bear"* bonds are index-linked debt instruments that allow investors to hedge against/speculate in general stock market movements. The first issue of bull and bear bonds was by SEK (Svensk Export Kredit) in June 1986 (see [136]). In short, a bull/bear bond issue is structured as follows:

The issuer issues two tranches, one bull tranche and one bear tranche. The bull bond, not surprising, increases in value with increases in the underlying stock market index. More precisely, the bull bond has a "strike" index level at maturity. At this level, the principal payable at maturity is at par. For every unit the index is above the strike level at maturity, the principal payable increases with a fixed amount, and for every unit the index is below the strike level the principal payable is reduced with the same fixed amount. The bear bond works in the same way as the bull bond with the obvious difference that the principal payable increases with decreases and decreases with increases in the underlying stock market index at maturity. Furthermore, bull and bear bonds are usually constructed with a cap and a floor which limit the extent to which investors can gain or lose. Bull and bear bonds from the same issue also somewhat offset each other. Thus, since the issuer issues both a bull and a bear tranche, the total principal repayment at maturity for the issuer does not change much with different outcomes in the level of the underlying stock index. If properly constructed, the combined effect of the bull tranche and the bear tranche can consequently closely resemble a fixed-rate straight debt issue from the view of the issuer. The reason for using a bull/bear bond issue is the possibility for the issuer to gain cheaper funding than from a comparable straight debt issue. The possible cheaper funding is due to the hedging/speculating tools which the issuer provides by issuing bull and bear bonds.

Bull and bear bonds can be linked to other indexes than stock market indexes. For example, in October 1986, the Kingdom of Denmark raised $120 million through the first bull/bear gold-linked Eurobond issue (see [136]).

A structure similar to the bull bond was the issue of *Stock Performance Exchange Linked (SPEL) bonds* by Guiness Finance in 1986 (see [136]). These bonds had a principal payable at maturity that was linked to the New York Stock Exchange (NYSE) composite index. An investor in a SPEL bond is guaranteed redemption at par. Furthermore, if the NYSE composite index is above a specified level at maturity the investors receive a premium. Moreover, the premium is bigger the more the index exceeds the specified level. If the index ends above the specified level, the principal payable to a SPEL bond holder at maturity is thus par + premium.

Structures with close resemblance to SPEL bonds are *Standard & Poor's Indexed Notes (SPINs)* and *All-ordinaries Share Price Riskless Index Notes (ASPRINs)*. For a SPIN the principal payable at maturity increases in the event of a rise in the S&P 500 index over the life of the bond. In a similar way, the principal payable at maturity on an ASPRIN is higher the higher the All-Ordinaries Index of the Sydney Stock Exchange is at maturity. It can be noted that both SPINs and ASPRINs, like SPEL bonds, gain from increases in the underlying indexes while they are protected against falls. The protection consists of the "floor" feature that the minimum redemption price at maturity is at par.

There have also been issues of *bonds linked to the value a bond of longer maturity.* An example is the issue, by Mitsui and Company in April 1986, of a 3-year bond indexed to the value of a specific U.S. Treasury 30-year bond (see [136] pp.308-310). The redemption price upon maturity for this bond is higher, the higher the price of the U.S. Treasury 30-year bond and vice versa. Furthermore, there have also been issues in which the redemption price upon maturity is lower, the higher the price of the underlying long-term bond and the other way around. One example of this type of bond is the issue by SEK in November 1986. That bond was also called "Ice Bear" (see [136]).

Supposedly, it can be of great interest for many investors to hedge against unanticipated movements in general inflation. For this, among other reasons, *inflation-index-linked debt instruments* have been issued. One example (taken from [136] p.312) of this type of instruments is the £30 million issue by the Nationwide Building Society in July 1986. This is a 35-year loan stock for which interest and capital are adjusted according to inflation every 6 months.

Countries and corporations that are dependent on a particular commodity may, for hedging purposes, want to issue debt linked to this commodity. Thus, it is not surprising that both companies and governments[47] have to a rather large extent issued *commodity-linked bonds* for the financing of their activities. (For an overview of commodity-linked financing see [115].)

Commodity prices have a long history of volatility, so it is not strange that commodity-linked financing is by no means a new phenomenon. The first commodity-linked bond was issued in 1863 by the Confederate States of America (see [115]). This bond issue was denominated in French francs and pounds sterling. The bond was also linked to the price of cotton. The Confederate States of America actually thus issued a dual-currency, cotton-linked bond (see [34]). In the 1920's, commodity-linked debt instruments were also available on the U.S. financial markets. The use of commodity-linked financing in the U.S. was, however, virtually prohibited by new regulations in the 1930's. The legal rights to use commodity-linked debt instruments were not restored until October 1977, with the Helms Amendment (see [34]). After this, their use and issued volume have increased rapidly (see [115]).

Some of the best known commodity-linked bonds issued by governments are the "Giscard" and the "Pinay" (see [136]). Both of these are issued by France (named after the finance ministers at the time of their issues) and linked to the value of gold (actually, the Pinay was linked to the value of the Napoleon gold coin). These bonds are, of course, issued to secure international investors against a weakening French franc, as hedges against inflation or for the purpose of raising funds at a lesser cost than otherwise possible {by

[47]Many developing countries are very sensitive to sharp fluctuations in the prices received for their primary commodity exports. Their terms of trade are also very susceptible to import price shocks, in particular to shocks in the price of petroleum (since petroleum is the most important commodity import for most developing countries). Thus, commodity-linked financing can be of great importance for many of these countries. Hitherto, the use of commodity-linked financial instruments has, however, been most prevalent in the industrialized part of the world (see [115]). But, for the reasons discussed above we can expect to observe more active use of this kind of financing by developing countries in the near future.

taking advantage of arbitrage possibilities in the longer-term side of the gold warrant market (see [115])}, and not due to France's dependence on the value of gold. Silver, oil and copper are among other commodities that government bonds have been linked to.

One of the first corporate issues of a commodity-linked bond was by the largest silver producer in the U.S., Sunshine Mining Company, in April 1980 (see [136]). The basic characteristics of this issue are as follows (see [18] and [115]):

Silver prices fluctuated widely in the late 1970's and early 1980's. As an example, the silver price per ounce was \$6 in 1979. After that it rose to a peak of \$50 in January 1980 only to fall back to \$33 in February 1980. In this volatile environment, Sunshine Mining Company decided to issue bonds that also provided a hedge against variations in its working capital (due to silver price fluctuations). Thus, in 1980 the company raised \$25 million with the issue of 15-year silver-indexed bonds with $8\frac{1}{2}\%$ coupon rate. The bonds make semi-annual coupon payments and each bond pays the "indexed principal amount" (the greater of the face amount of \$1000 or the market value of 50 ounces silver) upon maturity. Furthermore, the bonds can also be redeemed at the option of the issuer on or after April 15, 1985 at the indexed principal amount provided that this amount has exceeded \$2000 for a period of 30 consecutive calendar days (i.e., the silver price per ounce must have been greater than \$40 for 30 consecutive days). In addition to this, the issue is secured by 3.627% of the Sunshine Mining Company's share of the production of the Sunshine Mine in Idaho.

In the issue by the Sunshine Mining Company, it was only the principal payment that was affected by price changes in silver. There has, however, been a rather large number of issues where both the coupon payments and the principal payment or only the coupon payments vary, with price changes in the underlying commodity. An illustration of the latter of these types of issues is given by the following example (taken from [34] pp.83):

In 1988, Magma Copper Company issued *copper interest-indexed senior subordinated notes.* This 10-year debenture has embedded within it no less than 40 options on the price of copper, one for each quarterly coupon payment. This series of options makes the company's coupon payments vary with the price of copper. If the price of copper is high, then the coupon payment will also be high and the other way around. Thus, the design makes the coupon payments high when Magma Copper Company can sell copper for a high price (and therefore should have a good capability to pay) and vice versa[48].

Besides the issues described above there have been many other corporate bonds linked to different commodities (e.g., gold, oil, copper, zinc, natural gas and coal). Furthermore, many of these issues have been very innovative and exotic.

Before concluding this sub-subsection, a peculiar debt instrument issued by the securities brokerage firm Oppenheimer & Co. in 1981 will be mentioned (see [34]). This instrument had its principal repayment indexed to the volume of trading on the New York Stock Exchange. It was thus *debt linked to trading volume.*

A.3 Selected debt/equity hybrid innovations

The stream of innovations in the debt/equity hybrid area has, perhaps, not been as strong as the stream of innovations in the debt area. There have, however, been many new important financial products in the debt/equity hybrid market as well. As a matter of fact, these debt/equity hybrid innovations have, to a large extent, blurred the line of demarcation between debt and equity.

Many new debt/equity hybrid instruments are designed to reduce potential conflicts between bondholders and shareholders (this is, of course, also true for many other new financial products). These potential conflicts are often due to the fact that there are situations where it is possible for a company's management to increase shareholders' wealth at the expense of bondholders. Among the potential conflicts, the following can be mentioned (see [34] p.85):

[48]One interesting oil-linked bond is the issue by Shin Etsu (a Japanese chemical manufacturer) in 1991. This issue has a similar structure to the issue by Magma Copper Company. The difference is that the coupon payments float inversely with the price of oil (see [34]). Since Shin Etsu is a large buyer of oil related products, this design also makes the coupon payments high when the company is likely to have a good capacity to pay (due to low procurement costs) and the other way around.

- *The claims dilution problem* refers to cases where the value of outstanding bonds can be reduced by increasing debt or adding debt senior to the debt in question.
- When a company is close to bankruptcy, the company's management can transfer value from bondholders to shareholders by investing in ever riskier projects in attempts to save the firm. This phenomenon is called *the asset substitution problem.*
- In situations of low revenues and high interest payments, it can be tempting for managers to wait (until "better times", if they come) with value-adding projects and even basic maintenance and safety procedures. This is *the underinvestment problem.*

To protect themselves against problems of the above described types, corporate bondholders require compensation in form of a higher interest rate. Many debt/equity hybrids (and also many other new financial products) reduce the potential conflicts between bondholders and shareholders, and thereby lower the company's overall cost of capital.

As will be shown, another common driving force for new debt/equity hybrids (as well as for other new financial products) is to gain tax advantages.

A.3.1 Preferred stock innovations

A selected number of developments/extensions of the traditional preferred stock are described in this sub-subsection. These preferred stock innovations are listed in table 2.5.

U.S. tax regulations allow corporate investors to deduct 70% of the dividends they receive from unaffiliated corporations from their taxable income. Thus, due to the tax law, U.S. corporate cash managers have tax incentive to invest in preferred stock rather than in short term debt instruments (the interest on short term debt is fully taxable in the U.S.). This means that preferred stock can provide cheaper financing than debt for nontaxable corporate issuers since corporate investors, due to the tax advantage, are willing to accept a lower dividend yield. Traditionally fixed-dividend-rate preferred stock has, however, the drawback that rising interest rates can cause the price of the preferred stock to fall so much that the fall more than offsets the tax saving (see [51]).

SELECTED DEBT/EQUITY HYBRID INNOVATIONS
Preferred stock
• Adjustable Rate Preferred Stock (ARPS) • Auction Rate Preferred Stock • Convertible Adjustable Preferred Stock (CAPS) • Convertible Exchangeable Preferred Stock (CEPS) • Exchangeable Money-Market Preferred • Indexed Floating Rate Preferred Stock • Preferred Equity Redemption Cumulative Stock (PERCS) • Preferred Stock Convertible into Shares of Another Company • Remarketed Preferred Stock • Single Point Adjustable Rate Stock (SPARS) • Variable Cumulative Preferred Stock

Table 2.5: Selected Debt/Equity Hybrid Innovations: Preferred Stock.

A number of developments of the traditional preferred stock to deal with the problem of falling prices due to rising interest rates have been introduced. The first of these developments was the *Adjustable Rate Preferred Stock (ARPS)*. The first ARPS issue was by Chemical Bank in 1982 (see [136]). ARPS are constructed with the purpose of being priced closer to par than traditional preferred stock [and, therefore, reducing the risk of a falling price for the investor (henceforth called the price risk)]. This is done by adjusting the dividend rate quarterly according to a formula which uses market interest rates as input (see [136]). Typically, the dividend rate is reset each quarter, based on the highest of the 3-month Treasury bill rate, the 10-year Treasury bond rate or the 20-year Treasury bond rate plus or minus a specified spread (see [52])[49].

Almost immediately, rather large volumes of ARPS were issued by banks. This was due, in part, to the fact that regulators agreed that ARPS could be considered as primary capital, and banks at the time were under pressure to raise more primary capital. The method of adjusting the dividend rate and the large volume issued by banks led to ARPS sometimes being traded at a large discount. More precisely, at times of concern about

[49]Another similar concept is *indexed floating rate preferred stock*. The dividend rate on an indexed floating rate preferred stock is reset quarterly to a specified percentage of 3-month LIBOR (see [52]).

the safety of the banking system, the spread investors have required for pricing the ARPS at par has differed significantly from the spread calculated with the adjusting formula. Thus, now and then, ARPS have been priced significantly lower than their face values (see [136]). This means that ARPS reduced but did not eliminate the price risk.

An instrument constructed with the purpose of resolving the above described problem for ARPS is *Convertible Adjustable Preferred Stock (CAPS)*. In short, a CAPS can be described as an adjustable rate preferred stock with a conversion right added to it. The holder of a CAPS has the right to convert into common stock of the issuing company during the period after the announcement of each dividend rate for the next period. As a result, and also desired, CAPS have traded closer to par than ARPS. In spite of this progress, there have only been a few CAPS issues. One reason for this can be that issuers have disliked the idea of the possibility of being forced to issue common stock (see [51]).

An extension of the CAPS concept is the *Convertible Exchangeable Preferred Stock (CEPS)*. The first CEPS issue was by Martin Marietta in September 1982 (see [136]). As the name suggests, a CEPS has two main features. Firstly, it is convertible at the option of the CEPS holder into common stock of the issuer. Secondly, a CEPS is exchangeable at the option of the issuer for debt convertible into common stock in the issuing company. One reason for issuing CEPS is to gain tax advantages. They are especially attractive to issue for a company that is not at present in a tax-paying position, but expects to start paying tax during the life of the security. CEPS allow the issuer to convert into convertible debt, without having to pay additional underwriting commisions, and thus give the possibility of taking advantage of the interest tax shield when the company enters a tax-paying position (see [52]). CEPS become even more interesting if the company also wants to issue convertible debt, but for rating purposes is instead forced to issue equity (see [136]).

As mentioned earlier, CAPS never became a success. In a new attempt to eliminate the problem of ARPS (i.e., the price risk) *auction rate preferred stocks* were introduced in 1984 (see [136]). For auction rate preferred stocks, the dividend rate is reset by Dutch auction every 49 days. The reason for the 49 days interval between dividend rate-setting days is that 7 weeks is exactly enough whole weeks to fulfil the 46-day holding period required to claim the 70% dividend tax exclusion (see [51]). If the auction fails, the dividend rate is determined as a percentage of the composite commercial paper rate (see [136]).

At the time of the introduction of auction rate preferred stocks, there were beliefs that the concept was too complex to become popular. Auction rate preferred stocks, however, quickly achieved a dominating position on the U.S. preferred stock market. Moreover, soon after the introduction auction rate preferred stocks were sold under several different names, e.g., *Cumulative Auction Market Preferred stock (CAMP), Dutch Auction Rate Term Securities (DARTS), Market Auction rate PreferredS (MAPS), Money Market Preferreds (MMP)* and *Short Term Auction Rate preferredS (STARS)*. The many names are due to the fact that many securities firms have constructed and introduced their own products. Although there are many different names, the securities are more or less the same.

An extension of the auction rate preferred stock is the *exchangeable money market preferred*. This instrument is simply an auction rate preferred stock exchangeable into money market notes at the option of the issuer. Exchangeable Money Market Preferred was introduced in a issue for the U.S. insurance company AIG in 1985 (see [136]). One reason for the AIG issue was that the company needed capital for the financing of a fast growing business. AIG was not in a tax-paying position at the time of the issue, but expected to be a taxpayer in the future and thus wanted to have the option to convert the security into debt (see [136]).

Another attempt to improve the ARPS structure is the *Single Point Adjustable Rate Stock (SPARS)*. The dividend rate on a SPARS is adjusted automatically every 49 days to a specified percentage of the 60-day high-grade commercial paper rate. The aim of this dividend rate-setting technique is to achieve the same degree of liquidity as auction rate preferred stock, but with lower transaction costs since no auction is used. SPARS have, however, one problem in common with ARPS. The problem is that the fixed dividend rate-setting formula leads to a price risk for investors. If the issuer's credit worthiness falls, SPARS can trade significantly below par (see [51]). There have only been two SPARS issues, in part due to this problem (see [52]).

An interesting preferred stock variant is *remarketed preferred stock.* This structure is constructed with the aim to avoid the Dutch auction used for auction rate preferred stock, and still have a dividend rate-setting method that eliminates most of the price risk for investors. The method used for adjusting the dividend rate is to have a specified remarketing agent that at the end of each dividend period determines the rate that will cause the security to be priced at par.

As mentioned several times previously, ARPS have certain problems. Among the security types that are constructed with the purpose of eliminating these problems, are remarketed preferred stock and auction rate preferred stock. There is, however, no general agreement about which of the two that works best (see [52]). As a compromise, *variable cumulative preferred stock* was constructed. In principle, variable cumulative preferred stocks give the issuer the right at the beginning of each dividend period to decide whether to use Dutch auction or a remarketing agent for adjusting the dividend rate.

A rather recent development in the preferred stock area (the first issue was in August 1991), is *Preferred Equity Redemption Cumulative Stock (PERCS)* (or *mandatory conversion premium dividend preferred stock*) (see [52]). Briefly, this security is a preferred stock that has a dividend rate significantly higher than the dividend rate on the underlying common stock. In addition, a PERCS has an option to convert into common stock. The conversion option has, however, a capped share value. That is, (compared to ordinary common stock) an investor in PERCS gains a higher than normal dividend rate at the expense of a portion of the appreciation potential of the underlying common stock (see [52]).

Finally, one very interesting and exotic type of preferred stock is *preferred stock convertible into shares of another company.* This kind of security can be used by companies that have a large holding of equity of another company, which can back the issue. An example of an issue of preferred stock convertible into shares of another company is the issue by News Corporation in 1986 (see [136] p.326). In this issue, News Corporation offered preference shares convertible into ordinary shares of Reuters Holdings.

SELECTED DEBT/EQUITY HYBRID INNOVATIONS
Convertible debt
• ABC Security • Adjustable Rate Convertible Debt • Bond Convertible into Shares of Another Company • Cash Redeemable LYON • Commodity-Linked Convertible Bond • Convertible Reset Debenture • Cross-Currency Convertible • Equity Commitment Note • Equity Contract Note • Mandatory Convertible • Puttable Convertible Bond

Table 2.6: Selected Debt/Equity Hybrid Innovations: Convertible debt.

A.3.2 Convertible debt innovations

In this sub-subsection, some innovations in the area of debt convertible into common equity will be described briefly. These convertible debt innovations are listed in table 2.6. It should also be noted that a number of convertible debt and (closely related) convertible preferred stock innovations are described in earlier sub-subsections of this appendix. These are *Liquid Yield Option Notes (LYONs)* (in sub-subsection A.2.2), *Convertible Adjustable Preferred Stock (CAPS)* (in sub-subsection A.3.1), *Convertible Exchangeable Preferred Stock (CEPS)* (in sub-subsection A.3.1), *Preferred Equity Redemption Cumulative Stock (PERCS)* (in sub-subsection A.3.1) and *preferred stock convertible into shares of another company* (in sub-subsection A.3.1).

Convertible reset debentures, which were introduced in October 1983 (see [52]), provide investors with protection against deterioration in the issuer's credit quality or financial prospects within 2 years of issuance. The protection is accomplished by the fact that the coupon rate on the bonds must be adjusted upward, if necessary, 2 years after issuance by an amount sufficient enough to give the debentures a market value equal to their face values (see [51]). One reason for a deterioration of the issuer's credit standing can be through managers' attempts to increase stockholders' value at the expense of bondholders.

Thus, the construction of convertible reset debenture is likely to reduce agency costs (see [52]).

Another instrument type that is likely to reduce agency costs is *puttable convertible bond.* In short, a puttable convertible bond gives the holder the right to sell the bond back to the issuer prior to maturity, at specified dates and prices (see [52]).

One result of banks' strong demand for raising capital in 1982-1984 was the development of a category of instruments that may be called *mandatory convertibles* (see [136]). In contrast to ordinary convertible bonds (which give the holders the option to convert into equity) holders of mandatory convertibles are, one way or another, committed to convert into equity. After the introduction of mandatory convertibles, they quickly became frequently used and in 1984 mandatory convertible issues contributed to 40% of the capital raised by the top 50 U.S. bank-holding companies (see [136]).

The first issue of a security in the category of mandatory convertibles was by Manufacturers Hanover in April 1982 (see [136] p.323). This issue consisted of 10-year fixed-rate notes with detachable stock purchase contracts attached to them. The holders of the stock purchase contracts are obliged to buy shares of common stock up to the principal amount of the notes, at maturity of the fixed-rate notes. The price per share was set to the lower of $55.55 or the market price, but the price per share was never to be set at a price lower than $40. Thus, the holders of the stock purhase contracts are exposed to a risk from stock price movements.

In 1984, the mandatory convertible concept was developed by the introduction of the *equity contract note* (see [136]). The holder of a equity contract note is not exposed to any risk from movements in the issuing company's stock price. More precisely, an equity contract note must, at maturity, be exchanged for primary equity securities that have a market value equal to the principal amount of the note[50].

[50]Actually, the holder of a equity contract note can choose not to exchange them for equity at maturity. If the holder chooses not to exchange, the issuer will try to sell the equity on behalf of the holder. If the issuer, however, for some reason cannot find a buyer, the equity contract note holder is obliged to accept the equities.

Another security type belonging to the family of mandatory convertibles is the *equity commitment note*. The first issue of equity commitment notes was by J.P. Morgan in 1982 (see [136]). The basic feature of equity commitment notes is that the issuer is obliged to redeem the notes with the proceeds from an equity issue at some day in the future.

A fascinating financial instrument is *bond convertible into shares of another company*. This type of convertible bond has been used by several companies which have a holding in another company that is large enough to back an issue. An issue of bonds convertible into shares of another company often gives the issuing company a favourable financing (partly due to tax advantages) (see [136]).

Many issues of bonds convertible into shares of another company have been issued by companies which were in possession of a large block of stock after an unsuccessful takeover bid. An example of this is General Cinema Corporation's (GCC) issue of bonds convertible into shares of R.J. Reynolds (see [136] pp.325-326). The background to this issue is that GCC made a bid for Heublein. GCC was, however, outbid by R.J. Reynolds and GCC's 4 million shares in Heublein were automatically converted into a large holding of shares in R.J. Reynolds. Next, backed by this holding, GCC issued bonds convertible into shares of R.J. Reynolds.

LYONs, which are more thoroughly described in sub-subsection A.2.2, are basically zero coupon bonds convertible into common equity. A small extension of the LYON concept is the *cash redeemable LYON*. Briefly, this security is a LYON which is redeemable in cash for the market value of the underlying common stock, at the option of the issuer. Precisely as ordinary convertible bonds, if the issue converts, the issuer in effect will have sold tax deductible common equity. The option to redeem in cash, however, means that the issuer does not have to have its equity ownership interest diluted through conversion (see [52]).

Like LYONs (see sub-subsection A.2.2), *ABC securities* are a form of zero coupon bonds convertible into common equity. The non-interest-bearing nature of both of these instruments leads to the entire debt-service stream being converted into common equity, if holders convert. ABC securities are distinguished from LYONs by an interesting feature.

This feature is: If the price of the underlying common stock rises by more than a specified percentage (normally around 30%) from the date of issuance, then the dividends on the underlying common stock will be passed through to the holders of the ABC securities (see [52]).

A short-lived but amusing variation on the convertible debt theme was *adjustable rate convertible debt*. This security type was convertible debt on which the interest varied directly with the dividend rate on the underlying common stock. Adjustable rate convertible debt was clearly an attempt to issue equity disguised as debt for the purpose of benefiting from interest tax deductions. This attempt was, however, so obvious that the Internal Revenue Service (IRS) ruled that the security was equity for tax purposes after only three issues. After the ruling from IRS there have not been any issues of adjustable rate convertible debt (see [51]).

A *cross-currency convertible* is a convertible for which the bond is denominated in a currency different from the currency of the country where the underlying stock is issued. These types of convertibles are becoming increasingly popular and rather large volumes have been issued (see [73]). An example of a cross-currency convertible is the issue by the Swedish company SCA in 1988 (see [73] pp.94-97). This issue is denominated in ECU and the stock of SCA is denominated in Swedish crowns (the stock is listed on the Stockholm Stock Exchange). The complexity of the SCA convertible is further enhanced by the facts that it can be called by the issuer (actually, the issuer has two different kinds of option to call) and that it can be put by the holders.

Finally, it will only be mentioned that there have been issues of *commodity-linked convertible bonds*. An example is the issue for Semirara Coal Corporation in 1980 (a company from the Phillipine islands). After four years this issue could be converted into common shares or the market value of 30 tons of coal per 10,000-peso of principal (see [28] p.45).

SELECTED COMMON EQUITY INNOVATIONS
• Americus Trust
• Callable Equity
• Contingent Value Right (CVR)
• Dual-Track Stock
• Euroequity
• Multicurrency Equity
• Limited Partnership Units
• Puttable Equity
• Shares Linked to the Performance of a Subsidiary
• SuperShares
• Unbundled Stock Units (USU)

Table 2.7: Selected Common Equity Innovations.

A.4 Selected common equity innovations

During the last decade a rather large number of developments of ordinary common stock have been introduced on the financial markets. A selection of these developments, given in table 2.7, will be described briefly in this subsection.

The ideal stock would, of course, be a stock that goes up but never comes down. One of the first instruments that had some of this property was *puttable common stock.* The puttable common stock concept was constructed by financial engineers at Drexel Burnham Lambert for Arley Merchandise Corporation, and this issue came to the market in November 1984 (see [33]). Puttable common stock is exactly what the name suggests. Consequently, an issue of puttable common stock is an issue of ordinary common stock along with the right to sell the stock back to the issuer at a specified date to a specified price. Actually, for the purpose of ensuring a minimum holding period rate of return many puttable common stock issues have a schedule of increasing put prices (see [52]).

One driving force behind the development of puttable common stock was problems with initial public offerings (IPOs) (see [136]). Due to information asymmetry[51] invest-

[51]The information asymmetry prevails since managers often have information of new potentially profitable projects that cannot easily be communicated to the market. One hurdle for communication is, of course, that managers do not want to inform their company's competitors about these new promising

ment bankers are often forced to underprice IPOs. In such cases it can be valuable to issue puttable equity instead. This follows from the fact that the put feature reduces the information asymmetry[52], and thus also reduces investor uncertainty.

The "money-back guarantee" provided by puttable common stock is, of course, only as good as the issuing company's capability to fulfil it. Thus, an obvious improvement is to enhance the protection for investors with a third party guarantee. An instrument with this construction was introduced with *limited partnership units* in the Dean Witter Principal Guaranteed Fund L.P.. These instruments contained a money-back guarantee, in the form of put options, at the end of five years. The options were issued by a special-purpose "Guarantee Corporation" backed (for a fee) by an irrevocable letter of credit issued by Citibank, N.A. (see [33]).

A construction with a money-back guarantee by a reputable third party has advantages compared to puttable common stock (or convertible debt). Firstly, it gives, of course, the investors better protection. Secondly, the backing by the reputable third party is also likely to more effectively resolve the information problem of asymmetry. The third party requires, of course, a fee for providing the money-back guarantee. A company like Citibank, however, has economies of scale and scope in evaluating credit risks. Thus, if competitively priced, this kind of money-back guarantee can resolve a large part of the information asymmetry problem in the IPO market (see [33]).

Puttable common stock and limited partnership units limit the downside risk to investors. Investors would, of course, be even more happy if they could also be guaranteed that the stock would appreciate. *Contingent Value Rights (CVRs)* were constructed with the aim of somewhat fulfilling this objective (see [33]). The CVR concept can be illus-

projects in order to keep the competitive advantage. Another hurdle for communication is that some types of information are hard to communicate with credibility. The asymmetry problem is especially acute in the case of IPOs, since there is no established trading price or public track record of past earnings (see [33]).

[52]To be more precise, by bearing added risk (from the protection of new shareholders) existing shareholders signal their confidence about the prospects of the firm. This reduces the information asymmetry (see [33]).

trated with the first issue of the invention (see [33] pp.39-40).

This issue was for Dow Chemical Company (Dow) when it merged its wholly owned subsidiary Merrell Dow with Marion Laboratories in September 1989. The result of the merger was Marion Merrell Dow, Inc. (MMD). This transaction started with the fact that Dow bought 38.9% of Marion Laboratories at $38 per share in a tender offer. Next, Dow wanted to buy the remaining shares of Marion Laboratories with shares in the MMD. The problem was that Marion Laboratories stockholders did not agree that new MMD shares were worth the $38 each that Dow was assigning them. To resolve this problem Dow offered a CVR along with each new MMD share. Thus, for each remaining share of Marion Laboratories, the holder would receive a unit consisting of one share of the new MMD and one CVR. Now, to the design of the CVRs. Two years after the issue (i.e., September 30, 1991), each CVR gave the holder the right to a cash payment equal to the amount by which $45.77 exceeded the greater of $30 or the average price of the MMD stock for the 90 calendar days preceding September 18, 1991, and if the average MMD stock price exceeded $45.77, the CVRs would be worthless. That is, if nothing dramatic happened, holders of a unit (i.e., a MMD share and a CVR) would be guaranteed a value of at least $45.77 at expiration. Since the cash payment to a CVR holder is capped by $15.77, the CVR only provided a limited protection against losses by the investors. (As things turned out, Dow had to pay holders $11 per CVR.)

Unfortunately, there still remains a potential agency problem when using CVRs. This problem is: If the parent company, Dow in the example above, retains the possibility of manipulating the price of the new stock (for example through dividend policy or transfer pricing policy), it will be in a position to transfer wealth from the new company to its own shareholders (see [33]). One recent design that was developed in part to resolve this agency problem is *dual-track stock*. This financial product is best described by an example (see [33] p.41).

The first issue of dual-track stock was for Aramed Inc. and Gensia Pharmaceuticals Inc., in October 1991. Aramed was formed one month before the issue. The purpose in forming

Aramed was to use it to speed up the development of some of Gensia's promising new products. Aramed's whole substance derives from rights under a series of agreements with Gensia. The proceeds from the issue of Aramed stock units were to be used solely for the funding of Gensia's research in three related development programs. In return for this Aramed will own patent rights on the resulting products. Evidently, these relationships have the potential to create severe conflicts of interest. With the purpose of eliminating some of the potential for such conflicts of interest, the Aramed stock was offered in units. Each unit consisted of one share of callable common stock (see below) of Aramed, and one warrant to purchase a share of Gensia common stock. Thus, the potential for conflicts of interest was reduced by giving each part in the deal the possibility to take part in the other part's upside potential. Gensia has the option (through the callable common stock of Aramed) to reacquire all of Aramed's rights. The new stockholders of Aramed can take part of Gensia's upside potential by exercising their warrants.

The rather advanced package in the above example allows Gensia to spin-off and fund three research programs while still retaining an option to reacquire them. Thus, dual-track stock provides a new device for companies to spin-off entities like divisions, individual ventures, products or projects (see [33]).

Callable common stock, not surprisingly, is common stock with a stock repurchase option agreement added to it. The stock repurchase option agreement is often constructed with an exercise price that increases step-by-step over time. Furthermore, the agreement also often requires that all outstanding repurchase options should be exercised, if any are exercised. The construction of a stock repurchase option agreement implies that investors forego capital appreciation in excess of the exercise price.

Many issues of callable common stock are by subsidiary companys, but sold by the parent company. This allows the parent company to reacquire the subsidiary's shares at prespecified terms in the future, and thus regain a 100% capital and voting control. Many of these later issues also include warrants to purchase common stock in the parent company. Adding this type of warrant partly offsets the drawback of foregoing capital appreciation in excess of the exercise price for the investors. This is due to the fact that warrants to purchase shares of the parent company allow investors to gain from capital appreciation in the subsidiary through its impact on the value of the parent company's shares.

Exciting types of equity are *shares linked to the performance of a subsidiary.* This form of equity is best illustrated with an example (see [136] pp.326-327).

When General Motors (GM) bought Electronic Data Systems Corporation (EDS) in 1984, GM issued an instrument called class E shares. Class E shares represent ownership in GM, but have dividends (and thus a price) that are linked to the performance of EDS. This construction was intended to give holders of class E shares the security of owning GM, while still having the possibility to (directly) gain from good performance of EDS. Class E shares became a success, so in the later acquisition of Hughes Aircraft Corporation, GM issued a similar construction called class H shares.

An interesting feature of shares linked to the performance of a subsidiary is that they enable the marketplace to establish a separate market value for the subsidiary, while still allowing the issuing parent company to retain a 100% voting control. The instrument type can also be useful for an employee stock option plan or other incentive compensation schemes for employees in the subsidiary (see [51]).

A development that was introduced in the U.K. market in the 1980's is *multicurrency equity.* This type of equity is denominated in other currencies than the local currency of the country in which the issuing company is incorporated (see [136])[53]. A multicurrency equity issue is illustrated with an example taken from [136] pp.327-328.

In the late 1980's, the U.K. authorities gave the Scandinavian Bank permission to switch its equity capital from sterling into a basket of no less than four currencies. The new equity capital structure of the bank consisted of 50% U.S. dollars, 20% sterling, 15% deutsche mark and 15% Swiss francs. To accomplish this, four classes of shares were introduced, one for each currency. The reason for the Scandinavian Bank to invent this

[53]Before the introduction of multicurrency equity in the U.K., however, Singapore and Luxembourg permitted equity capital of local subsidiaries to be denominated in ECU (see [136]).

construction was that its balance sheet, to a large extent, consisted of non-sterling loans. Thus, with all equity capital in sterling shareholders had to put up new capital when sterling fell, only to maintain existing level of business.

Another rather recent development in the area of international equity is *Euroequity.* This security type does not represent a new kind of equity, the innovation is the structure of the market where the security type is traded. More precisely, the well-established and well-developed Euromarkets (on which Eurobonds have been traded for more than two decades) also provided a marketplace for equity (see [136]). The Euroequity market rapidly obtained depth and breadth with large volumes of Euroequity issues. The growth of the Euroequity market can, in part, be explained by two advantages of Euroequity (see [93]). Firstly, there are low issuing costs. Before the introduction of Euroequity, a company had to pay listing costs in every country it wanted to reach for raising equity capital. In contrast, the use of the Euromarkets allows Euroequity to be distributed through the large and active over-the-counter market that developed for Eurobonds in the 1970's and early 1980's. Secondly, there is good liquidity. Euroequity removes the registration barrier by providing a bearer instrument that can also be resold to U.S. investors. Thus, liquidity is enhanced both by allowing for anonymity that many European investors seek, and by allowing U.S. investors to trade in the instruments.

In October 1983, the first *Americus Trust* was introduced. It was offered to owners of AT&T shares. After the first Americus Trust more than two dozen other Americus Trusts have been established. A recent change in the U.S. tax law, however, made Americus Trusts unfavourable. Since this change in the tax law, no new Americus Trusts have been formed (see [52]).

Basically, an Americus Trust is a five-year unit trust to which outstanding shares of a particular company's common stock are contributed. The trust separates each contributed share into a PRIME component and a SCORE component. The PRIME component carries full dividend and voting rights, but it has only limited capital appreciation rights. The SCORE component carries the residual capital appreciation stream. More precisely, the SCORE component carries full capital appreciation rights above a stated price.

A construction similar, but more advanced, to Americus Trusts are *SuperShares.* SuperShares are created by financial engineers at Leland, O'Brien, Rubinstein Associates, and after more than three years of refining, SuperShares came to the market in the late spring 1992 (see [110]). SuperShares promise to provide mechanisms for institutional investors to protect their portfolios against general declines in the stockmarket. SuperShares also promise to provide other investors with a low-cost way to own all the stocks in the Standard & Poor's (S&P) 500.

There are four types of SuperShares. Two of them, Priority SuperShares and Appreciation SuperShares, are intended to divide the stream of total returns on a S&P 500 portfolio of common stocks into two components (over the next three years). Priority SuperShares receive all dividends plus the first 25 % of appreciation in the portfolio, and Appreciation SuperShares receive whatever is left over, if anything. The remaining two types of SuperShares, Protection SuperShares and Income and Residual SuperShares, are intended to divide the stream of returns from a portfolio of government securities (over the next three years) in a very interesting way. The value of Protection SuperShares will not be based on the performance of the portfolio of government securities, but on the performance of the S&P 500. If the S&P 500 goes up or remains level over the three years, the Protection SuperShares will expire worthless. If the S&P 500 declines the Protection SuperShares expires in the money. The value at expiration for a Protection SuperShare increases with $1 for every percentage decline in the S&P 500. If it goes down 30 percent or more, however, the Protection SuperShare will be worth $30. Thus, Protection SuperShares provide portfolio insurance. The Income and Residual SuperShares will receive all the income and whatever is left of the principal after the Protection SuperShares are paid off (see [52] and [110]).

Another concept which, like Americus Trusts and SuperShares, was aimed to give shareholders more freedom to choose among different components of the total return of common stock was *Unbundled Stock Units (USUs).* Despite aggressive marketing, however, the concept failed, and it was withdrawn from the marketplace before a single issue could be completed (see [52]).

The purpose of USUs was to divide the total return from common stock into no less than three components. Firstly, a 30-year base yield bond paying an interest rate equal to the dividend rate on the underlying common stock at the time the trust was formed. In addition to this, the first component had a limited capital appreciation right. Secondly, an 30-year instrument similar to a preferred stock. The dividend rate on this instrument was equal to the excess, if any, of the current dividend rate on common stock above the interest rate on the base yield bond. Thirdly, a 30-year warrant on the capital appreciation above the base yield bond's redemption value.

Chapter 3

Paper B: How "errors" in boundary conditions affect solutions when the implicit finite difference method is used

3.1 Introduction

Valuation by means of contingent claims analysis (CCA) is becoming increasingly popular. CCA is wide-ranging in that it can be applied to a very broad spectrum of problems. An excellent overview of different applications of CCA is given in [94].

CCA is a technique for determining the price of a security whose payoffs depend upon the evolution of one or more underlying state variables[1], i.e., a contingent claim. The origins of CCA are found in the theory of option pricing, and the major breakthrough came in a paper by *Black and Scholes* (see [12]).

When CCA is used, the value of the contingent claim is often obtained through the solution of a partial differential equation. The partial differential equation looks almost the same for every contingent claim. The features that distinguish one contingent claim

[1] This paper is, however, only concerned with the case of one underlying state variable.

from another are incorporated into the initial and boundary conditions[2] which are imposed on the partial differential equation.

It is impossible to find a closed-form solution to the partial differential equation subject to the relevant side conditions (both initial and boundary conditions are referred to as side conditions), for most contingent claims. Therefore, the value of the contingent claim has often to be found by means of numerical methods. One of the most commonly used of these methods is the implicit finite difference method. This method is very flexible and can be used for pricing contingent claims with very different characteristics.

The side conditions are very important in valuation by means of CCA, since they distinguish one contingent claim from another. However, when it is said that the side conditions distinguish between different types of contingent claims, fundamental differences in the conditions are referred to. A very different question is how slight differences in the modelling of the side conditions affect the value of the same contingent claim.

How to model the side conditions to the partial differential equation for a given contingent claim is often not self-evident. In some cases, there is more than one reasonable alternative, while in other cases no satisfactory alternative can be found. The difficulty in modelling side conditions has been mentioned by many researchers in many articles, see for example [19], [21], [25], [71], [73], [74] and [95]. Most of these articles have discussed how to model correctly bankruptcy conditions or owners' and issuers' decision rules.

The initial condition is often the easiest of the side conditions to model. This follows from the fact that the initial condition is usually well specified in the prospectus or is self-evident from the construction of the contingent claim. Modelling the boundary conditions

[2]The initial condition gives the contingent claim value at maturity. (The name initial condition refers to the fact that finite difference schemes work backwards in time. They start at maturity and work towards current time.) The boundary conditions give conditions on the price function of the contingent claim at the boundaries of the domain of the price function of the contingent claim in the dimension of the underlying state variable. In other words, the price function is defined for a given set of values of the underlying state variable (at a given point in time). The boundary conditions give conditions on the price function of the contingent claim at the boundaries of this set (at each point in time). This will be clarified later in this article.

is often much more troublesome[3]. Thus, "errors"[4] in the boundary conditions are not unusual. The research task of this paper is to investigate how errors in the boundary conditions affect the solution, when the implicit finite difference method is used[5].

To the knowledge of the author, there has not been any work regarding the research task of the paper. Furthermore, the research task has an academic interest since it has an unclear status in the academic literature.

Some researchers seem to have the apprehension that modelling errors in the boundary conditions of the magnitude likely to occur do not have any noticeable effects on the

---

[3]This is partly due to the fact that many pricing problems within CCA are pure initial value problems. However, in these cases, when some numerical methods (such as the implicit finite difference method) are used approximate boundary conditions have to be imposed on the partial differential equation.

[4]"Errors" is used, since there usually is not a single true way to model the boundary conditions. Instead, there may be several more or less good possibilities to model the boundary conditions for the same contingent claim. Thus, one interpretation of error is as the deviation between two more or less reasonable modelling decisions. So, the research task of the paper also tries to answer the question "How do different ways of modelling the boundary conditions, for the same contingent claim, affect the solution, when using the implicit finite difference method?".

[5]If the aforementioned discussion is unclear things may be clarified by *Ackoff's* (see [1] pp.139-140) classification of model errors. According to Ackoff, the ways in which a model can be in error are:

1. The model may contain variables which are not relevant; that is, have no effect on the outcome. Their inclusion in the model, then, makes the outcome depend on factors on which it has no dependence in reality.

2. The model may not include variables which are relevant; that is ones that do affect the outcome.

3. The function/functions that relate the controllable and uncontrollable variables to the outcome may be incorrect.

4. The numerical values assigned to the variables may be inaccurate.

How errors in (or different ways of modelling) the boundary conditions affect the solution (the outcome in Ackoff's taxonomy) is going to be investigated in this paper. Furthermore, the function that relates the parameters to the outcome is obtained as the solution to the partial differential equation subject to side conditions. Thus, the appearance of the function is affected by the way the boundary conditions are modelled. Hence, how errors (deliberate or not) of type 3 in Ackoff's classification affect the solution is (implicitly) investigated in the paper.

results. This apprehension manifests itself by the fact that the boundary conditions are not specified or are incompletely specified in many of the articles in which the implicit finite difference method is used (a survey of "how boundary conditions are treated in published financial articles in which the implicit finite difference method is used" is given in section 3.2). Consequently, it is clear that in spite of the knowledge of the modelling difficulties there is almost no discussion about whether errors in (different ways of modelling) the boundary conditions have any impact on contingent claim values, when the implicit finite difference method is used. This lack of discussion is possibly due to many researchers having independently reached the conclusion that exactly how the boundary conditions are modelled is usually not important. As an example, an oral discussion with Peter Jennergren at the Stockholm School of Economics gave the following information:

Jennergren has for several years given a basic course in numerical methods applied to option pricing problems. When examining solutions to problems handed in by students, Jennergren noticed that even large errors, especially in the lower boundary condition, seldom influence the results noticeably.

On the other hand, some researchers seem to have the apprehension that different ways of modelling the boundary conditions have major effects when the implicit finite difference method is used. This apprehension manifests itself through the arguments for methods where the modelling of boundary conditions can be avoided[6], and by the arguments against the implicit finite difference method (see e.g., [69])[7].

---

[6]The problem of how to model the boundary conditions can be avoided by some approximating techniques, e.g., binomial models and explicit finite difference methods.

[7]In a petition for the *explicit* finite difference method, *Hull and White* (see [69]) conclude that one disadvantage of the implicit finite difference method is that it requires specification of the boundary conditions. Since, many contingent claim problems naturally are initial value problems, use of boundary conditions will inevitably, in these cases, introduce errors. Hull and White formulate this fact in the following way: "One reason for the extra efficiency of the explicit method is that most derivative security pricing problems are initial value problems, *not* boundary value problems. Errors are introduced by the redundant boundary conditions in implicit methods." (see [69] p.99).

The purpose of this paper is to try to shed some light on the unclear question just discussed.

How errors in the boundary conditions affect the solutions is, however, not only of academic interest. It also has clear practical implications. If valuation by means of CCA and the implicit finite difference method is to be used for commercial purposes, it is important to know how robust the method is with respect to errors in the boundary conditions.

Before going into the research task of this paper, it can be helpful, as a background, to give a description of CCA and the finite difference methods applied to pricing problems within CCA. Such a background is given in the first part of this paper. The background aims to make it easier to follow the subsequent investigation, and it can also be used as a source of reference (in particular to finite difference methods applied to pricing problems within CCA). The background gives a fairly thorough and unified description of:

- The assumptions underlying contingent claims analysis.
- The derivation of the fundamental partial differential equation, i.e., the partial differential equation that the price of the contingent claim must satisfy.
- The numerical solution of the fundamental partial differential equation with finite difference methods, mainly the implicit finite difference method. A number of aspects of valuation with the use of the implicit finite difference method are treated in a novel fashion (at least, to this author's knowledge).

In this paragraph, the organization of the remainder of the paper is described. The content of section 3.2 is an overview of how boundary conditions are usually specified in published financial articles where the implicit finite difference method is used. In section 3.3, the fundamental partial differential equation is derived. Section 3.4 gives a

When Hull and White claim that this approximation of an initial value problem with a mixed initial and boundary value problem is a disadvantage for the implicit finite difference method, they implicitly claim that errors (of the size likely to occur with proper modelling) in the boundary values cause errors in the results of a magnitude that cannot be neglected.

description of finite difference methods, and in particular the implicit finite difference method. The main part of the investigation of how errors in the boundary conditions affect the solutions has been done through studies of specific cases. Nevertheless, a simple general analysis has been made. This general analysis has been done more to structure and give an understanding of what is involved than to represent a complete analysis of the problem. The general analysis is given in section 3.5. The empirical studies of specific cases are given in section 3.6, and the conclusions from this investigation are given in section 3.7.

3.2 How boundary conditions are usually treated, when the implicit finite difference method is used

A fairly concise overview of how the boundary conditions have been treated in published financial articles where the implicit finite difference method has been used is given in this section. The sample on which the overview is based consists of 19 articles[8]. Of those articles, 18 were published in well-established journals and the last article is from an article collection in the form of a book.

The articles in the sample were published during the time period 1977 to 1990. Furthermore, the journal articles are from a rather wide range of journals. More precisely, 5 articles are published in *The Journal of Finance* and *Journal of Financial and Quantitative Analysis* respectively, 2 articles are published in *The Journal of Financial Economics* and 1 article is published in each of *Journal of Business*, *The Accounting Review*, *Journal of International Financial Management and Accounting*, *The Journal of Risk and Insurance*, *Omega* and *BETA*.

The implicit finite difference method is used to value a fairly large number of different

[8]More precisely, the articles in the sample are [3], [17], [19], [20], [21], [22], [23], [25], [27], [31], [38], [39], [54], [72], [73], [74], [86], [95], and [122].

CONTINGENT CLAIM TYPES IN THE SAMPLE	# ARTICLES
• European call option on stock	3
• European put option on stock	2
• American call option on stock	2
• American put option on stock	3
• Warrant	1
• American call option on futures contract	1
• American put option on futures contract	1
• European call option on default-free bond	1
• European put option on default-free bond	1
• American call option on default-free bond	1
• American put option on default-free bond	1
• Savings bond	1
• Retractable bond	2
• Extendable bond	1
• Callable bond	1
• Convertible bond with firm value as state variable	4
• Convertible bond with stock value as state variable	2
• Liquid Yield Option Note (LYON)	1
• Equity-linked insurance contract	1
• Leveraged firm	1
• Division of firm value into values of common stock, preferred stock and bonds	1

Table 3.1: Contingent claim types valued with the help of the implicit finite difference method, in articles belonging to the sample.

types of contingent claims in the sample of articles. In table 3.1, all of these types of contingent claims are listed. The number of articles in the sample that each contingent claim type appears in is also given in table 3.1. (As table 3.1 shows, many sample articles value more than one contingent claim type.)

Now to the question: "How are the boundary conditions treated in published financial articles in which the implicit finite difference method is used?". The answer is that there is a disparity in the way the boundary conditions are described. The disparity is confirmed by the sample of articles used in the overview in this section.

In most of the sample articles (13 articles) the boundary conditions are specified mathematically. Of these 13 articles, a majority (10 articles) also have a short verbal explanation of the boundary conditions. One sample article only has a verbal specification of the boundary conditions. Finally, in as many as 5 articles, the boundary conditions are either incompletely specified or not specified at all.

What types of boundary conditions, then, are used in the sample articles (where specified)? At all but 3 boundaries, either a condition directly on the contingent claim value or a condition on the first derivative of the contingent claim value with respect to the underlying state variable is used. Furthermore, a condition directly on the contingent claim value is used at equally many boundaries as a condition on the first derivative of the contingent claim value with respect to the underlying state variable. By far, the most common combination is to use a condition directly on the contingent claim value at the lower boundary (the boundary with the lowest value of the underlying state variable) and a condition on the first derivative of the contingent claim value with respect to the underlying state variable at the upper boundary (the boundary with the highest value of the underlying state variable)[9].

The three remaining boundaries, all lower boundaries, are specified as a condition on a linear combination of the first derivatives of the contingent claim value with respect to time and the underlying state variable respectively (in one of these three cases the contingent claim value is also included in the linear combination)[10].

---

[9]Denote the value of the contingent claim by $W(S,t)$, where S is the underlying state variable and t is calendar time. (The notation in this footnote is exactly identical to the notation that is used further on in the paper.) A condition directly on the contingent claim value at the lower boundary can be expressed as

$$W(S^l,t) = g^t(t),$$

where g^t is a known function and S^l is the lower boundary of the domain of $W(S,t)$ in the dimension of S. A condition on the first derivative of the contingent claim value with respect to the underlying state variable at the upper boundary can be expressed as

$$\frac{\partial W(S^u,t)}{\partial S} = f^t(t),$$

where f^t is a known function and S^u is the upper boundary of the domain of $W(S,t)$ in the dimension of S. (More about this further on.)

[10]In these three cases, the value of the underlying state variable is equal to zero at the lower boundary.

Some sample articles mention that, in some cases, it is not self-evident how exactly to model the boundary conditions. More precisely, three sample articles (all three value contingent claims with the firm value as underlying state variable) mention the difficulty in correctly modelling the possibility for bankruptcy. Also, 5 sample articles mention that the assumptions made about investors' and issuers' decision rules affect the modelling of the side conditions in general.

Despite the difficulty in modelling the boundary conditions in some cases, alternative modelling choices are, however, not discussed in any of the sample articles. The lack of discussion about alternative modelling choices implies, of course, that no sample article mentions possible differences in the results due to different choices in the modelling of the boundary conditions.

In table 3.1, it can be seen that some contingent claim types are valued in more than one sample article. Therefore, it can be interesting to compare different articles' specification of the boundary conditions for the same contingent claim type. One should, however, not expect to find much difference in the modelling choices in a comparison like this. This is partly due to the fact that most finance articles that use the implicit finite difference method follow the same tradition and use many of the same references (this is true for all articles in the sample used in this section).

Let us start with stock options (warrants are also included in this group since the article that values warrants treats them as ordinary call options). In all sample articles, except one, where the boundary conditions are specified for a specific type of stock option they are specified in the same way. The lower boundary condition is modelled as a condition directly on the option value and the upper boundary condition is modelled as a condition on the first derivative of the option value with respect to the stock price for all different kinds of stock option types, in these articles.

The lower boundary condition is then achieved by setting the underlying state variable equal to zero in the partial differential equation that the value of the contingent claim must satisfy. Furthermore, in two of the three cases the underlying state variable is the interest rate. Also, in both of these latter cases the interest rate process is modelled in a way that allows zero as a possible lowest value of the interest rate.

Due to transformation of variables, the stock options valued in the remaining article are not completely comparable to the stock options valued in the other sample articles. For all stock options in the remaining article (this article values European call and put options as well as American call and put options) both boundaries are, however, modelled as a condition on the first derivative of the option value with respect to the stock price.

Besides the reason previously mentioned, the strong consensus of how to model the boundary conditions for stock options is due to the simplicity of the instruments.

Next, the two articles that value convertible bonds with the stock price as underlying state variable[11] are taken into consideration. It should be noticed that valuations of convertible bonds are often not entirely comparable, due to different put, call and conversion features.

To the point, the two articles that value convertible bonds with the stock price as underlying state variable model the boundary conditions somewhat differently. One of the articles models both boundary conditions as conditions directly on the convertible bond value, while the other article models both boundaries as conditions on the first derivative of the convertible bond value with respect to the underlying stock price.

Finally[12], the boundary conditions are compared in the sample articles that value convertible bonds with the firm value as underlying state variable. In table 3.1, it can be seen that this type of contingent claim is valued in no less than four of the articles in the sample. In one of those articles, the boundary conditions are not specified. Thus, there are three articles to consider. In all three articles, the boundary conditions are modelled in exactly the same way. This is, however, more or less to be expected. This follows from the fact that one of the articles is the classical paper by *Brennan and Schwartz* (see [21]), and the other two articles are influenced by that article.

---

[11]The sample article that values LYONs also can be considered to belong to the group of sample articles that value convertibles with the stock price as underlying state variable. It is, however, not exactly clear how the boundary conditions are modelled in that article.

[12]In table 3.1, it can be seen that retractable bonds are valued in two sample articles. The boundary conditions are, however, not specified in one of those articles. Thus, there are not any two pairs of boundary conditions to compare in the case of retractable bonds.

3.3 Derivation of the fundamental partial differential equation

Although the price of the contingent claim can depend upon the evolution of many state variables, this paper will only be concerned with the case of one underlying state variable.

In order to derive the partial differential equation that the price of the contingent claim must satisfy, the following assumptions are introduced (see also [100] and [101]):

A.1 There are no transaction costs, taxes or problems with indivisibilities of assets.

A.2 There are a sufficient number of investors with comparable wealth levels, so that each investor believes that he/she can buy and sell as much of an asset as he/she wants at the market price.

A.3 There exists an exchange market for borrowing and lending at the same rate of interest.

A.4 Short-sales of all assets, with full use of the proceeds, is allowed.

A.5 Trading in assets takes place continuously in time.

A.6(a) There is a riskless asset whose rate of return per unit time is a known function of calendar time, t. Denote this rate of return with $r(t)$.

A.7(a) The underlying state variable is the price of a traded security. The dynamics for the price of the underlying security, S, through time can be described by the stochastic differential equation,

$$dS = [\mu(S,t)S - D^t(S,t)]dt + \sigma(S,t)Sdz, \tag{3.1}$$

where

- $\mu(S,t) =$ the instantaneous expected rate of return on the underlying security per unit time,

- $D^t(S,t)$ = the instantaneous cash payout to the owners of the underlying security per unit time,
- $\sigma^2(S,t)$ = the instantaneous variance per unit time of the rate of return on the underlying security,
- $z(t)$ = a Wiener process, that is $E(dz) = 0$, $Var(dz) = dt$, $Cov(dz(t), dz(s)) = 0 \quad s \neq t$ and $dz \sim N(0, dt)$ where N indicates the normal distribution.

A.8 It is assumed that investors prefer more to less. It is also assumed that investors agree upon $\sigma(S,t)$, but it is *not* assumed that they necessarily agree on $\mu(S,t)$.

A.9 The price of the contingent claim is a function of the price of the underlying security and time. Denote the price of the contingent claim by $W(S,t)$. $W(S,t)$ is a real valued continuous function such that $\frac{\partial W}{\partial t}$, $\frac{\partial W}{\partial S}$ and $\frac{\partial^2 W}{\partial S^2}$ are continuous[13]. Furthermore, $W(S,t)$ also has the following properties:

- The owners of the contingent claim will receive an instantaneous payout per unit time, $d^t(S,t) \; t_0 \leq t < T + t_0$, where t_0 is the current time and T is the time to maturity (from t_0) for $W(S,t)$.
- Denote the upper boundary of the domain of $W(S,t)$ in the price dimension by S^u. For any t ($t_0 \leq t < T + t_0$), if $S(t) = S^u$, then $W_S(S^u, t) = f^t(t)$[14],

[13] In the remainder of this paper, partial derivatives will often be denoted with the help of subindexes. That is, the following fairly standard notation for partial derivatives of a function $f(x,y)$ will be used: $\frac{\partial f(x,y)}{\partial x} = f_x(x,y)$, $\frac{\partial f(x,y)}{\partial y} = f_y(x,y)$, $\frac{\partial^2 f(x,y)}{\partial x^2} = f_{xx}(x,y)$, $\frac{\partial^2 f(x,y)}{\partial y^2} = f_{yy}(x,y)$ and $\frac{\partial^2 f(x,y)}{\partial x \partial y} = f_{xy}(x,y)$.

[14] In this paper it is assumed that a condition on the first derivative of the contingent claim price with respect to the price of the underlying security can be used at the upper boundary, and a condition directly on the contingent claim price can be used at the lower boundary. As seen in section 3.2, it is rather common to model the boundary conditions like this. It is also like this that the boundary conditions are modelled in the numerical valuations later on in this paper.

Another fairly common type of boundary condition is to set the second derivative of the contingent claim price with respect to the price of the underlying security equal to zero at one or both of the boundaries.

where f^t is a known function. f^t gives a condition on $W(S,t)$ at the upper boundary.

- Denote the lower boundary of the domain of $W(S,t)$ in the price dimension by S^l[15]. For any t ($t_0 \leq t < T + t_0$), if $S(t) = S^l$ ($S^l < S^u$), then $W(S^l,t) = g^t(t)$, where g^t is a known function. g^t gives the value of $W(S,t)$ at the lower boundary[16].

- For $t = T+t_0$, the value of $W(S,T+t_0)$ is given by $W(S,T+t_0) = v(S)$, where v is a known function. v gives the value of the contingent claim at maturity. [$W(S,T+t_0) = v(S)$ is called the initial condition since finite difference methods work backwards through time. They start at maturity and iterate towards current time.]

Many of the aforementioned assumptions are not necessary for the model to hold (see [100]). In particular, the "perfect market" assumptions **A.1-A.4** can be substantially weakened.

[15] In this paper it is assumed that the upper and lower boundary of the domain of $W(S,t)$ are constants. In a more general setting the upper and lower boundary can, however, be functions of time.

[16] As mentioned earlier, many pricing problems within CCA are pure initial value problems. In these cases there are no natural boundaries and thus no natural functions that give conditions on the price function of the contingent claim at some upper and lower boundary for the underlying state variable. That is, there are no natural functions $f^t(t)$ and $g^t(t)$. When using the implicit finite difference method, however, boundaries and conditions on the price function of the contingent claim at these boundaries have to be specified. Let us exemplify this.

When the price process of the underlying security is modelled as in equation (3.1) [with $D^t(S,t) = cS$ and with c, μ and σ as constants] it is easy to verify that $S \in]0, \infty[$ (e.g., see [42] or [89]). Usually, the price function of the contingent claim is defined on the whole interval of S and thus no natural boundaries or boundary conditions exist.

To be able to use the implicit finite difference method the true domain of the price function has to be approximated with $S \in [0, R]$, where R is a sufficiently high real number. Then approximate conditions on the price function of the contingent claim at the boundaries of the approximate domain are modelled. It is exactly this approach that is used in the numerical valuations later on in this paper.

In order to derive the partial differential equation that the price of the contingent claim must satisfy the following lemma will be used:

lemma 3.1 (Itô's) [17] *Let $u(x,t)$ be a real valued continuous non-random function with continuous partial derivatives $\frac{\partial u}{\partial t} = u_t$, $\frac{\partial u}{\partial x} = u_x$ and $\frac{\partial^2 u}{\partial x^2} = u_{xx}$. Suppose that $x(t)$ is a real valued diffusion process with stochastic differential,*

$$dx = f(x,t)dt + b(x,t)dy,$$

where $y(t)$ is a Wiener process and f and b are real valued functions of x and t. Then $u(x,t)$ has a differential given by,

$$\begin{aligned} du &= [u_t(x,t) + u_x(x,t)f(x,t) + \frac{1}{2}u_{xx}(x,t)b^2(x,t)]dt \\ &\quad + u_x(x,t)b(x,t)dy. \end{aligned}$$

Apply Itô's lemma at $W(S,t)$, where dS follows equation (3.1). This gives

$$\begin{aligned} dW &= \{W_t(S,t) + W_S(S,t)[\mu(S,t)S - D^t(S,t)] \\ &\quad + \frac{1}{2}W_{SS}(S,t)\sigma^2(S,t)S^2\}dt + W_S(S,t)\sigma(S,t)Sdz. \end{aligned} \tag{3.2}$$

In the remainder of the derivation of the fundamental partial differential equation, the arguments in the functions will not be written out explicitly.

Assume that the cash amount α_1 is invested in the contingent claim. The instantaneous return from this investment is

$$\alpha_1(dW/W) + \alpha_1(d^t/W)dt. \tag{3.3}$$

Assume also that the cash amount α_2 is invested in the underlying security. This investment gives the instantaneous return

$$\alpha_2(dS/S) + \alpha_2(D^t/S)dt. \tag{3.4}$$

Form a portfolio of the investments equation (3.3) and equation (3.4), and choose the amounts α_1 and α_2 such that all risk disappears. Since the portfolio is risk-free it can be

[17] For a more mathematically correct formulation and a proof of Itô's lemma see [89].

financed by borrowing the amount $(\alpha_1 + \alpha_2)$ at the instantaneous riskfree interest rate, $r(t)$. The portfolio consisting of equation (3.3) and equation (3.4) (with the appropriate choices of α_1 and α_2) financed with the borrowed amount $(\alpha_1 + \alpha_2)$ at the riskfree interest rate does not require any net investment and is also without risk. Therefore, to avoid the possibility of arbitrage profits the return on the portfolio reduced by its cost of financing must be zero. Thus,

$$\begin{aligned} \alpha_1[(dW/W) + (d^t/W)dt] + \alpha_2[(dS/S) + (D^t/S)dt] & \\ -(\alpha_1 + \alpha_2)rdt &= 0. \end{aligned} \tag{3.5}$$

Insert (3.1) and (3.2) into (3.5), and after a slight reordering of terms this gives

$$\begin{aligned} \alpha_1 \frac{[W_t + W_S(\mu S - D^t) + (1/2)W_{SS}\sigma^2 S^2 + d^t]dt}{W} + \alpha_2 \mu dt & \\ +(\frac{\alpha_1 W_S \sigma S}{W} + \alpha_2 \sigma)dz - r(\alpha_1 + \alpha_2)dt &= 0. \end{aligned} \tag{3.6}$$

The only stochastic element in equation (3.6) comes from the Wiener process z. So if the amounts α_1 and α_2 are chosen such that the coefficient of dz always is zero, then all risk in the portfolio vanish. Therefore set

$$\frac{\alpha_1 W_S \sigma S}{W} + \alpha_2 \sigma = 0. \tag{3.7}$$

Use equation (3.7) to find the solution to α_2 in terms of α_1. Insert this solution into equation (3.6). This establishes

$$\frac{\alpha_1}{W}[W_t + (rS - D^t)W_S + \frac{1}{2}W_{SS}\sigma^2 S^2 - Wr + d^t]dt = 0. \tag{3.8}$$

From equation (3.8) it can be seen that the partial differential equation that $W(S,t)$ must satisfy if arbitrage profits should be prevented is

$$\begin{aligned} \frac{1}{2}W_{SS}(S,t)\sigma^2(S,t)S^2 + [r(t)S - D^t(S,t)]W_S(S,t) & \\ -r(t)W(S,t) + W_t(S,t) + d^t(S,t) &= 0. \end{aligned} \tag{3.9}$$

Equation (3.9) is a parabolic partial differential equation. Although an equation with variable coefficients, as equation (3.9), can switch from a parabolic to either an elliptic or hyperbolic form, equation (3.9) remains parabolic (see appendix BI for a further discussion of this).

A complete description of the partial differential equation (3.9) also requires specifications of side conditions. It is these conditions that distinguish one contingent claim from another. In this case, equation (3.9) is subject to the fairly general boundary and initial conditions given in assumption **A.9**, and restated below

$$
\begin{aligned}
W_S(S^u,t) &= f^t(t) & t_0 \le t < T+t_0, && (3.10)\\
W(S^l,t) &= g^t(t) & t_0 \le t < T+t_0, && (3.11)\\
W(S,T+t_0) &= v(S) & S^l \le S \le S^u. && (3.12)
\end{aligned}
$$

In most applications of CCA the following extensions of assumptions **A.6**(a) and **A.7**(a) are made:

A.6(b) The instantaneous riskless rate of interest, r, is known and constant.

A.7(b) The instantaneous variance per unit time of the rate of return on the underlying security, σ^2, is constant[18].

It is often more convenient to express equation (3.9) subject to (3.10), (3.11) and (3.12) in terms of time to maturity, τ, rather than in terms of calendar time, t. The relation between t and τ is

$$\tau = T - (t - t_0) \qquad t_0 \le t \le T + t_0. \qquad (3.13)$$

It is also rather common that[19]

$$S^l = 0 \qquad t_0 \le t < T + t_0. \qquad (3.14)$$

[18]When r and σ are assumed to be independent of time [$D^t(S,t)$ must also be independent of time] the numerical solution of (3.9) subject to (3.10), (3.11) and (3.12), with the implicit finite difference method is simplified. Assumptions **A.6**(b) and **A.7**(b) [if $D^t(S,t)$ is also independent of time] make the coefficient matrix (see subsection 3.4.2) constant in all iterations. This greatly reduces the computational effort, when the implicit finite difference method is used.

[19]Equation (3.14) simply says that the price of the contingent claim, $W(S,t)$, is defined for all non-negative prices of the underlying security less than or equal to S^u (also, see the discussion regarding the boundaries in the footnotes in assumption **A.9**).

With the use of assumptions **A.6**(b) and **A.7**(b), and the substitution of equations (3.13) and (3.14) into equation (3.9) subject to (3.10), (3.11) and (3.12), the partial differential equation and its side conditions can be rewritten as

$$\begin{aligned} \frac{1}{2}H_{xx}(x,\tau)\sigma^2x^2 + [rx - D(x,\tau)]H_x(x,\tau) & \\ -rH(x,\tau) - H_\tau(x,\tau) + d(x,\tau) &= 0, \end{aligned} \tag{3.15}$$

subject to,

$$H_x(x^u,\tau) = f(\tau) \qquad 0 < \tau \leq T, \tag{3.16}$$

$$H(0,\tau) = g(\tau) \qquad 0 < \tau \leq T, \tag{3.17}$$

$$H(x,0) = v(x) \qquad 0 \leq x \leq x^u. \tag{3.18}$$

To derive equation (3.15) subject to (3.16), (3.17) and (3.18),

$$\begin{aligned} S(t) &= x(\tau), \\ W(S,t) &= H(x,\tau), \\ W_S(S,t) &= H_x(x,\tau), \\ W_{SS}(S,t) &= H_{xx}(x,\tau), \\ W_t(S,t) &= -H_\tau(x,\tau), \\ D^t(S,t) &= D(x,\tau), \\ d^t(S,t) &= d(x,\tau), \\ S^u &= x^u, \\ S^l &= x^l \quad (S^l = x^l = 0), \\ f^t(t) &= f(\tau), \\ g^t(t) &= g(\tau), \end{aligned}$$

have been used.

The partial differential equation (3.15) subject to the boundary and initial conditions (3.16), (3.17) and (3.18) is valid for many significant valuation problems. In spite of advances made in recent years, many of these problems lack straightforward closed-form solutions.

The various approaches that have been suggested for calculating contingent claim prices, when there is no closed-form solution, include analytical approximation, compound option methods, series solutions, numerical integration, binomial models and finite difference methods. For further discussions about these methods see [54], [67], [68] and [82].

3.4 The implicit finite difference method

The implicit finite difference method is one of many finite difference methods that can be used to solve partial differential equations numerically. These methods have been used for a long time in the natural sciences and in technical applications.

The major breakthrough for the implicit finite difference method in the field of finance came little over a decade ago in a paper by *Eduardo Schwartz* (see [122]). The implicit finite difference method was then, in the late seventies, further popularized by *Brennan and Schwartz* (see [19], [20], [21] and [25]). Nowadays, the method is one of the most commonly used ones for solving numerically the partial differential equation (3.15) subject to the boundary and initial conditions (3.16), (3.17) and (3.18). Examples of CCA applications, where the implicit finite difference method has been used, are [19], [21], [25], [27], [38], [73] and [95] (cf. also section 3.2).

More importantly, the implicit finite difference method is perhaps that approximating technique that can be used to solve (3.15) subject to the boundary and initial conditions (3.16), (3.17) and (3.18) which has the best chances to become widely used for practical purposes. The main reasons for this are that the implicit finite difference method is very flexible and has good stability and convergence properties[20] (see [54], [82] and [126]).

3.4.1 Finite difference methods

Subsection 3.4.2 gives a relatively extensive description of the implicit finite difference method. Before that, it can be of interest to put the implicit finite difference method

[20]The terms stability and convergence are explained in one footnote each in subsection 3.4.1.

into a larger context. This is done by very briefly describing finite difference methods, for solving parabolic partial differential equations, in general.

Equation (3.15) subject to conditions (3.16), (3.17) and (3.18) is the problem that is going to be solved. For the convenience of the readers, the problem is restated below,

$$\begin{aligned}\frac{1}{2}H_{xx}(x,\tau)\sigma^2x^2 + [rx - D(x,\tau)]H_x(x,\tau) \\ -rH(x,\tau) - H_\tau(x,\tau) + d(x,\tau) &= 0\end{aligned}$$

subject to

$$\begin{aligned} H_x(x^u,\tau) &= f(\tau) && 0 < \tau \leq T, \\ H(0,\tau) &= g(\tau) && 0 < \tau \leq T, \\ H(x,0) &= v(x) && 0 \leq x \leq x^u. \end{aligned}$$

For the numerical solution of equation (3.15) subject to conditions (3.16), (3.17) and (3.18) by finite difference methods, a grid of mesh points $(x,\tau) = (ih, jk)$ is introduced. h and k are mesh parameters which are small and thought of as tending to zero, and i and j are integers, $i \geq 0$ and $j \geq 0$. An approximate solution to (3.15) subject to (3.16), (3.17) and (3.18) at these mesh points is then looked for. This approximate solution will be denoted by $H_{i,j}$. $H_{i,j}$ is obtained by solving a problem in which the derivatives in equation (3.15) have been replaced by finite difference quotients. In order to formalize this, the following notations are introduced:

$$\begin{aligned} h &= \Delta x, && && (3.19) \\ x_i &= ih && i = 0,1,\ldots,n, \quad (x \in [0,nh]; x^u = nh) && (3.20) \\ k &= \Delta\tau, && && (3.21) \\ \tau_j &= jk && j = 0,1\ldots,m, \quad (\tau \in [0,mk]; T = mk) && (3.22) \\ H(x_i,\tau_j) &= H(ih,jk) = H(i,j) \approx H_{i,j}. && && (3.23) \end{aligned}$$

To obtain a general finite difference approximation to equation (3.15), the following difference quotients are introduced[21]:

$$\partial_x^f H(i,j) = \frac{[H(i+1,j) - H(i,j)]}{h} = H_x(i,j) + O(h), \tag{3.24}$$

$$\partial_x^b H(i,j) = \frac{[H(i,j) - H(i-1,j)]}{h} = H_x(i,j) + O(h), \tag{3.25}$$

$$\partial_x H(i,j) = \frac{[H(i+1,j) - H(i-1,j)]}{2h} = H_x(i,j) + O(h^2), \tag{3.26}$$

$$\partial_x^f \partial_x^b H(i,j) = \frac{[H(i+1,j) - 2H(i,j) + H(i-1,j)]}{h^2} = H_{xx}(i,j) + O(h^2), \tag{3.27}$$

$$\partial_\tau^f H(i,j) = \frac{[H(i,j+1) - H(i,j)]}{k} = H_\tau(i,j) + O(k), \tag{3.28}$$

$$\partial_\tau^b H(i,j) = \frac{[H(i,j) - H(i,j-1)]}{k} = H_\tau(i,j) + O(k). \tag{3.29}$$

Above, $\partial_\bullet^b$ denotes a backward difference approximation, $\partial_\bullet^f$ a forward difference approximation and $\partial_\bullet$ a central difference approximation. $O(\bullet)$ represents the order of the errors in the difference approximations. A fairly general finite difference approximation to equation (3.15), at grid point $(i, j+1)$, is given by

$$\begin{aligned} \theta\partial_\tau^f H_{i,j} + (1-\theta)\partial_\tau^b H_{i,j+1} &= \\ \theta\{\frac{[\sigma^2(ih)^2]}{2}\partial_x^f\partial_x^b H_{i,j} + (rih - D_{i,j+1})\partial_x H_{i,j} - rH_{i,j+1} + d_{i,j+1}\} &+ \\ (1-\theta)\{\frac{[\sigma^2(ih)^2]}{2}\partial_x^f\partial_x^b H_{i,j+1} + (rih - D_{i,j+1})\partial_x H_{i,j+1} - rH_{i,j+1} + d_{i,j+1}\}, & \end{aligned} \tag{3.30}$$

where $\quad 0 \leq \theta \leq 1$[22].

---

[21]The operators ∂_x^f, $\partial_x^b, \ldots$ work exactly the same way at the discrete approximations $H_{i,j}$ as at the exact values $H(i,j)$. Thus, $\partial_x^f H_{i,j} = \frac{(H_{i+1,j} - H_{i,j})}{h}$, $\partial_x^b H_{i,j} = \frac{(H_{i,j} - H_{i-1,j})}{h}, \ldots$ and so on.

[22]Since equation (3.30) is a finite difference approximation equation to (3.15) at grid point $(i, j+1)$, the difference quotients with subindex i, j may be confusing. (The difference quotient $\partial_\tau^f H_{i,j}$ is, however, not confusing since $\partial_\tau^f H_{i,j} = \partial_\tau^b H_{i,j+1}$.) The explanation to this is as follows:

The explicit finite difference method (see sub-subsection 3.4.1.1) is obtained by setting $\theta = 1$ in equation (3.30). When using the explicit finite difference method in pricing problems within CCA, the tradition is to approximate the partial derivatives of H with respect to x at grid point $(i, j+1)$ with $\partial_x^f \partial_x^b H_{i,j}$ and $\partial_x H_{i,j}$ (e.g., see [22] p.463, [67] p.238 or [69] p.89). Thus, this tradition gives that the "explicit finite difference approximation" at grid point (i,j) can be written as

$$\partial_\tau^b H_{i,j} = \partial_\tau^f H_{i,j-1} =$$

In order to obtain an approximate solution to equation (3.15) subject to conditions (3.16), (3.17) and (3.18) finite difference methods work backwards through time. The initial condition (3.18) gives the exact values of $H_{i,0}$ $i = 0, \ldots, n$. Given $H_{i,0}$ $i = 0, \ldots, n$, equation (3.30) together with discrete versions of the boundary conditions (3.16) and (3.17) provides a way to find $H_{i,1}$ $i = 0, \ldots, n$, and so on ... until $H_{i,m}$ $i = 0, \ldots, n$ are found. The current price of the contingent claim is $H_{i_0,m}$, where $i_0 h = x(T)$ and $x(T)$ is the current price of the underlying security.

By altering the coefficient θ, a variety of finite difference methods with different characteristics can be achieved. Three important cases will be described very briefly below.

$$\frac{[\sigma^2(ih)^2]}{2}\partial_x^f\partial_x^b H_{i,j-1} + (rih - D_{i,j})\partial_x H_{i,j-1} - rH_{i,j} + d_{i,j}. \tag{3.31}$$

Another possible, and perhaps more straightforward, explicit finite difference approximation to equation (3.15) at grid point (i, j) is given by

$$\partial_\tau^f H_{i,j} =$$
$$\frac{[\sigma^2(ih)^2]}{2}\partial_x^f\partial_x^b H_{i,j} + (rih - D_{i,j})\partial_x H_{i,j} - rH_{i,j} + d_{i,j}. \tag{3.32}$$

Brennan and Schwartz have shown that the finite difference approximation in equation (3.31) can be given a very nice economic interpretation (see [22]). The interpretation is based on risk-neutral pricing (e.g., see [42] and [61]), and is as follows:

At each node in a trinomial tree, the contingent claim value can be calculated as the expected value of the contingent claim values (that are possible, conditional upon being at the node under consideration) in the next time step, discounted at the riskfree interest rate. Thus, the contingent claim value at current time is obtained by working backwards through time, from the maturity of the contingent claim to current time, and at each node in the tree performing this riskfree discounting.

Furthermore, the finite difference approximation in equation (3.32) can be given a similar (and very closely related) economic interpretation. In this case, the interpretation is that the contingent claim value at each node in a trinomial tree can be calculated with the help of something similar to state prices times the values in the three possible states of the world that can occur in the next time step. Thus, the contingent claim value at current time can be obtained by step by step working backwards through time, from the maturity of the contingent claim to current time, using this "state price approach" at each node in the trinomial tree.

These descriptions will be approximations for the simple heat flow problem

$$u_\tau = u_{xx} \qquad \tau > 0; 0 < x < 1 \tag{3.33}$$

subject to

$$u(x,0) = f(x), \tag{3.34}$$

$$u(0,\tau) = 0, \tag{3.35}$$

$$u(1,\tau) = 0, \tag{3.36}$$

rather than for equation (3.15) subject to conditions (3.16), (3.17) and (3.18). There are two reasons for describing the finite difference methods for equation (3.33) subject to conditions (3.34), (3.35) and (3.36) instead of describing them for equation (3.15) subject to conditions (3.16), (3.17) and (3.18). The first is to simplify the notation. The second reason is that the stability[23] and convergence[24] results are not so simple when using finite difference methods to approximate equation (3.15) subject to conditions (3.16), (3.17) and (3.18), as when approximating equation (3.33) subject to conditions (3.34), (3.35) and (3.36)[25].

The latter reason above may seem irrelevant, but the results for the simple heat flow equation give indications for the properties of the methods also for more complex differential equations[26].

---

[23]In short, the problem of stability concerns the unstable growth or stable decay of the errors in the arithmetical operations needed when a numerical method is used. Each calculation carried out introduces a round-off error. Generally, a numerical method is stable when the cumulative effect of all the rounding errors introduced when the method is used is negligible (see [126] p.57).

[24]A finite difference method is said to be convergent when the approximate solution established with the help of the method approaches the exact solution as both $\Delta x = h$ and $\Delta\tau = k$ approach zero (see [126] p.55).

[25]The requirements to guarantee stability and convergence that are given in sub-subsections 3.4.1.1, 3.4.1.2 and 3.4.1.3 are also valid for equation (3.33) on a more general region and subject to more general boundary condititions (than conditions (3.35) and (3.36)). See [53] for more details about this topic.

[26]The exact conditions for stability and convergence, for different finite difference methods, when approximating equation (3.15) subject to conditions (3.16), (3.17) and (3.18) must be determined in

One further motivation for describing the finite difference methods applied to equation (3.33) subject to conditions (3.34), (3.35) and (3.36) is that equation (3.15) in many cases can be transformed into the simple heat flow equation (3.33) (the transformed side conditions might, of course, have different forms). A fairly general example of this is given in appendix BII.

Apply difference quotients (3.27), (3.28) and (3.29) to equation (3.33). This gives the finite difference approximation

$$\theta \partial_\tau^f U_{i,j} + (1-\theta)\partial_\tau^b U_{i,j+1} = \theta \partial_x^f \partial_x^b U_{i,j} + (1-\theta)\partial_x^f \partial_x^b U_{i,j+1}, \tag{3.37}$$

where $0 \leq \theta \leq 1$ and U is the solution to the difference equation used to approximate the partial differential equation (3.33).

With equations (3.27), (3.28) and (3.29) equation (3.37) can be restated as

$$U_{i,j+1} - U_{i,j} = \frac{k}{h^2}[\theta(U_{i+1,j} - 2U_{i,j} + U_{i-1,j}) + (1-\theta)(U_{i+1,j+1} - 2U_{i,j+1} + U_{i-1,j+1})]. \tag{3.38}$$

3.4.1.1 The finite difference method when $\theta = 1$

With $\theta = 1$ equation (3.38) becomes,

$$U_{i,j+1} = U_{i,j} + \frac{k}{h^2}(U_{i+1,j} - 2U_{i,j} + U_{i-1,j}). \tag{3.39}$$

The method to solve numerically equation (3.33) using the approximation in equation (3.39) is called the explicit finite difference method. The method is called explicit because it is possible to explicitly find the solutions to the unknown values, $U_{i,j+1}$ $i = 1, \ldots, n-1$, in terms of the known values, $U_{i,j}$ $i = 0, \ldots, n$, by using equation (3.39) in every iteration. This feature, that one can explicitly find the solutions to the unknown

each separate case. Actually, stability refers to difference problems *with* initial and boundary conditions rather than to difference equations alone.

Anyway, in most option pricing problems the method converges if it is stable (see [69] p.92). {If convergence and stability are understood in the sense of Lax and Richtmyer it can be shown that for a difference equation approximating a differential equation in the formal sense, convergence and stability are equivalent (see [53] pp.135-137).} Further, the unstability of a method (for given choices of $\Delta\tau$ and Δx), in a specific case, is often obvious from the numerical results.

values, $U_{i,j+1}$ $i = 1, \ldots, n-1$, gives the explicit method the advantage of being computationally very simple. The main drawback of the explicit finite difference method is its bad convergence and stability properties[27].

To compare different finite difference methods the mesh ratio λ is introduced, where

$$\lambda = \frac{k}{h^2}. \tag{3.40}$$

Remember that $k = \Delta\tau$ and $h = \Delta x$. It can be shown that the requirement for the explicit finite difference method to be stable and convergent (in this case) is that $0 < \lambda \leq (1/2)$ (see [126]). This means that the time step $k = \Delta\tau$ necessarily is very small because the method is valid only for $k \leq (1/2)h^2$, and $h = \Delta x$ must be kept small in order to attain reasonable accuracy.

3.4.1.2 The finite difference method when $\theta = 0$

With $\theta = 0$ equation (3.38) changes to,

$$\frac{k}{h^2}U_{i+1,j+1} - (1 + 2\frac{k}{h^2})U_{i,j+1} + \frac{k}{h^2}U_{i-1,j+1} = -U_{i,j}. \tag{3.41}$$

The numerical method to solve equation (3.33) with the approximaton in equation (3.41) is called the (fully) implicit finite difference method.

The implicit finite difference method iterates through time. In each iteration the solutions to the $n+1$ unknowns, $U_{i,j+1}$ $i = 0, \ldots, n$, have to be found. Equation (3.41) describes a connection between the known value, $U_{i,j}$, and the three unknown values $U_{i-1,j+1}, U_{i,j+1}$ and $U_{i+1,j+1}$ $i = 1, \ldots, n-1$. This means that equation (3.41) together with the two boundary conditions give $n+1$ equations that can be used to find the solutions to the $n+1$ unknowns. The solution technique, when using the implicit finite difference method, is described in more detail in section 3.4.2.

The disadvantage of the implicit finite difference method, compared to the explicit finite difference method, is that in every iteration we have to solve $n+1$ simultaneous

[27] It can, however, be shown that the convergence and stability properties for the explicit finite difference method can be improved, for the standard CCA problem, by a logarithmic transformation ($y = \ln x$) (e.g., see [22] p.462).

equations. Nevertheless, this disadvantage need not be so severe since the coefficient matrix is constant through all iterations in many applications. It is certainly constant in this simple case (for more details see section 3.4.2).

The advantage of the implicit finite difference method, compared to the explicit finite difference method, is its good stability and convergence properties. In this case the implicit finite difference method is stable and convergent for all values of the mesh ratio, $\lambda = (k/h^2)$ (see [126]). This means that stability and convergence do not have to be worried about, and that it is not essential to choose the time step $\Delta\tau$ much smaller than Δx. This implies that the number of iterations can be considerably reduced.

There is, however, another problem with the implicit finite difference method that makes it suitable to choose the time step , $\Delta\tau$, smaller than the step in the price of the underlying security, Δx. As the difference quotients [i.e., equations (3.26), (3.27) and (3.29)] are chosen, the implicit method is only first order accurate in time while it is second order accurate in the price of the underlying security. This means that the error in the time discretization will dominate, unless the time step is chosen smaller than the step in the price of the underlying security. It would thus be desirable to find a method which is also second order accurate in time. The following sub-subsection describes such a method.

3.4.1.3 The finite difference method when $\theta = \frac{1}{2}$

With $\theta = \frac{1}{2}$ equation (3.38) changes to

$$-\frac{k}{h^2}U_{i-1,j+1} + (2+2\frac{k}{h^2})U_{i,j+1} - \frac{k}{h^2}U_{i+1,j+1} = \frac{k}{h^2}U_{i-1,j} + (2-2\frac{k}{h^2})U_{i,j} + \frac{k}{h^2}U_{i+1,j}. \quad (3.42)$$

The numerical method to solve the partial differential equation (3.33) with the approximation in equation (3.42) is the well-known Crank-Nicholson method.

The Crank-Nicholson method has two main advantages. Firstly, the method is stable and convergent without any restrictions on the mesh ratio, $\lambda = (k/h^2)$ (in this case). Secondly, the method is second order accurate in both time and the price of the underlying security (see [53]).

The disadvantage of the Crank-Nicholson method is that it is more complicated than the explicit or the (fully) implicit finite difference method.

In spite of the advantages of the Crank-Nicholson method it is seldom used in applications of contingent claims analysis.

Since the main purpose of this paper is to investigate how errors in the boundary conditions affect the solutions when the implicit finite difference method is used, this method will be described in more detail in the next subsection.

3.4.2 The implicit finite difference method in detail

This subsection will give a fairly extensive description of the implicit finite difference method applied on the partial differential equation (3.15) subject to conditions (3.16), (3.17) and (3.18). In subsection 3.4.1, it was shown that equation (3.30) was a fairly general finite difference approximation to equation (3.15). Moreover, subsection 3.4.1 also showed that the "implicit finite difference approximation" was obtained by setting $\theta = 0$ in equation (3.30).

Using the above facts and the difference quotients (3.26), (3.27) and (3.29) gives the difference approximation

$$(a_i - D_{i,j+1}\frac{k}{2h})H_{i-1,j+1} + b_i H_{i,j+1} + (c_i + D_{i,j+1}\frac{k}{2h})H_{i+1,j+1} = H_{i,j} + kd_{i,j+1}, \qquad (3.43)$$

where

$$\begin{aligned}
a_i &= \frac{rik}{2} - \frac{\sigma^2 i^2 k}{2}, \\
b_i &= \sigma^2 i^2 k + (1 + rk), \\
c_i &= -\frac{rik}{2} - \frac{\sigma^2 i^2 k}{2}, \\
i &= 1, \ldots, n-1, \\
j &= 0, \ldots, m-1.
\end{aligned}$$

The subsequent description will be simplified by assuming that the instantaneous cash payout to the owners of the underlying security is a constant yield. Also, it will be assumed that the owners of the contingent claim will receive a constant instantaneous payout per unit time. That is,

$$c = \frac{D(x,\tau)}{x(\tau)} = constant, \tag{3.44}$$

and

$$d(x,\tau) = d = constant. \tag{3.45}$$

The simplification in equation (3.44) is the important one. With this simplification [in addition to assumptions **A.6**(b) and **A.7**(b) in section 3.3], the coefficient matrix $\mathbf{A}$ (the matrix $\mathbf{A}$ is defined later in this subsection) is the same in all iterations. This will greatly reduce the computational effort to obtain an approximate solution to $H(x,\tau)$. In many applications of CCA the simplifications, in equation (3.44) and equation (3.45), are quite natural and not so strong violations of reality. For example, consider the valuation of a convertible bond with the stock as the underlying security. In this case, equation (3.44) is equivalent to assuming a constant continuous dividend yield, and equation (3.45) is equivalent to assume a constant rate of continuous coupon payments to the convertible bondholders.

With the simplifications in equations (3.44) and (3.45) equation (3.43) changes to

$$a_i H_{i-1,j+1} + b_i H_{i,j+1} + c_i H_{i+1,j+1} = H_{i,j} + dk, \tag{3.46}$$

where

$$\begin{aligned}
a_i &= \frac{ki(r-c)}{2} - \frac{\sigma^2 i^2 k}{2}, \\
b_i &= \sigma^2 i^2 k + (1 + rk), \\
c_i &= \frac{ki(c-r)}{2} - \frac{\sigma^2 i^2 k}{2}, \\
i &= 1, \ldots, n-1, \\
j &= 0, \ldots, m-1.
\end{aligned}$$

Equation system (3.46) gives $(n-1)$ linear (independent) equations for the solution of the $(n+1)$ unknown contingent claim values $(H_{0,j+1}, \ldots, H_{n,j+1})$. Consequently, two more linear [and linearly independent of the equations in (3.46)] are needed for the purpose of uniquely solving the unknowns. These equations are given by discrete versions of the boundary conditions (3.16) and (3.17). That is,

$$\begin{aligned}
\frac{(H_{n,j+1} - H_{n-1,j+1})}{h} &= f([j+1]k) \qquad & j = 0,\ldots,m-1, \qquad (3.47)\\
H_{0,j+1} &= g([j+1]k) \qquad & j = 0,\ldots,m-1. \qquad (3.48)
\end{aligned}$$

In order to get a more compact notation equation (3.46) is rewritten in matrix form[28],

$$\begin{bmatrix}
b_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\
a_2 & b_2 & c_2 & 0 & 0 & \cdots & 0 \\
0 & a_3 & b_3 & c_3 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & & & \vdots \\
0 & \cdots & \cdots & 0 & a_{n-2} & b_{n-2} & c_{n-2} \\
0 & \cdots & \cdots & 0 & 0 & a_{n-1} & (b_{n-1} + c_{n-1})
\end{bmatrix}
\begin{bmatrix}
H_{1,j+1} \\ H_{2,j+1} \\ H_{3,j+1} \\ \vdots \\ H_{n-2,j+1} \\ H_{n-1,j+1}
\end{bmatrix}
=
\begin{bmatrix}
H_{1,j} + dk - a_1 g([j+1]k) \\
H_{2,j} + dk \\
H_{3,j} + dk \\
\vdots \\
H_{n-2,j} + dk \\
H_{n-1,j} + dk - hc_{n-1} f([j+1]k)
\end{bmatrix}, \qquad (3.49)$$

where $j = 0,\ldots,m-1$.

[28]When equation (3.49) was derived, the facts that $H_{n,j+1} = H_{n-1,j+1} + hf([j+1]k)$ and $H_{0,j+1} = g([j+1]k)$ [see equations (3.47) and (3.48)] were used in order to reduce the size of the equation system. This explains the terms $-a_1 g([j+1]k)$ and $-hc_{n-1}f([j+1]k)$ in the first and last equation of the right hand side of equation (3.49) respectively. It also explains the term $(b_{n-1} + c_{n-1})$ in the lower right corner of the coefficient matrix.

Introduce the following definitions:

$$\mathbf{A} = \begin{bmatrix} b_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & \cdots & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a_{n-2} & b_{n-2} & c_{n-2} \\ 0 & \cdots & \cdots & 0 & 0 & a_{n-1} & (b_{n-1}+c_{n-1}) \end{bmatrix}, \tag{3.50}$$

$$\mathbf{H_{j+1}} = \begin{bmatrix} H_{1,j+1} \\ H_{2,j+1} \\ H_{3,j+1} \\ \vdots \\ H_{n-2,j+1} \\ H_{n-1,j+1} \end{bmatrix} \qquad j = 0,\ldots,m-1, \tag{3.51}$$

$$\mathbf{HL_{j+1}} = \begin{bmatrix} H_{1,j} + dk - a_1 g([j+1]k) \\ H_{2,j} + dk \\ H_{3,j} + dk \\ \vdots \\ H_{n-2,j} + dk \\ H_{n-1,j} + dk - hc_{n-1} f([j+1]k) \end{bmatrix} \qquad j = 0,\ldots,m-1. \tag{3.52}$$

With the help of equations (3.50), (3.51) and (3.52), equation (3.49) can be rewritten as

$$\mathbf{AH_{j+1}} = \mathbf{HL_{j+1}} \qquad j = 0,\ldots,m-1. \tag{3.53}$$

Equation (3.53) gives $n-1$ linear equations to solve the $n-1$ unknown contingent claim prices $(H_{1,j+1}, H_{2,j+1}, \ldots, H_{n-1,j+1})$ in each iteration[29]. It can be verified that the coefficient matrix $\mathbf{A}$ is strictly diagonally dominant[30]. The diagonal dominance of $\mathbf{A}$ implies the following properties:

[29]The remaining two unknowns, $H_{0,j+1}$ and $H_{n,j+1}$ can be calculated with the help of the boundary conditions. These conditions are incorporated into equation system (3.53) to make it possible to compute $(H_{1,j+1}, H_{2,j+1}, \ldots, H_{n-1,j+1})$.

[30]The matrix $\mathbf{A}$ is not mathematically restricted to be diagonally dominant. It is, however, impossible

Firstly, **A** is nonsingular. So, the solution exists and is unique. Secondly, Gauss elimination, with no use of any form of pivoting, can be used for the solution of equation (3.53), without any unnecessary loss of accuracy (see [49] and [127]).

A discrete form of the initial values, i.e., equation (3.18), can be written as

$$H_{i,0} = v(ih) \qquad\qquad i = 0, \ldots, n. \tag{3.54}$$

The solution methodology when using the implicit finite difference method is as follows:

With equation (3.54), it is straightforward to find $\mathbf{HL_1}$. Then equation (3.53), with $j = 0$, gives a way to find $\mathbf{H_1}$. With $\mathbf{H_1}$ it is easy to find $\mathbf{HL_2}$ and so forth and so on. This continues until $\mathbf{H_m}$ is found. The desired value is now $H_{i_0,m}$, where $i_0 h = x(T)$ and $x(T)$ is the current price of the underlying security. If $i_0 h < x(T) < (i_0+1)h$ an approximation of the requested value, $H(x(T), T)$, can be achieved by means of interpolation.

The solution methodology is, in other words, an iteration backwards through time, where equation (3.53) is solved in every iteration.

From this solution methodology, it is easy to understand the value of having a matrix **A** that does not change when the method iterates backwards through time. If **A** remains the same in all iterations it has only to be inverted once, while **A** has to be inverted (in principle) in every iteration if **A** depends on j[31].

The solution methodology described above is directly applicable to European type contingent claims. The method has, however, to be slightly modified for American type contingent claims.

---

to find a realistic economic problem in which **A** is not strictly diagonally dominant (at least, with a proper specification of the problem).

[31] **A** is, of course, not inverted. **A** is tridiagonal and diagonal dominant. Thus, Gauss elimination and back-substitution are used in order to solve equation (3.53). The row multipliers can, however, be found once and for all, if **A** remains unchanged.

3.4.3 Extension of the implicit finite difference method to the valuation of American type contingent claims

In the case of an American type contingent claim, the partial differential equation (3.15) is subject to the further condition[32]

$$H(x, \tau^+) = \pi^\tau(x, H(x, \tau^-)) \qquad 0 < \tau \leq T; 0 < x < x^u, \tag{3.55}$$

where the superindex on π^τ is used to show that the function π^τ may alter with τ. τ^+ denotes the time (in chronological time) immediately before an early exercise opportunity, and τ^- denotes the time immediately after (also in chronological time) the early exercise opportunity.

One possible interpretation of equation (3.55) is that the owner of H has the opportunity to exercise the contingent claim at every value of τ. This means that π^τ is a maximum value function, where the owner of H maximizes the value of the contingent claim. Naturally, the value of the contingent claim before the decision, $H(x, \tau^+)$, is dependent on the value of the contingent claim after the decision (if not exercised), $H(x, \tau^-)$, and the price of the underlying security, x. This explains the arguments in π^τ.

Take an American call option on a stock as a concrete example of the discussion above. For an American call option condition (3.55) can be stated as

$$H(x, \tau^+) = max[x - Œ, H(x, \tau^-)], \tag{3.56}$$

where x is the stock price and Œ is the exercise price[33].

[32]In condition (3.55) it can be seen that the early exercise condition is not defined on the upper and lower boundary. An alternative interpretation is that the early exercise opportunity is taken into consideration when the boundary conditions (3.16) and (3.17) are specified.

[33]In the case of an American call option without any dividend payments it can, however, be shown that it can never be optimal to choose to exercise early (see e.g., [67]). Furthermore, for the case of an American call option with discrete dividend payments it can be shown that it can only be optimal to exercise at a time immediately before the stock goes ex-dividend (see e.g., [67]). For this reason, condition (3.56) for an American call option with discrete dividend payments (and also modelled with discrete dividend payments, and not with a dividend yield) is often stated as $H(x, \tau^+) = max[x - Œ, H(x - div_\tau, \tau^-)]$,

There are, at least, two approaches to include condition (3.55) into the solution algorithm, when using the implicit finite difference method.

3.4.3.1 Approach 1

If condition (3.55) is to be taken into consideration when the implicit finite difference method is used, condition (3.55) has to be rewritten in a discrete form,

$$H_{i,j} = \pi^j(ih, H^-_{i,j}) \qquad i = 1, \ldots, n-1; j = 1, \ldots, m. \tag{3.57}$$

In equation (3.57), $H^-_{i,j}$ $i = 1, \ldots, n-1$ denotes the approximate prices of the contingent claim in iteration j (i.e., when $\tau = jk$), *without* taking condition (3.57) into consideration. Consequently, $H_{i,j}$ $i = 1, \ldots, n-1$ are the approximate prices of the contingent claim, *with* condition (3.57) taken into consideration.

Approach 1 is very straightforward. $H^-_{i,j}$ $i = 1, \ldots, n-1$ are solved in iteration j, with the method for European type contingent claims. That is with equation (3.53). Then condition (3.57) is used to obtain the approximate prices, $H_{i,j}$ $i = 1, \ldots, n-1$, for the American type contingent claim. Naturally, it is the prices for the American type contingent claim that should be used in $\mathbf{HL_{j+1}}$, equation (3.52), when $H^-_{i,j+1}$ $i = 1, \ldots, n-1$ are solved in the next iteration[34].

Approach 1 can be interpreted as an assumption that the owners of the contingent claim check the early exercise condition (3.55) only at discrete points in time, i.e., at $\tau = jk; j = 1, \ldots, m$. This means that between the discrete time points jk and $(j+1)k$ the partial differential equation (3.15) subject to conditions (3.16) and (3.17) is permitted

where div_τ is the *discrete* dividend payment to the stockholders at time τ. The condition $H(x, \tau^+) = max[x - Œ, H(x - div_\tau, \tau^-)]$ has only to be taken into consideration at the discrete dividend dates.

[34]The derivative condition at the upper boundary [see equations (3.16) and (3.47)] implies that if it is assumed that the early exercise condition (3.55) is incorporated into the specifications of the boundary conditions (as mentioned in a footnote earlier in subsection 3.4.3), it has also to be assumed that $\frac{d\pi^\tau(x^u, H(x^u, \tau^-))}{dx} = H_x(x^u, \tau^-)$ if the modifications of the implicit finite difference method in subsubsections 3.4.3.1 and 3.4.3.2 should be entirely consistent. This is, however, approximately true for many common contingent claims (with a proper specification of the grid), e.g., American call option, American put option and convertible bond.

to evolve freely through time, *without* considering condition (3.55). Then at time point $(j+1)k$ the owner of H makes the decision whether exercising or not[35].

3.4.3.2 Approach 2

The approach discussed in this sub-subsection is a slight extension of the approach suggested by *Brennan and Schwartz* (see [20] pp.452-453 and [21] pp.1708-1709), and it is also similar to the approach proposed by *Courtadon* (see [38] pp.83-86). If equation (3.49) is observed it can be concluded that it is not possible to solve the unknowns, $H_{i,j}$ $i = 1, \ldots, n-1$, independently of each other, in iteration j. All unknowns are linked together in the tridiagonal equation system. Approach 2 is an extension of approach 1 that takes this feature into consideration.

Before describing approach 2, the assumption that condition (3.55) can be rewritten as

$$H(x, \tau^+) = max[\xi^\tau(x), H(x, \tau^-)] \qquad 0 < \tau \leq T; 0 < x < x^u \qquad (3.58)$$

is introduced. The superindex, τ, in equation (3.58) indicates that the function ξ^τ may alter with τ. Condition (3.58) is a very natural modification of condition (3.55). Since H is an American type contingent claim, the owner of H can choose to do nothing or to exercise. The value of doing nothing is given by $H(x, \tau^-)$, and the value of exercising is given by $\xi^\tau(x)$. The owners of H want to maximize the value of H. Therefore they choose to exercise if $\xi^\tau(x) > H(x, \tau^-)$. From this condition (3.58) follows directly.

In order to describe a solution algorithm, a discrete version of equation (3.58) has to be formulated. This discrete version is

$$H_{i,j} = max[\xi^j(ih), H_{i,j}^-] \qquad i = 1, \ldots, n-1; j = 1, \ldots, m. \qquad (3.59)$$

[35] Approach 1 may look as an ad hoc based solution method. It can, however, be shown (see [32]) that the approach to only take condition (3.55) into consideration at discrete points in time will yield solutions that uniformly converge to the true solution as $\Delta\tau \longrightarrow 0$. By the true solution is meant the solution to equation (3.15) subject to conditions (3.16), (3.17), (3.18) and (3.55).

Also, make the further assumption that either

$$\frac{d\xi^{\tau}(x)}{dx} \geq \frac{\partial H(x,\tau^{-})}{\partial x} \qquad 0 < \tau \leq T; 0 < x < x^{u}, \tag{3.60}$$

or[36]

$$\frac{d\xi^{\tau}(x)}{dx} \leq \frac{\partial H(x,\tau^{-})}{\partial x} \qquad 0 < \tau \leq T; 0 < x < x^{u}. \tag{3.61}$$

Approach 2 when condition (3.60) is fulfilled is somewhat different from approach 2 when condition (3.61) is fulfilled. Let us start with assuming that condition (3.60) is fulfilled.

Look again at equation (3.49). Since the matrix **A** is strictly diagonally dominant, equation (3.49) can, by simple Gauss elimination without any pivoting, be transformed into

$$\begin{bmatrix} \beta_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \beta_2 & c_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \beta_3 & c_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \beta_{n-2} & c_{n-2} \\ 0 & \cdots & \cdots & 0 & 0 & 0 & \beta_{n-1} \end{bmatrix} \begin{bmatrix} H_{1,j} \\ H_{2,j} \\ H_{3,j} \\ \vdots \\ H_{n-2,j} \\ H_{n-1,j} \end{bmatrix} = \begin{bmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \\ \vdots \\ \varrho_{n-2} \\ \varrho_{n-1} \end{bmatrix}, \tag{3.62}$$

where $j = 1, \ldots, m$. In equation (3.62) the coefficients of the transformed system, β_i $i = 1, \ldots, n-1$, and the right hand side of the transformed system, ϱ_i $i = 1, \ldots, n-1$, are achieved by subtracting successively from each equation a suitable multiple of the preceding equation. That is, $\beta_1 = b_1, \beta_2 = b_2 - \frac{a_2}{b_1}c_1, \beta_3 = b_3 - \frac{a_3}{\beta_2}c_2, \ldots, \beta_{n-1} = (b_{n-1} + c_{n-1}) - \frac{a_{n-1}}{\beta_{n-2}}c_{n-2}, \varrho_1 = H_{1,j-1} + dk - a_1 g(jk), \varrho_2 = H_{2,j-1} + dk - \frac{a_2}{b_1}\varrho_1, \varrho_3 = H_{3,j-1} + dk - \frac{a_3}{\beta_2}\varrho_2, \ldots, \varrho_{n-1} = H_{n-1,j-1} + dk - hc_{n-1}f(jk) - \frac{a_{n-1}}{\beta_{n-2}}\varrho_{n-2}$.

With the help of equation system (3.62), approach 2 works iteratively in the following manner:

---

[36]Condition (3.60) is often fulfilled for call option-like contingent claims, and condition (3.61) is often fulfilled for put option-like contingent claims. Consider an American call option as an example. Equation (3.56) indicates that

$$\frac{d\xi^{\tau}(x)}{dx} = 1 \geq \frac{\partial H(x,\tau^{-})}{\partial x}.$$

Thus, condition (3.60) is fulfilled in the case of an American call option.

The last equation in equation system (3.62) gives a way to compute the contingent claim value at grid point $(n-1, j)$ *before* taking the early exercise condition (3.59) into consideration. That is, with the terminology of condition (3.57), it is possible to calculate $H^-_{n-1,j}$ with the help of the last equation in equation system (3.62). Then, insert $H^-_{n-1,j}$ into the early exercise condition (3.59). This gives $H_{n-1,j}$. Next, with $H_{n-1,j}$ it is possible to calculate $H^-_{n-2,j}$ with the help of the last but one equation in equation system (3.62)[37]. Insert $H^-_{n-2,j}$ into condition (3.59). This gives $H_{n-2,j}$. The iterative procedure just described continues until all unknown contingent claim values $H_{n-1,j}, \ldots, H_{1,j}$ are calculated.

Equation (3.58) and assumption (3.60) imply that if a q such that $H^-_{q,j} > \xi^j(qh)$ is found then no decision to exercise early will occur for any $i \leq q$. Thus, the computational burden when using approach 2 with assumption (3.60) fulfilled can be slightly reduced in the following fashion:

Start up with the successive back-substitution and test of the early exercise condition, exactly as previously described. This continues until a $q > 1$ is found such that $H^-_{q,j} > \xi^j(qh)$. Next, since no decision to exercise early will occur for any $i \leq q$, use back-substitution *without* checking the early exercise condition (3.59) to find the solutions to $H_{q-1,j}, \ldots, H_{1,j}$.

Let us now consider approach 2 when condition (3.61) is fulfilled. In this case equa-

[37]It should, however, be noted that $H^-_{i,j}$ $i = 1, \ldots, n-2$ when using approach 2 with condition (3.60) fulfilled need not be equal to $H^-_{i,j}$ $i = 1, \ldots, n-2$ when using approach 1. This follows from the fact that $H^-_{i,j}$ $i = 1, \ldots, n-2$ in approach 2, with condition (3.60) fulfilled, may be affected by "early exercises" at higher levels in the grid (i.e., at higher i values). As will be seen later, when using approach 2 with condition (3.61) fulfilled $H^-_{i,j}$ $i = 2, \ldots, n-1$ can be affected by early exercises at lower levels in the grid (i.e., at lower i values).

tion (3.49) is transformed into

$$\begin{bmatrix} \beta_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ a_2 & \beta_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & a_3 & \beta_3 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a_{n-2} & \beta_{n-2} & 0 \\ 0 & \cdots & \cdots & 0 & 0 & a_{n-1} & \beta_{n-1} \end{bmatrix} \begin{bmatrix} H_{1,j} \\ H_{2,j} \\ H_{3,j} \\ \vdots \\ H_{n-2,j} \\ H_{n-1,j} \end{bmatrix} = \begin{bmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \\ \vdots \\ \varrho_{n-2} \\ \varrho_{n-1} \end{bmatrix}, \tag{3.63}$$

where $j = 1, \ldots, m$. In equation (3.63) the coefficients of the transformed system, β_i $i = 1, \ldots, n-1$, and the right hand side of the transformed system, ϱ_i $i = 1, \ldots, n-1$, are achieved by subtracting successively from each equation a suitable multiple of the succeeding equation. That is, $\beta_{n-1} = (b_{n-1} + c_{n-1}), \beta_{n-2} = b_{n-2} - \frac{c_{n-2}}{(b_{n-1}+c_{n-1})} a_{n-1}, \beta_{n-3} = b_{n-3} - \frac{c_{n-3}}{\beta_{n-2}} a_{n-2}, \ldots, \beta_1 = b_1 - \frac{c_1}{\beta_2} a_2$, $\varrho_{n-1} = H_{n-1,j-1} + dk - hc_{n-1} f(jk)$, $\varrho_{n-2} = H_{n-2,j-1} + dk - \frac{c_{n-2}}{(b_{n-1}+c_{n-1})} \varrho_{n-1}$, $\varrho_{n-3} = H_{n-3,j-1} + dk - \frac{c_{n-3}}{\beta_{n-2}} \varrho_{n-2}, \ldots, \varrho_1 = H_{1,j-1} + dk - a_1 g(jk) - \frac{c_1}{\beta_2} \varrho_2$.

The first equation in equation system (3.63) gives a way to compute the contingent claim value at grid point $(1, j)$ *before* taking the early exercise condition (3.59) into consideration, i.e., $H^-_{1,j}$. Then, insert $H^-_{1,j}$ into the early exercise condition (3.59). This yields $H_{1,j}$. Next, with $H_{1,j}$ it is possible to calculate $H^-_{2,j}$ with the help of the second equation in equation system (3.63). Insert $H^-_{2,j}$ into condition (3.59). This yields $H_{n-2,j}$. This iterative procedure continues until all unknown contingent claim values $H_{1,j}, \ldots, H_{n-1,j}$ have been calculated.

Equation (3.58) and assumption (3.61) imply, of course, that the computational burden when using approach 2 can be slightly reduced also in this case. More precisely, if a q such that $H^-_{q,j} > \xi^j(qh)$ is found then no decision to exercise early will occur for any $i \geq q$. Thus, approach 2 with assumption (3.61) fulfilled can be simplified as follows:

Start up with the successive forward substitution and test of the early exercise condition, exactly as previously described. This continues until a $q < n-1$ is found such that $H^-_{q,j} > \xi^j(qh)$. Then, since no decision to exercise early will occur for any $i \geq q$, use forward substitution *without* checking the early exercise condition (3.59) to find the solutions to

$H_{q+1,j}, \ldots, H_{n-1,j}$.

Approach 2 can be interpreted as an assumption that the owner may take the early exercise condition into consideration at every point in (continuous) time, and an effort is made to approximate this.

Approximation by means of approach 1 and approach 2 ought to give nearly the same result in most cases (at least with a proper specification of the grid). This matter should, however, be examined in a separate investigation.

3.5 A general analysis

As mentioned earlier, the purpose of this paper is to get some knowledge about how errors in the boundary conditions affect the approximate solutions to the value of a contingent claim, when using the implicit finite difference method. The predominant part of this investigation will be through studies of specific cases. Nevertheless, a cursory analytical treatment of the question under investigation will be done in this section.

This analytical treatment is carried out primarily to give an understanding of what is involved rather than as a complete analysis of the problem. That is, the main purpose of this section is to give hints about how to structure and carry out the case studies.

From equation (3.53) it is known that the unknown contingent claim prices $(H_{1,j}, H_{2,j}, \ldots, H_{n-1,j})$ in iteration j can be expressed as

$$\mathbf{AH_j} = \mathbf{HL_j} \qquad\qquad j = 1, \ldots, m,$$

where $\mathbf{A}, \mathbf{H_j}$ and $\mathbf{HL_j}$ are defined in equations (3.50), (3.51) and (3.52).

It has also been concluded that the matrix $\mathbf{A}$ has full rank. That is, equation (3.53) can be rewritten as

$$\mathbf{H_j} = \mathbf{BHL_j} \qquad\qquad j = 1, \ldots, m, \tag{3.64}$$

where $\mathbf{B} = \mathbf{A}^{-1}$. $\mathbf{A}^{-1}$ is the inverse matrix to matrix $\mathbf{A}$.

In order to make the subsequent analysis more tractable, the following vector notations are introduced:

$$\mathbf{C_j} = \begin{bmatrix} a_1 g(jk) \\ 0 \\ 0 \\ \vdots \\ 0 \\ hc_{n-1} f(jk) \end{bmatrix} \qquad j = 1, \ldots, m, \tag{3.65}$$

$$\mathbf{D} = \begin{bmatrix} dk \\ dk \\ dk \\ \vdots \\ dk \\ dk \end{bmatrix}. \tag{3.66}$$

From equations (3.51), (3.52), (3.65) and (3.66) it can be concluded that

$$\mathbf{HL_j} = \mathbf{H_{j-1}} + \mathbf{D} - \mathbf{C_j} \qquad j = 1, \ldots, m. \tag{3.67}$$

With equation (3.67), equation (3.64) can be restated as

$$\mathbf{H_j} = \mathbf{BH_{j-1}} + \mathbf{BD} - \mathbf{BC_j} \qquad j = 1, \ldots, m. \tag{3.68}$$

Repeated use of equation (3.68) yields that the approximations of the contingent claim values in iteration j can be written as

$$\begin{aligned} \mathbf{H_j} = \mathbf{BH_{j-1}} + \mathbf{BD} - \mathbf{BC_j} = \mathbf{B}(\mathbf{BH_{j-2}} + \mathbf{BD} - \mathbf{BC_{j-1}}) + \mathbf{BD} - \mathbf{BC_j} &= \\ \mathbf{B}^2\mathbf{H_{j-2}} + \mathbf{B}^2\mathbf{D} + \mathbf{BD} - \mathbf{B}^2\mathbf{C_{j-1}} - \mathbf{BC_j} = \ldots &= \\ \mathbf{B}^j\mathbf{H_0} + \sum_{l=1}^{j} \mathbf{B}^l\mathbf{D} - \sum_{l=1}^{j} \mathbf{B}^{j+1-l}\mathbf{C_l}& \\ j = 1, \ldots, m.& \end{aligned} \tag{3.69}$$

In equation (3.69) is

$$\mathbf{H_0} = \begin{bmatrix} v(h) \\ v(2h) \\ v(3h) \\ \vdots \\ v([n-2]h) \\ v([n-1]h) \end{bmatrix}, \tag{3.70}$$

and $v(ih)$ $i = 1, \ldots, n-1$ are the values of the contingent claim at maturity. Thus, $\mathbf{H_0}$ is the vector of (correct) initial values.

Define the vector of errors in the boundary conditions at time $\tau = jk$ by

$$\mathbf{\Delta_j} = \begin{bmatrix} a_1 \Delta g_j \\ 0 \\ 0 \\ \vdots \\ 0 \\ h c_{n-1} \Delta f_j \end{bmatrix} \qquad j = 1, \ldots, m, \tag{3.71}$$

where a_1 and c_{n-1} are defined in equation (3.46). Δg_j is the error in the lower boundary condition at time $\tau = jk$. Δf_j is the error in the upper boundary condition at time $\tau = jk$. The subindexes at Δf_j and Δg_j denote that the errors may be of different size at different time steps.

Denote the right hand side of equation (3.53), when there have been errors in the boundary conditions in iterations $1, \ldots, j-1$, by $\mathbf{HL_j^{**}}$. Equations (3.53), (3.64) and (3.71) give

$$\mathbf{H_j^{**}} = \mathbf{BHL_j^{**}} - \mathbf{B\Delta_j} \qquad j = 1, \ldots, m. \tag{3.72}$$

In equation (3.72), $\mathbf{H_j^{**}}$ is the approximate contingent claim values after iteration j, when there have been errors in the boundary conditions in every iteration done.

If equation (3.67) is observed it can be concluded that

$$\mathbf{HL_j^{**}} = \mathbf{H_{j-1}^{**}} + \mathbf{D} - \mathbf{C_j} \qquad j = 1, \ldots, m. \tag{3.73}$$

Applying equation (3.72) and equation (3.73) repeatedly yields

$$\begin{aligned}
\mathbf{H_j^{**}} = \mathbf{BHL_j^{**}} - \mathbf{B\Delta_j} = \mathbf{B(H_{j-1}^{**} + D - C_j)} - \mathbf{B\Delta_j} &= \\
\mathbf{B^2HL_{j-1}^{**}} - \mathbf{B^2\Delta_{j-1}} + \mathbf{BD} - \mathbf{BC_j} - \mathbf{B\Delta_j} &= \\
\mathbf{B^2(H_{j-2}^{**} + D - C_{j-1})} + \mathbf{BD} - \mathbf{BC_j} - \mathbf{B^2\Delta_{j-1}} - \mathbf{B\Delta_j} &= \\
\mathbf{B^2H_{j-2}^{**}} + \mathbf{B^2D} + \mathbf{BD} - \mathbf{B^2C_{j-1}} - \mathbf{BC_j} - \mathbf{B^2\Delta_{j-1}} - \mathbf{B\Delta_j} = \ldots &= \\
\mathbf{B}^j\mathbf{H_0} + \sum_{l=1}^{j}\mathbf{B}^l\mathbf{D} - \sum_{l=1}^{j}\mathbf{B}^{j+1-l}\mathbf{C_l} - \sum_{l=1}^{j}\mathbf{B}^{j+1-l}\mathbf{\Delta_l}& \\
j = 1, \ldots, m.& \qquad (3.74)
\end{aligned}$$

Define the error vector $\mathbf{R}$ by

$$\mathbf{R} = \mathbf{H} - \mathbf{H}^{**}. \qquad (3.75)$$

With equations (3.69), (3.74) and (3.75) the formula for the propagation of the errors in the boundary conditions is

$$\mathbf{R_j} = \mathbf{H_j} - \mathbf{H_j^{**}} \;=\; \sum_{l=1}^{j}\mathbf{B}^{j+1-l}\mathbf{\Delta_l} \qquad j = 1, \ldots, m. \qquad (3.76)$$

Thus, the amplification matrix, $\mathbf{B} = \mathbf{A^{-1}}$, solely determines the propagation of errors in the boundary conditions.

3.6 Investigation of specific cases

It was shown in section 3.5 that the influence from errors in the boundary conditions heavily depends on the properties of the amplification matrix. The amplification matrix, $\mathbf{B}$, is the inverse of matrix $\mathbf{A}$, i.e., $\mathbf{B} = \mathbf{A}^{-1}$. From equations (3.46) and (3.50) it is known that

$$\mathbf{A} \;=\; \begin{bmatrix}
b_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\
a_2 & b_2 & c_2 & 0 & 0 & \cdots & 0 \\
0 & a_3 & b_3 & c_3 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & & & \vdots \\
0 & \cdots & \cdots & 0 & a_{n-2} & b_{n-2} & c_{n-2} \\
0 & \cdots & \cdots & 0 & 0 & a_{n-1} & (b_{n-1} + c_{n-1})
\end{bmatrix},$$

where

$$\begin{aligned}
a_i &= \frac{ki(r-c)}{2} - \frac{\sigma^2 i^2 k}{2}, \\
b_i &= \sigma^2 i^2 k + (1 + rk), \\
c_i &= \frac{ki(c-r)}{2} - \frac{\sigma^2 i^2 k}{2}.
\end{aligned}$$

The elements in the amplification matrix, $\mathbf{B}$, will depend on the same parameters as the elements in the matrix $\mathbf{A}$. Thus, the propagation of errors in the boundary conditions will depend on the same parameters as the elements in matrix $\mathbf{A}$. These parameters are:

- $r =$ the instantaneous rate of return per unit time on the riskless asset. r is assumed to be constant.
- $\sigma^2 =$ the instantaneous variance per unit time of the rate of return on the underlying security. σ^2 is assumed to be constant.
- $c =$ the instantaneous cash payout per unit time to the owners of the underlying security expressed as a yield. c is assumed to be constant.
- $k = \Delta\tau =$ the time step used when approximating the partial differential equation that the value of the contingent claim must satisfy with difference quotients.
- $i =$ the index that denotes the value of the underlying security at the grid point in question, i.e., $x_i = ih \;\; i = 0, 1, \ldots, n$.

Of the aforementioned parameters, the values of r, σ^2 and c are exogenously given, i.e., they are not modelling decisions. Nevertheless, the value of parameter k and the range of i are modelling decisions. It is by these choices that the modeller can give the method the desired accuracy.

Section 3.5 showed that the amplification matrix solely determines the error propagation. Since the elements of the amplification matrix are functions of the parameters discussed above, it is obvious that the influence from errors in the boundary conditions is likely to vary with different values for the parameters. So, for the instruments which are going to be investigated, valuations will be performed with different sets of parameter values.

This paper will only treat two instruments since there are many parameters to vary. The instruments are European call option on a stock and European put option on a stock. The restriction to look at only two instruments may not necessarily be bad. This is because the influence from errors in the boundary conditions depends on the amplification matrix, and the amplification matrices look almost the same in most (standard) option pricing problems.

3.6.1 Description of the instruments

The instruments that are used for the case studies are European call option on a stock and European put option on a stock. These instruments have analytical solutions. This implies that it is possible to compare the solutions from the implicit finite diffence method with the analytical solutions in order to get an apprehension of the accuracy of the numerical method.

3.6.1.1 European call option

The partial differential equation that the price of a European call option on a stock that pays a continuous dividend yield must satisfy looks like

$$\frac{1}{2}H_{xx}(x,\tau)\sigma^2x^2 + (r-c)xH_x(x,\tau) - rH(x,\tau) - H_\tau(x,\tau) = 0, \tag{3.77}$$

subject to

$$\lim_{x\to\infty} H_x(x(\tau),\tau) = e^{-c\tau} \qquad 0 < \tau \leq T, \tag{3.78}$$

$$H(0,\tau) = 0 \qquad 0 < \tau \leq T, \tag{3.79}$$

$$H(x,0) = max[x - Œ, 0], \tag{3.80}$$

where Œ is the exercise price (the remaining notation is explained earlier in the text).

The solution to equation (3.77) subject to conditions (3.78)-(3.80) is

$$H(x,\tau) = x(\tau)e^{-c\tau}\Phi(d1) - Œe^{-r\tau}\Phi(d2), \tag{3.81}$$

where

$$\begin{aligned}
d1 &= \frac{ln(x(\tau)/\text{Œ}) + (r - c + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \\
d2 &= d1 - \sigma\sqrt{\tau},
\end{aligned}$$

and $\Phi(\bullet)$ is the standardized cumulative normal distribution function. Equation (3.81) is the well-known modification of Black & Scholes formula for the case where the underlying stock pays a continuous dividend yield (e.g., see [67] p.135).

The next step, in order to use the implicit finite difference method, is to derive discrete versions of equations (3.77)-(3.80). Using the method described in section 3.4, the following equation is arrived at[38]:

$$a_i H_{i-1,j+1} + b_i H_{i,j+1} + c_i H_{i+1,j+1} = H_{i,j}, \tag{3.82}$$

where

$$\begin{aligned}
a_i &= \frac{ki(r-c)}{2} - \frac{\sigma^2 i^2 k}{2}, \\
b_i &= \sigma^2 i^2 k + (1 + rk), \\
c_i &= \frac{ki(c-r)}{2} - \frac{\sigma^2 i^2 k}{2}, \\
i &= 1, \ldots, n-1, \\
j &= 0, \ldots, m-1,
\end{aligned}$$

[38]When the price process of the underlying stock is modelled as in equation (3.1) [with $D^t(S,t) = cS$ and with c, μ and σ as constants] it is easy to verify that $x \in]0, \infty[$ [This is true if the current stock price $x(T)$ is greater than 0. If $x(T) = 0$ then $x(\tau) = 0$ for all $0 < \tau \leq T$.] (e.g., see [42] or [89]). Furthermore, the price function of the option is defined on the whole interval of x.

To be able to use the implicit finite difference method, we approximate the true domain of the price function in the dimension of x with $x \in [0, nh]$. The maximum value of the index in the stock price dimension, n, must be chosen large enough for the limiting condition on the option price [equation (3.78)] to be well approximated at the approximate upper boundary. The option value at the lower boundary is self-evident, since the properties of the price process of the underlying stock [equation (3.1) with $D^t(S,t) = cS$ and with c, μ and σ as constants] give that if $x(s) = 0$ then $x(\tau) = 0$ for all $\tau > s$.

The discussion above holds equally well when the boundary conditions for a European put option are modelled in sub-subsection 3.6.1.2.

subject to

$$\begin{aligned}
H_{n,j} - H_{n-1,j} &= he^{-cjk} & j = 1, \ldots, m, \quad (3.83)\\
H_{0,j} &= 0 & j = 1, \ldots, m, \quad (3.84)\\
H_{i,0} &= max[ih - Œ, 0] & i = 0, \ldots, n. \quad (3.85)
\end{aligned}$$

3.6.1.2 European put option

The price of a European put option must satisfy partial differential equation (3.77) subject to

$$\begin{aligned}
\lim_{x \to \infty} H_x(x(\tau), \tau) &= 0 & 0 < \tau \leq T, \quad (3.86)\\
H(0, \tau) &= Œe^{-r\tau} & 0 < \tau \leq T, \quad (3.87)\\
H(x, 0) &= max[Œ - x, 0]. & \quad (3.88)
\end{aligned}$$

The solution to equation (3.77) subject to conditions (3.86) - (3.88) is well-known (e.g., see [67] p.135), and is

$$H(x, \tau) = Œe^{-r\tau}\Phi(-d2) - x(\tau)e^{-c\tau}\Phi(-d1), \quad (3.89)$$

where $d1, d2$ and $\Phi(\bullet)$ are defined in equation (3.81).

In order to derive a discrete version of equation (3.77) subject to conditions (3.86)-(3.88), the method described in section 3.4 is used. This gives equation (3.82) subject to

$$\begin{aligned}
H_{n,j} - H_{n-1,j} &= 0 & j = 1, \ldots, m, \quad (3.90)\\
H_{0,j} &= Œe^{-rjk} & j = 1, \ldots, m, \quad (3.91)\\
H_{i,0} &= max[Œ - ih, 0] & i = 0, \ldots, n. \quad (3.92)
\end{aligned}$$

3.6.2 The base case

As a reference there is a base case, for all valuations in this paper. The parameter values in the base case are as follows:

- Option parameters:
  - riskfree rate = r = 0.1
  - volatility = σ = 0.3
  - dividend yield = c = 0.05
  - exercise price = Œ = 50.00
  - current stock price = $x(T)$ = 60.00
  - time to maturity (from t_0) = T = 0.5

  Of the above parameters, the riskfree rate, the volatility and the dividend yield affect the amplification matrix, $\mathbf{B} = \mathbf{A}^{-1}$. Thus, it is these option parameters that are going to be varied. Valuations are performed for riskfree rates in the range from 0.02 to 0.30, for volatilities in the range from 0.05 to 0.75 and for dividend yields in the range from 0.00 to 0.20.

- Grid parameters:
  - $\Delta x = h$ =1.0
  - $\Delta \tau = k$ =0.0025
  - max. index in the price dim. = n = 400
  - max. index in the time dim. = m = 200

  Of the grid parameters, $\Delta\tau$ and the range of the index in the price dimension affect the amplification matrix. Thus, the options are going to be valued for different values of these grid parameters. Valuations are performed for different values of n in the range from 20 to 1600. Since the upper boundary for the stock price is 400 in all valuations this means that the valuations are done for different $\Delta x = h$ in the range from 0.25 to 20. Furthermore, valuations are performed for different values of m in the range from 25 to 800. Since the time to maturity is 0.5 year in all valuations this means that the valuations are done for different $\Delta\tau = k$ in the range from 0.02 to 6.25×10^{-4}.

In every case, when the value of one specific parameter is changed all the other parameters remain as in the base case.

Also, in the case studies the following abbreviations are used:

- B & S-value = the values from the closed-form solutions, i.e., from equations (3.81) and (3.89)[39].
- IMP = the values from the implicit finite difference method without any errors in the boundary conditions.
- IMPERR = the values from the implicit finite difference method with errors in the boundary conditions.
- ERR1 = IMP - B & S-value.
- ERR2 = IMPERR - IMP.

3.6.3 Numerical results

The principle for how errors in the boundary conditions are introduced is the same in all valuations. This principle will be described below.

From equation (3.79) and equation (3.87), it is clear that conditions directly on the option price are used at the lower boundary, for both a European call option and a European put option. When an error is introduced in the lower boundary condition, a constant is added to the "true" option price. Using this approach, it can be established that the discrete lower boundary condition with error will be

$$H_{0,j} = \Delta \qquad\qquad j = 1, 2, \ldots, m \tag{3.93}$$

[39]It can be noted that the standardized cumulative normal distribution function, $\Phi(\bullet)$, in equations (3.81) and (3.89) is also evaluated with the help of numerical approximation. Thus, when option values obtained with the help of equations (3.81) and (3.89) are compared to option values obtained with the help of the implicit finite difference method, values from two different numerical approximation techniques are actually compared to each other.

for a European call option, and

$$H_{0,j} = \text{Œ}e^{-rjk} + \Delta \qquad j = 1, 2, \ldots, m \tag{3.94}$$

for a European put option. Δ is a constant, in equations (3.93) and (3.94).

Equations (3.78) and (3.86) show that the conditions at the upper boundary are conditions on the first derivative of the option price with respect to x, both for a European call option and a European put option. The errors in the upper boundary conditions are introduced by adding a constant to the "true" value of the derivatives. This approach indicates that the discrete versions of the upper boundary conditions become

$$H_{n,j} - H_{n-1,j} = h(e^{-cjk} + \Delta) \qquad j = 1, 2, \ldots, m \tag{3.95}$$

for a European call option, and

$$H_{n,j} - H_{n-1,j} = h\Delta \qquad j = 1, 2, \ldots, m \tag{3.96}$$

for a European put option. Δ is a constant, in equations (3.95) and (3.96).

3.6.3.1 European call option, errors in the boundary conditions

As a first study of how errors in the boundary conditions affect the solutions, errors are introduced in both the lower and upper boundary conditions at the same time. In this first study the constant error is set to be equal to unity, that is $\Delta = 1$. With this choice of Δ several valuations are performed, with different values of the relevant parameters. Whenever one parameter is varied, the other parameters remain as in the base case (see subsection 3.6.2). Valuations are performed for parameter values in the ranges described in 3.6.2. The results from these valuations are presented in table 3.2.

Before the effect from the errors in the boundary conditions is analysed, it can be interesting to observe two facts. Firstly, from the column for ERR1 in table 3.2 it can be concluded that the implicit finite difference method has good accuracy in all valuations except where the grid is very coarse, i.e., where n or m is small. Secondly, from the same column it can also be observed that solutions from the implicit finite difference method converge to the closed-form solutions when n or m increases.

The column for ERR2 in table 3.2 is now referred to. In this column it can be seen that the introduction of errors with $\Delta = 1$ in both the upper and lower boundary condition does not give any errors in the solutions (At any rate, not sufficiently significant to affect the fourth decimal of ERR2.). This is despite the fact that valuations are performed for very wide ranges of parameter values.

There is, however, one exception. It occurs when the volatility is as high as 0.75, and with this high volatility ERR2 is equal to 0.0083 (a more thorough discussion of this high value of ERR2 is given a little further on). Certainly, 0.0083 is a small number but compared to 0.0000 it is large. The fact that ERR2 when $\sigma = 0.75$ is much larger than ERR2 for any other set of parameter values implies that there exist some unfortunate sets of parameter values that give much less reduction of the effect from errors in the boundary conditions than usual.

Reasonably, the errors in the solutions that are due to errors in the boundary conditions will be larger the nearer the current stock price, $x(T)$, is to one of the boundaries. Table 3.3, which shows ERR2 for different values of $x(T)$ (the rest of the parameters are as in the base case, and $\Delta = 1$ for both boundary conditions), confirms this. Also, table 3.3 shows that the reduction of the effect from errors in the upper boundary condition is much slower than the reduction of the effect from errors in the lower boundary condition[40].

In order to offer a better understanding of how errors in the boundary conditions affect the solutions, the effects from errors in the lower boundary condition and the effects from errors in the upper boundary condition will be studied separately from now on. However, the manner in which the errors are introduced will still be described by equations (3.93)-(3.96).

More precisely, what will be investigated is how the effect from errors in one of the boundary conditions is reduced (or increased) as the distance (in the price dimension) from the boundary under investigation is increased, for the current time ($j = m$). In order to

[40]The specification of the grid used (described in subsection 3.6.2) would certainly be very inappropriate if the current stock price, $x(T)$, was equal to 360 (or 300). Instead, the cut off value (nh) should be chosen much higher (than the 400 used) when $x(T) = 360$ [or $x(T) = 300$].

Riskfree rates in the range 0.02-0.30

r	B & S	IMP	IMPERR	ERR1	ERR2
0.02	10.4113	10.4087	10.4087	-0.0026	0.0000
0.06	11.1633	11.1603	11.1603	-0.0030	0.0000
0.10	11.9270	11.9237	11.9237	-0.0033	0.0000
0.12	12.3125	12.3090	12.3090	-0.0035	0.0000
0.20	13.8706	13.8664	13.8664	-0.0043	0.0000
0.30	15.8321	15.8268	15.8268	-0.0053	0.0000

Volatilities in the range 0.05-0.75

σ	B & S	IMP	IMPERR	ERR1	ERR2
0.05	10.9571	10.9569	10.9569	-0.0002	0.0000
0.15	11.0107	11.0103	11.0103	-0.0004	0.0000
0.30	11.9270	11.9237	11.92371	-0.0033	0.0000
0.50	14.1467	14.1409	14.1409	-0.0058	0.0000
0.75	17.3840	17.3755	17.3838	-0.0085	0.0083

Dividend yields in the range 0.00-0.20

c	B & S	IMP	IMPERR	ERR1	ERR2
0.00	13.2207	13.2172	13.2172	-0.0035	0.0000
0.02	12.6950	12.6915	12.6915	-0.0035	0.0000
0.06	11.6766	11.6733	11.6733	-0.0033	0.0000
0.12	10.2335	10.2308	10.2308	-0.0027	0.0000
0.20	8.4724	8.4707	8.4707	-0.0016	0.0000

Max ind. price dim. range from 20 - 1600

n	B & S	IMP	IMPERR	ERR1	ERR2
20	11.9270	12.2729	12.2729	0.3458	0.0000
80	11.9270	11.8736	11.8736	-0.0534	0.0000
200	11.9270	11.9173	11.9173	-0.0097	0.0000
400	11.9270	11.9237	11.9237	-0.0033	0.0000
1600	11.9270	11.9257	11.9257	-0.0013	0.0000

Max ind. time dim. range from 25 - 800

m	B & S	IMP	IMPERR	ERR1	ERR2
25	11.9270	11.9155	11.9155	-0.0115	0.0000
50	11.9270	11.9201	11.9201	-0.0069	0.0000
100	11.9270	11.9225	11.9225	-0.0045	0.0000
200	11.9270	11.9237	11.9237	-0.0033	0.0000
800	11.9270	11.9246	11.9246	-0.0024	0.0000

Table 3.2: Error in the boundary conditions, European call option.

Note: Explanations to the notation in table 3.2 are given in subsection 3.6.2

Current stock prices in the range 20.0-360.0						
	$x(T) = 20.0$	$x(T) = 60.0$	$x(T) = 100.0$	$x(T) = 200.0$	$x(T) = 300$	$x(T) = 360$
ERR2	0.0000	0.0000	0.0000	0.0142	5.0060	28.0880

Table 3.3: The effect from errors in the boundary conditions, for different values of $x(T)$.

be able to perform the desired study, the following reduction factors are introduced:

- Reduction factors, lower boundary:

$$\begin{aligned} REDL_1 &= \frac{\Delta}{H^e_{1,m} - H_{1,m}}. \\ REDL_i &= \frac{H^e_{i-1,m} - H_{i-1,m}}{H^e_{i,m} - H_{i,m}} \qquad i = 2, \ldots, n. \end{aligned} \tag{3.97}$$

$$MULL_i = \prod_{l=1}^{i} REDL_l = \frac{\Delta}{H^e_{i,m} - H_{i,m}} \qquad i = 1, \ldots, n. \tag{3.98}$$

- Reduction factors, upper boundary:

$$REDU_i = \frac{H^e_{i+10,m} - H_{i+10,m}}{H^e_{i,m} - H_{i,m}} \qquad i \in \Omega. \tag{3.99}$$

$$MULU_i = \prod_{l=i, l \in \Omega}^{n} REDU_l = \frac{H^e_{n,m} - H_{n,m}}{H^e_{i,m} - H_{i,m}} \qquad i \in \Omega, \tag{3.100}$$

where

$$\Omega = \Big\{ i \in \{1, 2, \ldots, n-10\} : \frac{(n-10) - i}{10} = integer \Big\}.$$

In equations (3.97) - (3.100) denote:

- $H_{i,m}$ $i = 1, \ldots, n$ are the option values at current time from the implicit finite difference method without errors in the boundary conditions.

- $H^e_{i,m}$ $i = 1, \ldots, n$ are the option values at current time from the implicit finite difference method with errors in the boundary condition under investigation.

As is clear from equation (3.99), $REDU_{n-10} = \Delta/(H^e_{n-10,m} - H_{n-10,m})$ is not used. The reason for this is that a derivative condition is used at the upper boundary. The successive use of an error equal to Δ in the value of the derivative leads to a $(H^e_{n,m} - H_{n,m})$ rather different than Δ.

From equation (3.97) it can be seen that $REDL_i$ gives a measure of how much the effect from errors in the lower boundary condition decreases (or increases) when moving one step further away from the lower boundary, when $j = m$. $MULL_i$ gives a measure of the total reduction of the effect from errors in the lower bondary condition i steps away from the lower boundary, when $j = m$. Equation (3.99) establishes that $REDU_i$ is a measure of the reduction of the effect from errors in the upper boundary condition when moving 10 steps further away from the upper boundary, when $j = m$. Due to the fact that the error reduction is much slower for errors in the upper boundary condition than the error reduction for errors in the lower boundary condition, an interval of 10 steps has been chosen when analysing the reduction of the effect from errors in the upper boundary condition[41]. $MULU_i$ is a measure of the total reduction of the effect from errors in the upper boundary condition $n - i$ grid points away from the upper boundary, when $j = m$.

The general analysis in section 3.5 shows that the propagation of errors in the boundary conditions only depends on the amplification matrix, $\mathbf{B} = \mathbf{A}^{-1}$. Also, $REDL_i$ and $REDU_i$ are relative measures. These facts imply that $REDL_i, MULL_i, REDU_i$ and $MULU_i$ get the same numerical values irrespective of what numerical value is chosen for Δ. The remaining parameter values have, of course, to be the same.

As a reference case, the reduction factors are computed for parameter values as in the base case [except for $x(T)$]. The base case is described in subsection 3.6.2. The results from these computations are given in table 3.4.

Lower boundary, reference case

In the column for $REDL_i$ in table 3.4, $REDL_i$ is large for low values of $x(T)$, and then declines rather fast as $x(T)$ increases. That is, $REDL_i$ decreases with the distance from the lower boundary.

[41]There is, however, one exception from this. The exception is when $n = 20$. When n is equal to 20 then Δx is equal to 20. So, a step of 10 grid points in the price dimension is the same as a step of $10 \times \Delta x = 200$ price units. Since the whole grid in the price dimension is 400 price units, this step is far too large. Thus, when $n = 20$ the step for the calculation of $REDU_i$ (and, of course, also $MULU_i$) is only one grid point.

In the column for $MULL_i$ in table 3.4, $MULL_i$ increases rapidly with increasing $x(T)$ even though $REDL_i$ decreases rather fast. This means that the total reduction of the effect from errors in the lower boundary condition is very large only a few grid points away from the lower boundary. An intuitive explanation of this is as follows:

When the price process of the underlying asset is modelled as a geometric Brownian motion (as is the case in the valuations in this paper), the probability for the price to be very close to zero is small. Thus, due to their low probabilities, the values of the contingent claim when the price of the underlying asset is close to zero have little influence on the current contingent claim price. Actually, when modelling the price process of the underlying asset as a geometric Brownian motion, the probability for the price to be equal to zero (which is the lower boundary for the price of the underlying asset in the numerical valuations in this paper) is zero [at least, if $S(t_0) = x(T) > 0$]. Hence, this fact provides an intuitive explanation of the observation that the total reduction of the effect from errors in the lower boundary condition is very large only a few grid points away from the lower boundary.

Upper boundary, reference case

The column for $REDU_i$ in table 3.4 indicates that the reduction of the effect from errors in the upper boundary condition is rather slow. It is in any case much slower than the reduction of the effect from errors in the lower boundary condition[42]. Also, in contrast to $REDL_i$, $REDU_i$ increases with the distance from the upper boundary.

The total reduction of the effect from errors in the upper boundary, $MULU_i$, is small

[42] As earlier mentioned, matrix A is strictly diagonally dominant (for all realistic economic applications). Also, it is known that Gauss elimination (with no use of pivoting) applied to diagonally dominant matrices does not lead to any growth in the matrix elements (see [49]).

It seems that the diagonal dominance of **A** and the relation between the diagonal element and the elements beside the diagonal element is also important for the error reduction. Near the lower boundary where the error reduction is very quick, the relative diagonal dominance, $\frac{|b_i|}{(|a_i|+|c_i|)}$, is very large. The relative diagonal dominance is, however, near unity (that, is very small) near the upper boundary, where the error reduction is slow.

near the upper boundary, but increases rather fast when moving away from the upper boundary, as can be seen from table 3.4. This circumstance explains, at least partially, the high values for ERR2 for high values of $x(T)$ in table 3.3. $MULU_i$ is, however, very large for reasonable values of $x(T)$ [This is due to the fact that "the real" $x(T)$ will be somewhere far from the upper boundary with a sensible specification of the grid.].

An intuitive explanation for $MULU_i$ rather rapidly increasing when moving away from the upper boundary is as follows:

When modelling the price process of the underlying asset as a geometric Brownian motion (as is the case in the valuations in this paper) the probability for the price to be very high is small. Thus, if the upper boundary of the grid is specified to be large enough (in the dimension of the price of the underlying asset), the probability for the price of the underlying asset to reach the upper boundary is small. Thus, due to their low probabilities the values of the contingent claim at the upper boundary have small influence when calculating the current contingent claim price. Hence, even large errors in the upper boundary condition generally do not have any noticeable effects on the solution.

One peculiar feature of table 3.2 is the value for ERR2 when the volatility is equal to 0.75. This peculiarity suggests that the reduction factors may vary a lot with different sets of parameter values. To investigate this, the reduction factors are computed for different values of the parameters. As usual, when one parameter is altered the values of the remaining parameters are as in the base case (see subsection 3.6.2).

Lower boundary

Look at tables 3.4, 3.5 and 3.6. The absolute value of the total reduction factor [for given values of $x(T)$], $MULL_i$, seems to be affected by parameter changes in the following way:

- The absolute value of $MULL_i$ seems to increase as the riskfree interest rate increases.
- When the volatility increases, the absolute value of $MULL_i$ decreases.
- The absolute value of $MULL_i$ decreases with higher values of the dividend yield.

Reduction factors for the lower and the upper boundary						
Lower boundary				Upper boundary		
$x(T)$	$REDL_i$	$MULL_i$		$x(T)$	$REDU_i$	$MULU_i$
1.0	104.8108	104.8		390.0	1.1673	1.167
2.0	32.4962	3.406×10^{3}		380.0	1.1820	1.380
3.0	19.5763	6.668×10^{4}		370.0	1.1984	1.654
4.0	14.1904	9.462×10^{5}		360.0	1.2166	2.012
5.0	11.2362	1.063×10^{7}		350.0	1.2370	2.489
6.0	9.3698	9.961×10^{7}		340.0	1.2598	3.135
7.0	8.0837	8.052×10^{8}		330.0	1.2853	4.029
8.0	7.1453	5.752×10^{9}		320.0	1.3140	5.295
9.0	6.4261	3.696×10^{10}		310.0	1.3464	7.129
10.0	5.8607	2.166×10^{11}		300.0	1.3829	9.858
11.0	5.4035	1.171×10^{12}		290.0	1.4244	14.04
12.0	5.0262	5.884×10^{12}		280.0	1.4717	20.67
13.0	4.7094	2.771×10^{13}		270.0	1.5259	31.53
14.0	4.4397	1.230×10^{14}		260.0	1.5883	50.09
15.0	4.2072	5.176×10^{14}		250.0	1.6605	83.17
16.0	4.0048	2.073×10^{15}		240.0	1.7448	145.1
17.0	3.8270	7.932×10^{15}		230.0	1.8440	267.6
18.0	3.6694	2.911×10^{16}		220.0	1.9616	524.9
19.0	3.5289	1.027×10^{17}		210.0	2.1026	1.104×10^{3}
20.0	3.4028	3.495×10^{17}		200.0	2.2736	2.509×10^{3}
21.0	3.2890	1.150×10^{18}		190.0	2.4839	6.232×10^{3}
22.0	3.1858	3.662×10^{18}		180.0	2.7465	1.712×10^{4}
23.0	3.0917	1.132×10^{19}		170.0	3.0807	5.273×10^{4}
24.0	3.0056	3.403×10^{19}		160.0	3.5152	1.854×10^{5}
25.0	2.9265	9.959×10^{19}		150.0	4.0947	7.590×10^{5}
26.0	2.8535	2.842×10^{20}		140.0	4.8915	3.713×10^{6}
27.0	2.7861	7.918×10^{20}		130.0	6.0286	2.238×10^{7}
28.0	2.7235	2.156×10^{21}		120.0	7.7265	1.729×10^{8}
29.0	2.6654	5.748×10^{21}		110.0	10.4090	1.800×10^{9}
30.0	2.6111	1.501×10^{22}		100.0	14.9618	2.693×10^{10}
60.0		7.591×10^{31}		60.0		5.065×10^{17}

Table 3.4: Reduction factors: European call option, base case.

- Different values of n give identical values for both $REDL_i$ and $MULL_i$, when the error reduction between grid points with same value of index i is studied. This follows from the general analysis in section 3.5, which showed that the amplification matrix solely determines the error propagation. Therefore, since the matrix elements of the amplification matrix do not depend on Δx it does not matter what n is as long as the values for index i are the same[43]. The values for index i are the same when the reduction of the effect from errors in the lower boundary condition is investigated. If $MULL_i$ is, however, observed for a given value of $x(T)$ it can be seen that $MULL_i$ increases with higher values of n.
- The absolute value of $MULL_i$ increases as the value of m increases.

Even if the absolute value of $MULL_i$ changes for all changes in the parameter values the variations in $MULL_i$ are only large for variations in the volatility. The absolute value of $MULL_i$ decreases fast when the volatility increases. But, even in the case where the volatility is as high as 0.75, $MULL_i$ fast becomes very large when moving away from the lower boundary.

In tables 3.4, 3.5 and 3.6, it is also interesting to observe that $REDL_i$ immediately starts to decrease as $x(T)$ increases, in most cases. In the case where $r = 0.30$ and in the case where $c = 0.00$ $REDL_i$ is, however, negative with increasing absolute value for low values of $x(T)$. But, when the value of $x(T)$ increases then $REDL_i$ becomes positive and starts to decrease. Moreover, in the case where $\sigma = 0.05$, $REDL_i$ is negative with increasing absolute value for the whole range of $x(T)$ in table 3.5. But, also in this case, $REDL_i$ will become positive with decreasing absolute value as the value of $x(T)$ becomes higher[44].

[43]This argumentation is, of course, not completely true since all equations in equation system (3.53) are interconnected. However, the tridiagonal structure of the matrix $\mathbf{A}$ and the values of a_i, b_i and c_i make the argumentation almost true.

[44]Look at the expressions for a_i, b_i and c_i in equation (3.46). For all possible sets of parameter values b_i is positive and increases with increasing values of the index i. Also, in all valuations where the reduction factors are calculated, c_i is negative for all values of i (Except for the case where $c = 0.2$. In this case, c_1

Reduction factors for the lower boundary					
Riskfree rate = 0.02			Riskfree rate = 0.30		
$x(T)$	$REDL_i$	$MULL_i$	$x(T)$	$REDL_i$	$MULL_i$
1.0	34.2543	34.25	1.0	-27.5415	-27.54
2.0	19.9834	684.5	2.0	-61.4400	1.692×10^{3}
3.0	14.3079	9.794×10^{3}	3.0	217.4121	3.679×10^{5}
4.0	11.2566	1.102×10^{5}	4.0	40.2403	1.480×10^{7}
5.0	9.3506	1.031×10^{6}	5.0	22.5740	3.342×10^{8}
6.0	8.0467	8.295×10^{6}	6.0	15.8879	5.310×10^{9}
7.0	7.0984	5.888×10^{7}	7.0	12.3743	6.570×10^{10}
8.0	6.3776	3.755×10^{8}	8.0	10.2087	6.707×10^{11}
9.0	5.8112	2.182×10^{9}	9.0	8.7403	5.862×10^{12}
10.0	5.3543	1.168×10^{10}	10.0	7.6791	4.502×10^{13}
60.0		5.630×10^{29}	60.0		1.119×10^{36}
Volatility = 0.05			Volatility = 0.75		
$x(T)$	$REDL_i$	$MULL_i$	$x(T)$	$REDL_i$	$MULL_i$
1.0	-86.4019	-86.40	1.0	9.0774	9.077
2.0	-89.3442	7.719×10^{3}	2.0	4.9021	44.50
3.0	-93.8216	-7.243×10^{5}	3.0	3.6171	161.0
4.0	-99.1154	7.178×10^{7}	4.0	2.9887	481.1
5.0	-105.2099	-7.552×10^{9}	5.0	2.6142	1.258×10^{3}
6.0	-112.2218	8.476×10^{11}	6.0	2.3647	2.974×10^{3}
7.0	-120.3406	-1.020×10^{14}	7.0	2.1859	6.501×10^{3}
8.0	-129.8323	1.324×10^{16}	8.0	2.0513	1.333×10^{4}
9.0	-141.0655	-1.868×10^{18}	9.0	1.9460	2.595×10^{4}
10.0	-154.5590	2.887×10^{20}	10.0	1.8613	4.830×10^{4}
60.0		5.314×10^{130}	60.0		8.534×10^{10}
Dividend yield = 0.00			Dividend yield = 0.20		
$x(T)$	$REDL_i$	$MULL_i$	$x(T)$	$REDL_i$	$MULL_i$
1.0	-419.2686	-419.2	1.0	22.0747	22.07
2.0	52.8095	-2.214×10^{4}	2.0	15.0938	333.2
3.0	25.3355	-5.610×10^{5}	3.0	11.6447	3.880×10^{3}
4.0	16.9202	-9.492×10^{6}	4.0	9.5668	3.712×10^{4}
5.0	12.8420	-1.219×10^{8}	5.0	8.1749	3.034×10^{5}
6.0	10.4351	-1.272×10^{9}	6.0	7.1765	2.178×10^{6}
7.0	8.8468	-1.125×10^{10}	7.0	6.4250	1.399×10^{7}
8.0	7.7200	-8.687×10^{10}	8.0	5.8389	8.169×10^{7}
9.0	6.8790	-5.976×10^{11}	9.0	5.3688	4.386×10^{8}
10.0	6.2273	-3.721×10^{12}	10.0	4.9833	2.186×10^{9}
60.0		-3.590×10^{33}	60.0		4.330×10^{28}

Table 3.5: Reduction factors: European call option, lower boundary.

Reduction factors for the lower boundary						
$n = 20$				$n = 1600$		
$x(T)$	$REDL_i$	$MULL_i$		$x(T)$	$REDL_i$	$MULL_i$
20.0	104.8108	104.8		0.25	104.8108	104.8
40.0	32.4962	3.406×10^{3}		0.5	32.4962	3.406×10^{3}
60.0	19.5763	6.668×10^{4}		0.75	19.5763	6.668×10^{4}
80.0	14.1904	9.462×10^{5}		1.0	14.1904	9.462×10^{5}
100.0	11.2362	1.063×10^{7}		1.25	11.2362	1.063×10^{7}
120.0	9.3698	9.961×10^{7}		1.5	9.3698	9.961×10^{7}
140.0	8.0837	8.052×10^{8}		1.75	8.0837	8.052×10^{8}
160.0	7.1435	5.752×10^{9}		2.0	7.1435	5.752×10^{9}
180.0	6.4261	3.696×10^{10}		2.25	6.4261	3.696×10^{10}
200.0	5.8607	2.166×10^{11}		2.5	5.8607	2.166×10^{11}
60.0		6.668×10^{4}		60.0		3.729×10^{57}
$m = 25$				$m = 800$		
$x(T)$	$REDL_i$	$MULL_i$		$x(T)$	$REDL_i$	$MULL_i$
1.0	104.9731	105.0		1.0	104.7934	104.8
2.0	31.5712	3.314×10^{3}		2.0	32.5992	3.416×10^{3}
3.0	18.5129	6.135×10^{4}		3.0	19.6990	6.730×10^{4}
4.0	13.1062	8.041×10^{5}		4.0	14.3198	9.637×10^{5}
5.0	10.1659	8.175×10^{6}		5.0	11.3682	1.095×10^{7}
6.0	8.3263	6.806×10^{7}		6.0	9.5027	1.041×10^{8}
7.0	7.0720	4.813×10^{8}		7.0	8.2164	8.553×10^{8}
8.0	6.1651	2.968×10^{9}		8.0	7.2756	6.223×10^{9}
9.0	5.4809	1.626×10^{10}		9.0	6.5573	4.081×10^{10}
10.0	4.9478	8.047×10^{10}		10.0	5.9907	2.445×10^{11}
60.0		7.743×10^{26}		60.0		6.128×10^{32}

Table 3.6: Reduction factors: European call option, lower boundary.

To summarize, the absolute value of the total reduction of the effect from errors in the lower boundary condition, $MULL_i$, varies when the parameter values are changed. It is, however, only changes in the volatility that cause $MULL_i$ to vary a lot.

Moreover, in all cases the absolute value of $MULL_i$ quickly becomes very large, when moving away from the lower boundary. If errors in the lower boundary values are to have any perceivable impact on the solution then the errors have to be very large. As a matter of fact, the absolute values of $MULL_i$ are so large, for sensible values of $x(T)$ (a proper specification of the grid does not have the real current stock price at a grid point close to the lower boundary), that almost any value can be applied to the lower boundary. This confirms the observations made by Peter Jennergren when examining solutions to problems handed in by students, as was discussed in section 3.1.

Upper boundary

From the columns for $REDU_i$ in tables 3.4, 3.7 and 3.8, it can be seen that $REDU_i$ increases when moving further away from the upper boundary, in all cases.

It is, however, more interesting to observe the behaviour of the total reduction of the effect from errors in the upper boundary condition, $MULU_i$. From the columns for $MULU_i$ in tables 3.4, 3.7 and 3.8, it can be concluded that [for given values of $x(T)$]:

- $MULU_i$ decreases when the riskfree interest rate increases.
- $MULU_i$ decreases when the volatility increases.

has a *very* small positive value, but is negative for all other values of i.), and its absolute value increases with increasing values of i. Moreover, in all valuations where the reduction factors are calculated, except for three cases, a_i is negative for all values of i, and its absolute value increases with increasing values of i. These three exceptions are exactly the three cases where $REDL_i$ is negative with increasing absolute value for low values of $x(T)$. In the cases where $r = 0.30$, where $\sigma = 0.05$ and where $c = 0.00$ the value of a_i is positive for low values of the index i. As i becomes larger the term $-\sigma^2 i^2 k/2$ will, however, get a higher absolute value than the term $ki(r - c)/2$, and a_i becomes negative with increasing absolute value. When this happens, $REDL_i$ becomes positive with decreasing value as i increases even further, also in these cases. a_i becomes negative when $i = 3$ in the case where $r = 0.30$, when $i = 21$ in the case where $\sigma = 0.05$ and when $i = 2$ in the case where $c = 0.00$.

- $MULU_i$ increases when the dividend yield increases.
- The values of $MULU_i$ are not comparable for different values of n. But $MULU_i$ increases for increasing values of n, for a given value of $x(T)$.
- $MULU_i$ increases [except for very high values of $x(T)$] when the value of m increases.

Thus, $MULU_i$ varies with all changes of the parameter values. Yet, it is only changes in the volatility that cause the values of $MULU_i$ to change considerably. When σ increases the values of $MULU_i$ quickly decrease. Nevertheless, also in the worst case, i.e., when $\sigma = 0.75$, $MULU_i$ increases fairly rapidly with the distance from the upper boundary. As an example (from the worst case, $\sigma = 0.75$), $MULU_i = 1.744 \times 10^4$ when $x(T) = 60.0$.

If tables 3.5 and 3.6 are compared with tables 3.7 and 3.8, it can be concluded that the reduction of the effect from errors in the upper boundary condition is much slower than the reduction of the effect from errors in the lower boundary condition. Still, the upper boundary has one more bad property. This property arises from the use of a derivative condition.

The successive use of an error equal to Δ in the value of the derivative at the upper boundary leads to $(H^e_{n,m} - H_{n,m})$ becoming much larger than Δ. In the base case it can be established that

$$REL = \frac{H^e_{n,m} - H_{n,m}}{\Delta} = 67.45. \tag{3.101}$$

That is, $(H^e_{n,m} - H_{n,m})$ is 67.45 times higher than the constant error in the value of the derivatives, for the base case. The values of REL in all other cases are given in tables 3.7 and 3.8. From these tables, it can be observed that REL is fairly constant and rather high, between 50.0 and 90.0, for most sets of parameter values. There are, however, two exceptions. Both of these occur when the volatility is varied. Low volatility gives a small value of REL (σ=0.05 gives REL= 17.18), and high volatility gives a high value of REL (σ=0.75 gives REL= 145.4). Thus, changes in the volatility cause REL to vary a lot. REL increases quickly as the volatility increases. That is, high volatilities lead to two bad properties, both a high value of REL and a slow reduction of the effect from errors in the boundary conditions.

Reduction factors for the upper boundary					
Riskfree rate = 0.02			Riskfree rate = 0.30		
$REL = 60.75$			$REL = 86.23$		
$x(T)$	$REDU_i$	$MULU_i$	$x(T)$	$REDU_i$	$MULU_i$
390.0	1.1872	1.187	390.0	1.1286	1.129
380.0	1.2031	1.428	380.0	1.1404	1.287
370.0	1.2207	1.744	370.0	1.1535	1.485
360.0	1.2404	2.163	360.0	1.1683	1.735
350.0	1.2622	2.730	350.0	1.1850	2.056
340.0	1.2866	3.512	340.0	1.2037	2.474
330.0	1.3139	4.615	330.0	1.2248	3.030
320.0	1.3446	6.205	320.0	1.2486	3.783
310.0	1.3791	8.557	310.0	1.2756	4.826
300.0	1.4181	12.13	300.0	1.3062	6.304
60.0		2.452×10^{18}	60.0		1.160×10^{16}
Volatility = 0.05			Volatility = 0.75		
$REL = 17.18$			$REL = 145.4$		
$x(T)$	$REDU_i$	$MULU_i$	$x(T)$	$REDU_i$	$MULU_i$
390.0	2.0514	2.051	390.0	1.0729	1.073
380.0	2.8286	5.803	380.0	1.0763	1.155
370.0	4.2418	24.61	370.0	1.0800	1.247
360.0	6.8241	168.0	360.0	1.0840	1.352
350.0	11.6094	1.950×10^{3}	350.0	1.0883	1.471
340.0	20.6301	4.023×10^{4}	340.0	1.0929	1.608
330.0	37.9293	1.526×10^{6}	330.0	1.0979	1.765
320.0	71.6421	1.093×10^{8}	320.0	1.1034	1.948
310.0	138.3355	1.512×10^{10}	310.0	1.1093	2.161
300.0	272.1905	4.116×10^{12}	300.0	1.1158	2.411
60.0		9.332×10^{194}	60.0		1.744×10^{4}
Dividend yield = 0.00			Dividend yield = 0.20		
$REL = 72.51$			$REL = 54.11$		
$x(T)$	$REDU_i$	$MULU_i$	$x(T)$	$REDU_i$	$MULU_i$
390.0	1.1549	1.155	390.0	1.2120	1.212
380.0	1.1689	1.350	380.0	1.2287	1.489
370.0	1.1846	1.599	370.0	1.2471	1.857
360.0	1.2020	1.922	360.0	1.2677	2.354
350.0	1.2215	2.348	350.0	1.2906	3.038
340.0	1.2433	2.919	340.0	1.3162	3.999
330.0	1.2677	3.700	330.0	1.3449	5.378
320.0	1.2952	4.793	320.0	1.3771	7.406
310.0	1.3263	6.357	310.0	1.4134	10.47
300.0	1.3614	8.654	300.0	1.4544	15.22
60.0		1.889×10^{17}	60.0		1.073×10^{19}

Table 3.7: Reduction factors: European call option, upper boundary.

Reduction factors for the upper boundary					
$n = 20$			$n = 1600$		
$REL = 75.74$			$REL = 67.14$		
$x(T)$	$REDU_i$	$MULU_i$	$x(T)$	$REDU_i$	$MULU_i$
380.0	1.3588	1.359	397.5	1.0383	1.038
360.0	1.4327	1.947	395.0	1.0391	1.079
340.0	1.5267	2.972	392.5	1.0399	1.122
320.0	1.6475	4.896	390.0	1.0407	1.168
300.0	1.8041	8.833	387.5	1.0415	1.216
280.0	2.0093	17.75	385.0	1.0423	1.268
260.0	2.2814	40.49	382.5	1.0432	1.322
240.0	2.6471	107.2	380.0	1.0441	1.381
220.0	3.1462	337.2	377.5	1.0450	1.443
200.0	3.8400	1.295×10^{3}	375.0	1.0459	1.509
60.0		1.054×10^{11}	60.0		6.567×10^{17}
$m = 25$			$m = 800$		
$REL = 67.14$			$REL = 67.48$		
$x(T)$	$REDU_i$	$MULU_i$	$x(T)$	$REDU_i$	$MULU_i$
390.0	1.1681	1.168	390.0	1.1672	1.167
380.0	1.1827	1.382	380.0	1.1819	1.380
370.0	1.1989	1.656	370.0	1.1983	1.653
360.0	1.2168	2.015	360.0	1.2166	2.011
350.0	1.2366	2.492	350.0	1.2371	2.488
340.0	1.2585	3.136	340.0	1.2599	3.135
330.0	1.2828	4.023	330.0	1.2856	4.030
320.0	1.3098	5.269	320.0	1.3145	5.297
310.0	1.3398	7.059	310.0	1.3471	7.136
300.0	1.3732	9.694	300.0	1.3840	9.876
60.0		4.252×10^{13}	60.0		4.710×10^{18}

Table 3.8: Reduction factors: European call option, upper boundary.

The fact that the value of REL is rather large in all cases motivates the following advice:

In valuations where the value of the instrument at the upper boundary is fairly obvious, that value should be used rather than a derivative condition, when the implicit finite difference method is used. Suggestions for conditions on the value at the upper boundary that can be used when valuing some common contingent claim types are given in appendix BIII.

The strange value of ERR2 in table 3.2 is now referred to. This value is ERR2 = 0.0083 for the case where $\sigma = 0.75$ [the values of all the other parameters are as in the base case (see subsection 3.6.2)]. The reason for the high value of ERR2, in this case, is the combined effect of a high value of REL and a slow reduction of the effect from errors in the upper boundary condition. The following formula can be used to calculate ERR2[45] [for $x(T) = 60.0$]:

$$ERR2 = \frac{\Delta \times REL}{MULU_{60}}. \tag{3.102}$$

Use equation (3.102) to calculate ERR2 in the case where $\sigma = 0.75$. The values to use are $\Delta = 1, REL = 145.4$ and $MULU_{60} = 1.744 \times 10^4$. Plug these values into equation (3.102) and

$$ERR2 = 1 \times 145.4/(1.744 \times 10^4) = 0.0083,$$

which is equal to the strange value in table 3.2.

To summarize, the total reduction factor, $MULU_i$, varies with different sets of parameter values. It is, however, only variations in the volatility that cause $MULU_i$ to change a lot. $MULU_i$ decreases quickly when the volatility increases. Yet $MULU_i$ is large for reasonable values of $x(T)$, in all cases.

The derivative condition at the upper boundary has a bad property. The successive

[45] It should be noted that equation (3.102) is not completely true. When the values in table 3.2 were computed there were errors equal to unity, $\Delta = 1$, in both the upper and lower boundary condition. The quick reduction of the effect from errors in the lower boundary condition implies, however, that the impact from these errors can be ignored when $x(T) = 60.0$.

use of an error equal to Δ in the derivative condition makes the value of $(H^e_{n,m} - H_{n,m})$ much larger than Δ. This implies the following rule, as already mentioned earlier:

In valuations where the value of the instrument at the upper boundary is fairly obvious, that value should be used rather than a derivative condition, when using the implicit finite difference method.

The combined effect of a fairly slow error reduction and a large value of REL can, for reasonable errors in the derivative condition, give noticeable errors in the option values, in some extremely unfavourable cases.

3.6.3.2 European put option, errors in the boundary conditions

Table 3.9 shows the reduction factors, see equations (3.97)-(3.100), for a European put option. All the parameter values are as in the base case (see subsection 3.6.2), except for $x(T)$.

When the numbers in table 3.9 are compared to the numbers in the table with reduction factors for a European call option (i.e., table 3.4), the numbers are found to be identical. This is not surprising, however.

The general analysis in section 3.5 shows that the amplification matrix, $\mathbf{B} = \mathbf{A}^{-1}$, solely determines the propagation of errors in the boundary conditions. The amplification matrix in the case of a European call option is identical with the amplification matrix in the case of a European put option, when the parameter values are the same. Thus, the analysis of the effect from errors in the boundary conditions in sub-subsection 3.6.3.1 is equally relevant for a European put option.

The aim of, and the conclusion from, the investigation of European put options in this sub-subsection (i.e., sub-subsection 3.6.3.2) is to confirm that as long as the amplification matrices are nearly identical, errors in the boundary conditions will affect the solutions in a similar way when using the implicit finite difference method. Furthermore, the amplification matrices are nearly the same for all (standard) option pricing problems. Thus, the analysis in sub-subsection 3.6.3.1 is fairly general.

Reduction factors for the lower and the upper boundary						
Lower boundary				Upper boundary		
$x(T)$	$REDL_i$	$MULL_i$		$x(T)$	$REDU_i$	$MULU_i$
1.0	104.8108	104.8		390.0	1.1673	1.167
2.0	32.4962	3.406×10^{3}		380.0	1.1820	1.380
3.0	19.5763	6.668×10^{4}		370.0	1.1984	1.654
4.0	14.1904	9.462×10^{5}		360.0	1.2166	2.012
5.0	11.2362	1.063×10^{7}		350.0	1.2370	2.489
6.0	9.3698	9.961×10^{7}		340.0	1.2598	3.135
7.0	8.0837	8.052×10^{8}		330.0	1.2853	4.029
8.0	7.1453	5.752×10^{9}		320.0	1.3140	5.295
9.0	6.4261	3.696×10^{10}		310.0	1.3464	7.129
10.0	5.8607	2.166×10^{11}		300.0	1.3829	9.858
11.0	5.4035	1.171×10^{12}		290.0	1.4244	14.04
12.0	5.0262	5.884×10^{12}		280.0	1.4717	20.67
13.0	4.7094	2.771×10^{13}		270.0	1.5259	31.53
14.0	4.4397	1.230×10^{14}		260.0	1.5883	50.09
15.0	4.2072	5.176×10^{14}		250.0	1.6605	83.17
16.0	4.0048	2.073×10^{15}		240.0	1.7448	145.1
17.0	3.8270	7.932×10^{15}		230.0	1.8440	267.6
18.0	3.6694	2.911×10^{16}		220.0	1.9616	524.9
19.0	3.5289	1.027×10^{17}		210.0	2.1026	1.104×10^{3}
20.0	3.4028	3.495×10^{17}		200.0	2.2736	2.509×10^{3}
21.0	3.2890	1.150×10^{18}		190.0	2.4839	6.232×10^{3}
22.0	3.1858	3.662×10^{18}		180.0	2.7465	1.712×10^{4}
23.0	3.0917	1.132×10^{19}		170.0	3.0807	5.273×10^{4}
24.0	3.0056	3.403×10^{19}		160.0	3.5152	1.854×10^{5}
25.0	2.9265	9.959×10^{19}		150.0	4.0947	7.590×10^{5}
26.0	2.8535	2.842×10^{20}		140.0	4.8915	3.713×10^{6}
27.0	2.7861	7.918×10^{20}		130.0	6.0286	2.238×10^{7}
28.0	2.7235	2.156×10^{21}		120.0	7.7265	1.729×10^{8}
29.0	2.6654	5.748×10^{21}		110.0	10.4090	1.800×10^{9}
30.0	2.6111	1.501×10^{22}		100.0	14.9618	2.693×10^{10}

Table 3.9: Reduction factors: European put option, base case.

3.7 Conclusions

The analysis of "how errors in the boundary conditions affect the solution when the implicit finite difference method is used" produced the following results:

The total reduction of the effect from errors in the lower boundary condition increases rapidly when moving away from the lower boundary, in all cases. In fact, the total reduction of the effect from errors in the lower boundary condition becomes amazingly large only a few grid points away from the lower boundary. [In the worst case, i.e., for a very high volatility ($\sigma = 0.75$), the total reduction is much lower than in any other case. However, in this case the total reduction also soon becomes very large as the distance from the lower boundary increases.]

Thus, errors in the lower boundary condition have to be very large in order to have any perceivable impact on the solution. To be more concrete, almost any value can be placed at the lower boundary without affecting the solution. This confirms the observations made by Peter Jennergren when examining solutions to problems handed in by students, as discussed in section 3.1.

The reduction of the effect from errors in the upper boundary condition is much slower than the reduction of the effect from errors in the lower boundary condition. The total reduction of the effect from errors in the upper boundary condition is, however, very large for reasonable values of the current price of the underlying security. (With a sensible specification of the grid, the current price of the underlying security will be somewhere far away from the upper boundary.) This is also true for the worst case, i.e., for a very high volatility ($\sigma = 0.75$). This is in spite of the fact that the total reduction is considerably lower in this worst case than in any other case.

A derivative condition at the upper boundary has a bad property. The successive use of a constant error in the values of the derivative leads to a difference between the contingent claim values (for the current time) at the upper boundary with and without errors that is larger than the constant error. Moreover, this difference is many times larger (usually between 50 and 90 times larger) than the constant error in the values of the derivative. This phenomenon can be referred to as the enlargement effect. (High volatilities produce both slow reduction of the effect from errors in the upper boundary condition and a large enlargement effect.)

The combined effect of a fairly slow reduction of the effect from errors in the upper boundary condition and a large value of the enlargement effect can, for reasonable errors in the upper derivative condition, produce a noticeable error in the solution, in some extremely unfavourable cases.

The fairly slow reduction of the effect from errors in the upper boundary condition leads to the following advice:

Advice 1: The grid should be specified in a manner such that the current price of the underlying security is somewhere far away from the upper boundary (that is, far below the middle in the price dimension), when using the implicit finite difference method.

Also, the presence of the enlargement effect leads to the following advice:

Advice 2: In valuations where one has a fairly good idea of the value of the instrument at the upper boundary, that value should be used rather than a derivative condition, when using the implicit finite difference method. Suggestions for conditions on the value at the upper boundary that can be used when valuing some common contingent claim types are given in appendix BIII.

Even if the total reduction of the effect from errors in the boundary conditions changes, both for different sets of parameter values and depending on whether the errors are in the lower or in the upper boundary condition, the results of this investigation imply that errors in the boundary conditions of the magnitude likely to occur in any real valuation will not lead to an error in the solution of any practical significance. An intuitive explanation to this is as follows:

With a proper specification of the grid (i.e., ensuring that the specified size of the grid in the dimension of the underlying state variable is large enough), the probability for the underlying state variable to reach the lower or the upper boundary condition should be close to zero. Thus, due to their low probabilities, the value of the contingent claim at the lower and the upper boundary have small influence, when calculating the current contingent claim price. Hence, even completely erroneous boundary conditions do not generally give any noticeable effects on the solution.

Actually, when modelling the price process of the underlying asset as a geometric Brownian motion (as is the case in the valuations in this paper), the probability for the price to be equal to zero (which is the lower boundary for the price of the underlying asset in the numerical valuations in this paper) is zero [if $S(t_0) = x(T) > 0$]. Consequently, this gives an intuitive explanation for the fact that one can put almost anything at the lower boundary without affecting the solution in the valuations in this paper.

The results of this investigation imply that the argument against the implicit finite difference method based on the fact that the method requires specifications of boundary conditions is not very convincing. To be more precise, the argument is very weak.

The results of this study may also justify the lack of discussion of how the boundary conditions are modelled in many articles in which the implicit finite difference method is used. The exact modelling is simply not interesting, since differences in the modelling choice very rarely produce any differences in the result.

Finally, the implicit finite difference method seems to be robust enough, with respect to errors in the boundary conditions, to be well suited for commercial purposes.

Appendices:

BI Remarks on the classification of partial differential equations

The purpose of this appendix is to show that the partial differential equation that the value of any contingent claim, W, must satisfy (given the assumptions in section 3.3), is always a parabolic equation. It will thus be shown that

$$\begin{aligned}\frac{1}{2}W_{SS}(S,t)\sigma^2(S,t)S^2 + [r(t)S - D^t(S,t)]W_S(S,t) \\ -r(t)W(S,t) + W_t(S,t) + d^t(S,t) &= 0,\end{aligned} \tag{3.103}$$

is always a parabolic partial differential equation.

Hadamard (see [60]) calls a problem of mathematical physics *well-posed* if its solution exists, is unique and depends continuously on the data.

Partial differential equations can be classified according to the type of side conditions that must be imposed to produce a well-posed problem. In the case of linear differential equations of the second order in two independent variables, x and y, this classification is easy to describe. Every differential equation of this type is a special case of the equation,

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu + G = 0, \tag{3.104}$$

where the coefficients A, B, C, D, E, F and G can be functions of x and y.

Equation 3.104 is called *elliptic, hyperbolic* or *parabolic* if the determinant

$$det\begin{pmatrix} A & B \\ B & C \end{pmatrix} \tag{3.105}$$

is positive, negative or zero respectively.

This classification depends in general on the region of the $(x, y) - plane$ under consideration. That is, the same partial differential equation can be elliptic, hyperbolic or parabolic in different regions of the $(x, y) - plane$.

Apply the determinant criterion in equation (3.105) on equation (3.103). This gives

$$det\begin{pmatrix} \frac{1}{2}\sigma^2(S,t)S^2 & 0 \\ 0 & 0 \end{pmatrix} = 0. \tag{3.106}$$

From equation (3.106) it is obvious that equation (3.103) is parabolic in the whole $(S,t) - plane$.

BII Remarks on the transformation of the fundamental partial differential equation into the simple heat flow equation

This appendix gives a relatively universal example of how the general partial differential equation which the price of a contingent claim must satisfy can be transformed into the simple heat flow equation from physics. The example closely resembles the transformation used in the classical paper by *Black and Scholes* (see [12] pp.643-644). This appendix is, however, more general than the Black and Scholes article. This is due to the fact that equation (3.107) (see below) allows for instantaneous payments to the contingent claim holders [the term $d(x,\tau)$, which, however, is assumed to be constant later on in this appendix], and instantaneous payments to the holders of the underlying security [the term $D(x,\tau)$, which, however, is assumed to be a constant yield later on in the appendix].

A fairly general partial differential equation, for the price of the contingent claim, is given by

$$\begin{aligned} \frac{1}{2}H_{xx}(x,\tau)\sigma^2x^2 + [rx - D(x,\tau)]H_x(x,\tau) \\ -rH(x,\tau) - H_\tau(x,\tau) + d(x,\tau) &= 0. \end{aligned} \tag{3.107}$$

Assume that the underlying security pays a constant yield, and that the instantaneous payment to the holders of the contingent claim is constant. In other words it is assumed that

$$c = \frac{D(x,\tau)}{x(\tau)} = constant, \tag{3.108}$$

and

$$d(x,\tau) = d = constant. \tag{3.109}$$

Substitute equations (3.108) and (3.109) into equation (3.107). This gives

$$\begin{aligned}\frac{1}{2}H_{xx}(x,\tau)\sigma^2x^2 + (r-c)xH_x(x,\tau)&\\ -rH(x,\tau) - H_\tau(x,\tau) + d &= 0.\end{aligned} \tag{3.110}$$

In order to transform equation (3.110) into the simple heat flow equation, the following substitution is introduced:

$$H(x,\tau) = e^{-r\tau}u(f(x,\tau), g(\tau)) + \frac{d}{r}, \tag{3.111}$$

where

$$\begin{aligned} f(x,\tau) &= \frac{2}{\sigma^2}(r-c-\frac{\sigma^2}{2})[\ln x + (r-c-\frac{\sigma^2}{2})\tau], \\ g(\tau) &= \frac{2}{\sigma^2}(r-c-\frac{\sigma^2}{2})^2\tau. \end{aligned}$$

With the substitution in equation (3.111) the partial derivatives of $H(x,\tau)$ become

$$H_x = e^{-r\tau}\frac{2}{\sigma^2 x}(r-c-\frac{\sigma^2}{2})u_f, \tag{3.112}$$

$$H_{xx} = e^{-r\tau}[\frac{4}{\sigma^4x^2}(r-c-\frac{\sigma^2}{2})^2u_{ff} - \frac{2}{\sigma^2x^2}(r-c-\frac{\sigma^2}{2})u_f], \tag{3.113}$$

$$H_\tau = e^{-r\tau}[-ru + \frac{2}{\sigma^2}(r-c-\frac{\sigma^2}{2})^2u_f + \frac{2}{\sigma^2}(r-c-\frac{\sigma^2}{2})^2u_g]. \tag{3.114}$$

With the help of equations (3.112), (3.113) and (3.114), equation (3.110) can be transformed into

$$u_{ff} = u_g, \tag{3.115}$$

which is the simple heat flow equation from physics. In the derivation of equation (3.115) it is assumed that $(r-c-\frac{\sigma^2}{2}) \neq 0$.

BIII Suggestions for upper boundary values

In this appendix suggestions will be given for values to use at the upper boundary for some common contingent claim types.

Let us start with call options on a stock. First, consider a European call option on a stock that pays a continuous dividend yield[46], c. A reasonable condition on the upper boundary value of a European call option on a stock that pays a continuous dividend yield is

$$H(x^u, \tau) = x^u e^{-c\tau} - \text{Œ} e^{-r\tau},$$

and the discrete version of this condition is

$$H_{n,j} = nhe^{-cjk} - \text{Œ} e^{-rjk}.$$

An American call option on a stock that pays a continuous dividend yield can be exercised at any time. Also, an American option can never be worth less than its European counterpart. Thus, a reasonable condition on the value to use at the upper boundary when valuing an American call option on a stock that pays a continuous dividend yield is

$$H(x^u, \tau) = max[x^u - \text{Œ}, x^u e^{-c\tau} - \text{Œ} e^{-r\tau}].$$

Discretization of the above condition gives

$$H_{n,j} = max[nh - \text{Œ}, nhe^{-cjk} - \text{Œ} e^{-rjk}].$$

Next, consider call options on a stock that pays discrete dividend payments. Introduce the following notation:

- D_s $s = 1, \ldots, S$ are the discrete dividend payments that will be paid during the remaining life of the option.

- τ_s $s = 1, \ldots, S; \tau_s < \tau; \tau_1 < \tau_2 < \ldots < \tau_S$ are the times from current time to respective dividend payment day.

[46]Of course, the value of this type of option is easier to calculate with the analytical formula given in equation (3.81), than with the implicit finite difference method. A European call option on a stock that pays a continuous dividend yield is, however, a good starting point when a condition on the value of its American counterpart to use at the upper boundary is sought.

A reasonable condition on the value at the upper boundary for a European call option on a stock that pays discrete dividend payments is[47]

$$H(x^u, \tau) = x^u - \sum_{s=1}^{S} D_s e^{-r\tau_s} - \text{Œ}e^{-r\tau},$$

and the discrete version of the condition becomes

$$H_{n,j} = nh - \sum_{s=1}^{S} D_s e^{-r(j-j_s)k} - \text{Œ}e^{-rjk},$$

where j_s $s = 1, \ldots, S$ is the number of grid points from maturity to dividend payment D_s.

It is a well-known fact that it can only be optimal to exercise an American call option on a stock that pays discrete dividends at a time immediately before the stock goes ex-dividend (e.g., see [67] p.125)[48]. This means that the following condition on the value is reasonable to use at the upper boundary when valuing an American call option on a stock that pays discrete dividends:

$$H(x^u, \tau) = max[x^u - \text{Œ}e^{-r\tau_1}, x^u - D_1 e^{-r\tau_1} - \text{Œ}e^{-r\tau_2},$$
$$x^u - D_1 e^{-r\tau_1} - D_2 e^{-r\tau_2} - \text{Œ}e^{-r\tau_3}, \ldots, x^u - \sum_{s=1}^{S} D_s e^{-r\tau_s} - \text{Œ}e^{-r\tau}].$$

Thus, the discrete version of the upper boundary condition for an American call option that pays discrete dividend payments can be written as

$$H_{n,j} = max[nh - \text{Œ}e^{-r(j-j_1)k}, nh - D_1 e^{-r(j-j_1)k} - \text{Œ}e^{-r(j-j_2)k},$$
$$nh - D_1 e^{-r(j-j_1)k} - D_2 e^{-r(j-j_2)k} - \text{Œ}e^{-r(j-j_3)k}, \ldots, nh - \sum_{s=1}^{S} D_s e^{-r(j-j_s)k} - \text{Œ}e^{-rjk}].$$

[47] The value of a European call option on a stock that pays discrete dividend payments is, of course, easier to calculate with a slight modification of the Black & Scholes formula (e.g., see [67] pp.123-124), than to calculate with the help of the implicit finite difference method. But, also in this case the upper boundary condition for the European option is a good starting point for determining a condition on the upper boundary value of its American counterpart.

[48] It can also be noticed that in most cases when valuing an American call option on a stock which pays discrete dividends the only time for early exercise that needs to be considered is the time immediately prior to the dividend payment closest to maturity. Furthermore, to find that early exercise is never optimal is not uncommon when valuing American call options (e.g., see [67] p.126).

The next contingent claim types to consider are put options on a stock. Irrespective of whether it is a European or American put option or if the stock pays a continuous dividend yield[49] or discrete dividends[50]

$$H(x^u, \tau) = 0,$$

is an appropriate condition to use at the upper boundary. A discrete version of the condition is given by

$$H_{n,j} = 0.$$

The last contingent claim type to be discussed is convertible bond. Two different convertible bond issues can be rather different, due to different call, put and conversion features. This means that, for some convertible bond issues, there can exist more appropriate conditions on the value to use at the upper boundary than those suggested in this appendix. Furthermore, the conditions given are the simplest possible. Perhaps more complicated and slightly more accurate conditions can be derived if it is also taken into account how discrete coupon payments and discrete dividend payments affect the timing of the early exercise decision[51] (cf. the modelling of the upper boundary conditions for call options previously in this appendix.).

When valuing convertible bonds in a model with only one state variable the modeller can choose to use either the stock price or the firm value as underlying state variable. Let us start with the case where the stock price is used as underlying state variable.

[49]The value of a European put option on a stock that pays a continuous dividend yield is, of course, easier to calculate with the analytical formula given in equation (3.89), than to calculate with the help of the implicit finite difference method. Also, the value of a European put option on a stock that pays discrete dividends is easier to calculate with a modification of Black & Scholes formula (e.g., see [67] pp.123-124).

[50]It can be shown that it is never optimal to exercise an American put option on a stock that pays discrete dividends early if $D_s \geq \text{Œ}(1 - e^{-r(\tau_{s+1} - \tau_s)})$ for all $s < S$ and $D_S \geq \text{Œ}(1 - e^{-r(\tau - \tau_S)})$ (e.g., see [67] p.128).

[51]For example, it can be shown that it can only be optimal to convert a convertible bond immediately prior either to a dividend date or to an adverse change in the conversion terms, or at maturity (see [21]).

A reasonable condition on the value to use at the upper boundary when valuing a convertible bond with the stock price as underlying state variable is simply to set

$$H(x^u, \tau) = \alpha x^u,$$

where α = *the number of shares of stock that a bond certificate can be converted into.* Obviously, the discrete version of the condition to be used at the upper boundary is

$$H_{n,j} = \alpha n h.$$

Finally, when calculating the value of a convertible bond with the firm value as underlying state variable[52] the following condition on the value can be used at the upper boundary[53]:

$$H(v^u, \tau) = \frac{\alpha}{\beta + \kappa\alpha} v^u,$$

where β = *the number of shares of stock outstanding before conversion takes place*, α = *the number of shares of stock that a bond certificate can be converted into*, and κ = *the number of bond certificates outstanding.* Thus, the discrete version of the upper boundary condition for a convertible bond with the firm value as underlying state variable is

$$H_{n,j} = \frac{\alpha}{\beta + \kappa\alpha} n h.$$

[52]With the firm value as underlying state variable, all notation remains as previously, except that x is replaced by v everywhere, v being the value of the firm.

[53]When using this condition at the upper boundary, it is assumed that the company's capital structure consists solely of common stock and the convertible bond issue. If there are more instruments in the capital structure, the upper boundary condition has to be modified. Also, the solution methodology becomes much more complex since there are combinatorial interactions between the instruments (see [74]).

Chapter 4

Paper C: Two finite difference schemes for evaluation of contingent claims with three underlying state variables

4.1 Introduction

In 1973 a major breakthrough came in the field of finance, when *Black and Scholes* presented an ingenious way of valuing options (see [12]). The paper by Black and Scholes was followed by two seminal papers from *Merton* (see [99] and [100]). Together, these three papers make up the foundation of modern option pricing.

The theory of option pricing provides a unified framework for the valuation of many different kinds of securities. In fact, in terms of their valuation, many more or less complex financial instruments have attributes that make option pricing approaches superior to other (currently known) valuation methods. Furthermore, option pricing theory can be useful in contexts beyond the valuation of financial instruments, e.g., for capital budgeting decisions (see [94]).

The key concept in the option pricing framework is that the price of the asset to be valued depends solely upon the evolution of some underlying state variables and time. The theory, in other words, gives a framework for the valuation of contingent claims[1].

Valuation of a contingent claim through the use of option pricing theory, in continuous time, very often includes a derivation of a partial differential equation which the price of the contingent claim must satisfy. This fundamental partial differential equation is almost identical for different contingent claims. The features that distinguish one contingent claim from another affect the side conditions that the fundamental partial differential equation is subject to.

Unfortunately, the fundamental partial differential equation subject to the relevant side conditions lacks an analytical solution in most cases. Therefore, the solution to the fundamental partial differential equation often has to be obtained by means of numerical methods. Many different numerical methods have been used for the valuation of contingent claims, e.g., analytical approximation, series solution, numerical integration, Monte Carlo simulation, lattice approaches or finite difference methods. Finite difference methods are among the most commonly used, and are the methods which will be focused on in this paper.

There are many examples of valuation of contingent claims by numerical methods in the literature. These examples are, in an overwhelming majority, valuations of contingent claims with only one underlying state variable (see for example [54], [74], [95] or [122]). During the last two decades, however, there has been an explosion of new exotic financial instruments. Many of these instruments have features which make the assumption that the value of the instrument depends only on one underlying state variable and time an oversimplification. Indeed, to accurately price these exotic instruments, it is often necessary to assume that the value of the contingent claim depends on two or even more underlying state variables and time.

Naturally, as the number of state variables is increased, the likelihood of finding an analytical solution to the fundamental partial differential equation subject to the relevant

[1]Although contingent claim is a much wider concept than financial instrument, contingent claim and financial instrument will be used interchangeably in this paper.

side conditions is decreased. Thus, the probability that numerical methods will have to be used for the valuation of contingent claims also increases as the number of state variables increases. However, as the number of underlying state variables increases so does the degree of complexity and the need for computing power, when using numerical methods. Moreover, the degree of complexity and computing power needed increases substantially for every additional state variable.

For the case of contingent claims with two underlying state variables, the literature contains some examples of valuation by numerical methods (e.g., see [15], [69] or [73]). These examples are, however, considerably fewer than the examples of valuation of contingent claims with only one underlying state variable.

As for the case of contingent claims with three underlying state variables, there is as far as the author has been able to ascertain, only one published example of a valuation by numerical methods. This example is a valuation by means of a lattice approach (see [16]).

Finite difference schemes have proved to be very flexible numerical methods for the pricing of financial instruments with one and two underlying state variables. The schemes are flexible both with regard to the features of the instruments (puttable, callable, extendable etc.) as well as the type of stochastic processes that the underlying state variables follow. The flexibility of the finite difference schemes and the steady stream of new complex financial instruments imply that finite difference schemes for the valuation of contingent claims with three underlying state variables can supposedly be very useful. In this paper, two such schemes will be developed and tested. A brief outline of the paper will be given in the remainder of this section.

Valuations of contingent claims with three underlying state variables by means of finite difference schemes require considerable computing power. In fact, the valuations require so much computing power that they can hardly be executed on an ordinary PC. For this reason all "heavy" computations in this paper have been performed on a supercomputer of type Connection Machine Model CM-2000. A description of the Connection Machine CM-2000 System is given in section 4.3.

Because of the use of the supercomputer it may be argued that the relevance of the research performed in this paper is low, since it is commonly believed that there are very few advanced architecture computers in the financial community. This is, however, incorrect. Several financial companies are currently using advanced architecture computing, and others are likely to do the same at an increasing pace in the future. In section 4.2, a short overview of "The use of advanced architecture computing in the financial services industry" is given.

In section 4.4, the fundamental partial differential equation that the value of a contingent claim with three underlying state variables must satisfy is derived.

In order to gain some idea of the accuracy of the finite difference schemes in this paper, valuations of two financial instruments with three underlying state variables that also have analytical solutions will be performed. Thus, an idea of the schemes' accuracy can be given by comparing the numerical solution to the analytical solution for different sets of parameter values. The instruments that will be valued are European call options on the maximum of three risky assets and European call options on the minimum of three risky assets. Descriptions of these instruments and their analytical solutions are given section 4.5.

Preliminary evaluations showed that it would be necessary to make transformations of the underlying assets' price processes in order to achieve stability[2] for the difference schemes in this paper. These transformations and the resulting partial differential equation are described in section 4.6.

It should now be clear that the purpose of this paper is to develop and test two finite difference schemes which can be used when valuing contingent claims with three underlying state variables. These finite difference schemes are described in section 4.7.

The numerical evaluations are performed in section 4.8, and the conclusions from this investigation are given in section 4.9.

[2]In short, the problem of stability concerns the unstable growth or stable decay of the errors in the arithmetic operations needed when a numerical method is used. Each calculation carried out introduces a round-off error. Generally, a numerical method is stable when the cumulative effect off all the rounding errors introduced when the method is used is negligible (see [126] p.57).

4.2 The use of advanced architecture computing in the financial services industry

The purpose of this section of the paper is to give a brief insight into the use of/prospects for advanced architecture computing in the financial services industry. The section is largely based on two articles by *Zenios* (see [139] and [141]), and one article by *Worzel and Zenios* (see [137]). In particular [137], and a survey reported in that article, are referred to extensively.

First, it is necessary to explain what is meant by advanced architecture computers. Included in the group of advanced architecture computers are:

- *"Traditional" supercomputers:* The term traditional supercomputers refers to single processor and vector computers. The most well-known traditional supercomputers are different models of CRAY computers.

- *Parallel computers:* The term parallel computers refers to the class of computer architectures that has multiple processing units. Thus, the class of parallel computers includes a fairly broad spectrum of computers. An obvious example of a (fairly advanced) parallel computer is the Connection Machine described in section 4.3.

There are several reasons for believing that parallel computers will successively gain more and more of the "market" from traditional supercomputers. Among these reasons the following can be mentioned:

- The most advanced single processor computer architectures are close to their practical limits in terms of speed. As an example, the fastest single processor computer in the world, CRAY Y-MP, generates a signal every sixth nanosecond. With this signal frequency, components of the machine that are more than six feet apart cannot be kept synchronized (see [137]). Thus, to achieve significant improvements in computing speed multi-processor architectures have to be used.

- Parallel computing is in an early stage of the development process. For this reason, great improvements in the performance of parallel architectures for rather modest costs can be expected in the future. In comparison, traditional supercomputers are in the later stage of the development process. This implies that small improvements in performance come at a greater cost (see [141]).

- Many technical, physical and economical problems have a natural inherent parallelism (see [139]). Thus, the use of parallel architectures can simplify the development of solution algorithms for these types of problems.

- Parallel computers use multiple processors that utilize a common memory. Naturally, these types of architectures offer greater cost efficiency compared to traditional architectures. This is due, as *Zenios* neatly expresses it, to, "After all, only a tiny fraction of a computer's silicon is found in the processing unit." (see [139] p.4).

Before the use of advanced architecture computing in the financial services industry is discussed, it can be of interest to mention some industries for which this type of computing has become a necessity. Advanced architecture computing is extensively used, for example, in the aerospace and automobile industries. With advanced architecture computing, engineers in the aerospace industry can use computer models as complement to/substitute for expensive trials of physical models in wind tunnels (see [137]). Almost all of the world's car manufacturers use advanced architecture computing in functions such as structural analysis, crash simulation, and component design and testing (see [137]). Examples of other industries/fields on which advanced architecture computing has had a significant impact are chemical, pharmaceutical, computer manufacturing, weather forecasting, petroleum exploration and seismology (see [137]).

Within the financial services industry, however, advanced architecture computing has hitherto not been used to a great extent. There are several explanations for this, among which three are of major importance (see [137]). Firstly, it was not until recently that manufacturers of advanced computer architectures started to actively market their products to the financial services industry. Furthermore, the type of marketing used has been more suited for highly technical industries such as aerospace and automobile, than for the traditionally non-technical financial services industry. Secondly, advanced architecture computing has not been able to prove its potential value to financial institutions until the last couple of years. Thirdly, the use of advanced architecture computing often requires considerable "start up" costs in the form of education, and the development of new solution algorithms etc.

Many financial applications are, however, complex, and furthermore often require a high degree of timing. It would therefore seem that advanced architecture computing can certainly be of great value to the financial services industry. Some examples of application areas where the potential benefits to be derived from advanced architecture computing appear to be very great are pricing of complex financial products, trading strategies, portfolio management/immunization, econometric modelling and term structure modelling.

It is therefore not surprising that the financial services industry has started to show a continuously increasing interest in advanced architecture computing during the past few years. Some evidence of this can be found in the following facts:

At least six financial firms are reputed to be using CRAY computers. Among these are Goldman Sachs and Nomura Securities (the use by these two firms has been confirmed) (see [137]).

In 1991, Prudential-Bache Securities publicly acknowledged the purchase of a parallel computer. Furthermore, Prudential-Bache Securities later upgraded its original parallel computer to a more powerful one (see [137]). Blackstone Financial Management and Federal National Mortgage Association (FNMA) together with researchers at the HERMES Laboratory for Financial Modeling of the Wharton School of University of Pennsylvania are developing a model for the pricing of mortgage-backed securities. In this project, a massively parallel Connection Machine (see section 4.3) is being used (see [139]). American Express has reputedly ordered two of the latest and most advanced of the massively parallel Connection Machines (CM-5).

A number of mini-supercomputers are being used by Bear, Stearns & Co (see [137]). Finally, First Boston and the Blackstone Group are using distributed networks to handle their large computing needs (see [137]).

Before this section is concluded, some of the main results from a survey carried out by *Worzel and Zenios* will be reported[3]. Technical details and more complete results can be found in [137][4]. At the time of the survey (1991), only 4% of the responding financial companies' total computer processing was performed with advanced architecture computing. The companies, however, expect this figure to increase to nearly 20% within the next 5 years. Furthermore, more than 20% of the responding financial companies have already used some form of advanced architecture computing in either operational or experimental/research related activities.

The most common of the current application areas for advanced architecture computing are information processing, investment management and equity portfolio management. The potential application areas mentioned by responding firms were fairly varied. Many were, naturally, computer-intensive analytics, e.g., asset/liability management, investment management, portfolio immunization, term structure modelling and portfolio management for fixed income and equity securities.

The most important factors by far in decisions on whether or not to use advanced architecture computers were cost efficiency and software availability. The importance of software availability should be a strong signal to manufacturers. This is a consequence of the fact that manufacturers have used hardware-based marketing, with only limited effort put into the development of financial software.

The defective marketing of advanced architecture computers within the financial services industry is further confirmed by the fact that 41% of the responding companies replied that manufacturers had marketed advanced architecture computers with only a

[3]In their survey Worzel and Zenios used a somewhat broader definition of advanced architecture computers than that which is used in this paper. More precisely, they also included distributed networks and transputer boards in their definition.

[4]The results from the survey carried out by Worzel and Zenios may not be completely reliable. This is due to a low response rate. The low response rate was not surprising given a high degree of secrecy in the financial services industry. Worzel and Zenios claim, however, that the survey in combination with conversations with professionals in the financial services industry make their reported results fairly reliable.

low level of sales and marketing effort. Furthermore, another 33% of the responding firms had never been approached regarding the purchase of an advanced architecture computer.

This lack of marketing effort by computer manufacturers seems very unwise since there appears to be an enormous potential for selling advanced architecture computers to financial companies. For example, consider the fact that 41% of the responding companies replied that they were "marginally interested" in advanced architecture computing, but had not yet made any formal investigations. Moreover, of the responding companies that had evaluated advanced architecture computing, as many as 63% are currently using the technology, and a further 18.5% reported that they plan to buy a system in the future. These figures clearly suggest that a large part of the financial services industry is likely to start using advanced architecture computing once they have evaluated the technology.

4.3 The Connection Machine CM-2000 System

Valuation of contingent claims with more than two underlying state variables requires considerable computing power. The computer that is used for the valuations in this paper is a Connection Machine model CM-2000.

In this section, a short description of the CM-2000 machine will be given. The reason for giving this description is twofold. Firstly, the machine type is fairly new, different from traditional computers and unknown to many. Secondly, the CM-2000 Connection Machine offers the possibility to solve problems that would be impossible to solve with traditional computers. Information regarding the system can therefore be of interest to others.

4.3.1 A new paradigm of computation

The advent of massive parallelism represents a major revolution in supercomputing, and promises spectacular increases in computational speed. The Connection Machine Model CM-2000 is a fine grain massively parallel supercomputer. It uses a single instruction stream and a multiple data stream.

The key idea in the Connection Machine's programming paradigm is the association of processors with data. One application might associate one processor per node in a finite difference calculation while another application might associate one processor per random walk in a Monte Carlo simulation.

For this parallelism to be successful, there must be some notion of interprocessor connectivity, so that processors can communicate. This could be anything from nearest neighbour communication on a Cartesian grid for a finite difference calculation, to full interconnectivity (all processors talk to each other).

A single instruction set is broadcast to all the processors of the CM-2000 which accordingly manipulate their data in parallel. Instructions of this type can specify local operations, such as telling all processors to multiply two locally stored numbers; local communication, such as telling all processors to obtain data from a neighbour on a grid; or global or cooperative operations, such as adding up the elements of a particular array variable across all the processors of the machine. Inevitably, the Connection Machine programmer begins to think of the processor configuration as corresponding to that of the problem itself.

Thus, although this paradigm differs from what serial computer programmers have been used to, it is in many ways a conceptually simpler paradigm. The approach when using the Connection Machine is to make the computer look like the problem being solved.

4.3.2 Computational model of the Connection Machine Model CM-2000

This subsection gives a brief description of the hardware and the computational model of the CM-2000 Connection Machine system.

The Connection Machine Model CM-2000 is a data parallel computing system, and was specifically designed to handle the largest computational problems. Data parallel computing associates one processor with each data element. This computing style exploits the natural computational parallelism inherent in many data-intensive problems. It can significantly decrease the execution time of a problem, as well as simplify its programming.

The Connection Machine Model CM-2000 is an integrated system of hardware and software. The hardware elements of the system include front-end computers that provide the development and execution environment for the users' software, a parallel processing unit that executes the data parallel operations and a high-performance data parallel I/O system. Software elements begin with the standard operating system and program development environment of the front-end computer, and enhance that enviroment with extensions to standard languages and tools that facilitate data parallel program development.

The front-end computer is the users' gateway to the Connection Machine system. The front-end machine supports the operating environment; all programming is done via the front-end using its editors, compilers, debuggers and other utilities. The front-end machine executes the control structure of programs, issuing commands to the CM-2000 processors whenever necessary.

The central element in the system is the parallel processing unit, which contains:

- from 4K (4096) to 64K (65536) processors
- a sequencer that controls the data processors
- an interprocessor communications network
- zero or more I/O controllers and/or framebuffer modules

Each of the data processors can execute arithmetic and logical instructions, calculate memory addresses and perform interprocessor communication or I/O. In this respect, each data processor is very much like an ordinary serial computer. The difference is that the data processors do not retrieve instructions from their respective memories. Instead they are collectively under the control of a single microcoded sequencer. The task of the sequencer is to decode commands from the front-end computer and broadcast them to the data processors, which then execute the same instruction simultaneously. The memories of the parallel processors hold the bulk of data that can be usefully processed in parallel. Naturally, some small part of an application's data may be better processed serially. Such data resides in the memory of the front-end computer and is processed serially in the usual way.

Interprocessor communication is particularly important in data parallel processing. This communication is implemented by a special-purpose high-speed network. When data is needed, it is passed over the network to the appropriate processors. Processors that hold interrelated data elements store pointers to one another, thus supporting completely general patterns of communication. In addition, special hardware supports certain commonly used regular patterns of communication. Nearest-neighbor communication in a multi-dimensional rectangular grid is particularly efficient.

The Connection Machine system implements data parallel programming constructs directly in hardware and microcode. Parallel data structures are spread across the data processors, with a single element stored in each processor's memory. When parallel data structures contain more data elements than the system has processors (the normal situation), the system operates in virtual processor mode, presenting the user with a larger number of processors, each with a correspondingly smaller memory. This allows the user to write programs assuming the number of processors that is natural for the application, rather than forcing code to conform to the number of hardware processors available. Each hardware processor is made to simulate the appropriate number of virtual processors; as the program issues each parallel instruction, microcode causes it to be executed many times, once for each virtual processor. The same program can run without change on different numbers of hardware processors, but the more hardware the faster it runs.

High-speed transfers between peripheral devices and Connection Machine memory take place through the Connection Machine I/O system. All processors, in parallel, pass data to and from I/O buffers. The data is then moved between the buffers and the peripheral devices.

4.4 Derivation of the fundamental value equation

In order to derive the fundamental partial differential equation that the prices of the contingent claims in this paper satisfy, the following assumptions are made (see also [100] and [101]):

A.1 There are no transaction costs, taxes or problems with indivisibilities of assets.

A.2 There are a sufficient number of investors with comparable wealth levels so that each investor believes that he/she can buy and sell as much of an asset as he/she wants at the market price.

A.3 There is an exchange market for borrowing and lending at the same rate of interest.

A.4 Short-sales of all assets, with full use of the proceeds, is allowed.

A.5 Trading in assets takes place continuously in time.

A.6 There is a riskless asset whose rate of return per unit time is known and constant. Denote this rate of return by r.

A.7 The underlying state variables are prices of traded securities. The dynamics for the prices of the underlying securities are described by the following stochastic differential equations:

$$dS_i = \mu_i S_i dt + \sigma_i S_i dz_i \qquad i = 1, \ldots, n, \tag{4.1}$$

where

- t = calendar time, $t_0 \le t \le T + t_0$ where t_0 is current time and T is time to maturity from t_0.
- μ_i = S_i's instantaneous expected rate of return per unit time.
- σ_i^2 = the instantaneous variance per unit time of S_i's instantaneous rate of return. σ_i^2 $i = 1, \ldots, n$ are assumed to be constants.

- $z_i(t)$ = a Wiener process, that is $E(dz_i) = 0$, $Var(dz_i) = dt$ and $dz_i \sim N(0, dt)$ where $N(\bullet, \bullet)$ indicates the normal distribution. It is also assumed that $cov(dz_i(t), dz_j(t)) = \rho_{ij} dt$ where $\rho_{ij} = constant$ $i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$ and that $cov(dz_i(t), dz_j(s)) = 0$ $i, j = 1, \ldots, n; s \neq t$.

A.8 It is assumed that investors prefer more to less. It is also assumed that investors agree upon σ_i and ρ_{ij} $\forall i, j$, but it is *not* assumed that they necessarily agree upon μ_i $\forall i$.

A.9 The price of the contingent claim is a function of the prices of the underlying securities and time. Denote the price of the contingent claim by $W(S_1, \ldots, S_n, t)$. $W(S_1, \ldots, S_n, t)$ is a real valued continuous non-random function such as $\frac{\partial W}{\partial t}, \frac{\partial W}{S_i}$ $\forall i$ and $\frac{\partial^2 W}{\partial S_i \partial S_j}$ $\forall i, j$ are continuous[5]. Moreover, each individual contingent claim has its own properties. The properties can be formulated as side conditions, which distinguish different contingent claims from each other. The side conditions are not formulated in a general way in this section, but instead are formulated in section 4.5 explicitly for each of the two contingent claims that are valued in this paper.

However, the model holds even if many of the assumptions above are substantially weakened (see [100]).

The derivation of the fundamental partial differential equation will depend to a large extent upon the following lemma:

[5]In the remainder of this paper, partial derivatives will often be denoted with the help of subindexes. That is, the following fairly standard notation for partial derivatives of a function $f(x, y)$ will be used:

$$\begin{aligned}
\frac{\partial f(x,y)}{\partial x} &= f_x, \\
\frac{\partial f(x,y)}{\partial y} &= f_y, \\
\frac{\partial^2 f(x,y)}{\partial x^2} &= f_{xx}, \\
\frac{\partial^2 f(x,y)}{\partial y^2} &= f_{yy}, \\
\frac{\partial^2 f(x,y)}{\partial x \partial y} &= f_{xy}.
\end{aligned}$$

lemma 4.1 (Generalized Itô's) [6] *Assume that $u(x_1, \ldots, x_n, t)$ is a real valued continuous non-random function such that its partial derivatives*

$$
\begin{aligned}
u_t &= \frac{\partial u(x_1, \ldots, x_n, t)}{\partial t}, \\
u_{x_i} &= \frac{\partial u(x_1, \ldots, x_n, t)}{\partial x_i} \qquad i = 1, \ldots, n, \\
u_{x_i x_j} &= \frac{\partial^2 u(x_1, \ldots, x_n, t)}{\partial x_i \partial x_j} \qquad i, j = 1, \ldots, n,
\end{aligned}
$$

are continuous. Suppose that $x_i(t)$ $i = 1, \ldots, n$ are real valued diffusion processes with stochastic differentials

$$dx_i = f_i(x_1, \ldots, x_n, t)dt + b_i(x_1, \ldots, x_n, t)dy_i \qquad i = 1, \ldots, n,$$

where $y_i(t)$ $\forall i$ are Wiener processes, $E(dy_i dy_j) = \rho_{ij}dt$ $\forall i, j$, and f_i $\forall i$ and b_i $\forall i$ are real valued functions of $x_1, \ldots, x_n$ and t. Then $u(x_1, \ldots, x_n, t)$ has a differential given by

$$du = (u_t + \sum_{i=1}^{n} u_{x_i} f_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} u_{x_i x_j} b_i b_j \rho_{ij})dt + \sum_{i=1}^{n} u_{x_i} b_i dy_i.$$

In order to simplify the notation, the arguments in the functions are not written out explicitly in the remainder of the derivation.

With the help of the generalized Itô's lemma and equation (4.1) the differential of $W(S_1, \ldots, S_n, t)$ can be written as

$$dW = (W_t + \sum_{i=1}^{n} W_{S_i} \mu_i S_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{S_i S_j} \sigma_i \sigma_j \rho_{ij} S_i S_j)dt + \sum_{i=1}^{n} W_{S_i} \sigma_i S_i dz_i. \tag{4.2}$$

Invest the amount α_W in the contingent claim and the amounts $\alpha_{S_1}, \ldots, \alpha_{S_n}$ in the underlying assets. Also, choose the amounts $\alpha_W, \alpha_{S_1}, \ldots, \alpha_{S_n}$ so that all risk disappears. The portfolio described above contains no risk. To prevent arbitrage opportunities the portfolio must therefore earn the riskless rate of return. Thus, the following equation for the instantaneous return from the portfolio must hold:

$$\alpha_W \frac{dW}{W} + \sum_{i=1}^{n} \alpha_{S_i} \frac{dS_i}{S_i} = (\alpha_W + \sum_{i=1}^{n} \alpha_{S_i})rdt. \tag{4.3}$$

[6] For an even more general and also more mathematically correct formulation of the generalized Itô's lemma see [89].

The next step is to find amounts that eliminate all risk in the return from the portfolio constructed above. In order to eliminate the risks from the Wiener processes $z_i(t)$ $i = 1, \ldots, n$ it is obvious from equations (4.1), (4.2) and (4.3) that the amounts must be chosen as

$$\alpha_W \frac{W_{S_i} S_i \sigma_i}{W} + \alpha_{S_i} \sigma_i = 0 \qquad i = 1, \ldots, n. \tag{4.4}$$

It is possible to arbitrarily set

$$\alpha_W = 1. \tag{4.5}$$

With the choice in equation (4.5), equation (4.4) immediately yields

$$\alpha_{S_i} = -\frac{W_{S_i} S_i}{W} \qquad i = 1, \ldots, n. \tag{4.6}$$

Insertion of equations (4.5) and (4.6) into equation (4.3) and the use of equations (4.1) and (4.2) give

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{S_i S_j} \sigma_i \sigma_j \rho_{ij} S_i S_j + \sum_{i=1}^{n} W_{S_i} r S_i + W_t = rW. \tag{4.7}$$

Equation (4.7) is the fundamental partial differential equation that the price of every[7] contingent claim with the prices of the assets in assumption **A.7** as underlying state variables must satisfy. The features that distinguish different contingent claims from each other are incorporated into the side conditions that equation (4.7) is subject to.

It is usually more convenient to express the value of the contingent claim as a function of time to maturity, τ, instead of as a function of calendar time, t. That is,

$$\tau = T - (t - t_0) \qquad t_0 \le t \le T + t_0, \tag{4.8}$$

where t_0 is current time and T is time to maturity from t_0. Also, introduce the following relations:

$$S_i(t) = x_i(\tau) \qquad i = 1, \ldots, n, \tag{4.9}$$

$$W(S_1, \ldots, S_n, t) = H(x_1, \ldots, x_n, \tau). \tag{4.10}$$

[7]Naturally, it must be assumed that $W(S_1, \ldots, S_n, t)$ fulfils some regularity conditions. That is, it must be possible to use the generalized Itô's lemma.

With equations (4.8), (4.9) and (4.10), the fundamental partial differential equation (4.7) can be rewritten as

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} H_{x_i x_j}\sigma_i\sigma_j\rho_{ij}x_i x_j + \sum_{i=1}^{n} H_{x_i} r x_i - H_\tau = rH, \tag{4.11}$$

subject to the side conditions that are relevant for the contingent claim under consideration.

In this paper, only contingent claims with three underlying state variables are considered. Thus, for the purpose of this paper equation (4.11) can be simplified to

$$\begin{aligned} \frac{1}{2}H_{x_1x_1}\sigma_1^2x_1^2 + \frac{1}{2}H_{x_2x_2}\sigma_2^2x_2^2 + \frac{1}{2}H_{x_3x_3}\sigma_3^2x_3^2 + \\ H_{x_1x_2}\sigma_1\sigma_2\rho_{12}x_1x_2 + H_{x_1x_3}\sigma_1\sigma_3\rho_{13}x_1x_3 + H_{x_2x_3}\sigma_2\sigma_3\rho_{23}x_2x_3 + \\ H_{x_1}rx_1 + H_{x_2}rx_2 + H_{x_3}rx_3 - H_\tau &= rH, \end{aligned} \tag{4.12}$$

subject to the relevant side conditions.

4.5 Descriptions of the instruments

In order to gain a comprehension of the accuracy of the finite difference schemes in this paper, the valuation of two of the rare instruments with three underlying state variables that also have an analytical solution will be considered. Thus, an idea of the accuracy of the finite difference schemes can be achieved by comparing the computed values from the finite difference schemes to the values from the analytical formulas for different sets of parameter values[8].

The instruments that are valued are European call options on the maximum of three risky assets and European call options on the minimum of three risky assets. Descriptions of these instruments and the analytical formulas for their values are given in the next two subsections.

[8]In fact, the values from the analytical formulas are also achieved by a numerical method. This means that, in reality, values from different numerical methods are being compared. Further details on this point are given later in the paper.

4.5.1 European call option on the maximum of three risky assets

The price of a European call option on the maximum of three risky assets must satisfy the partial differential equation (4.12) subject to

$$
\begin{aligned}
H(x_1, x_2, x_3, 0) &= max[max(x_1, x_2, x_3) - Œ, 0], \\
0 \leq H(x_1, x_2, x_3, \tau) &\leq max(x_1, x_2, x_3), \\
0 \leq H(x_1, x_2, 0, \tau) &\leq max(x_1, x_2), \\
0 \leq H(x_1, 0, x_3, \tau) &\leq max(x_1, x_3), \\
0 \leq H(0, x_2, x_3, \tau) &\leq max(x_2, x_3), \\
0 \leq H(x_1, 0, 0, \tau) &\leq x_1, \\
0 \leq H(0, x_2, 0, \tau) &\leq x_2, \\
0 \leq H(0, 0, x_3, \tau) &\leq x_3, \\
H(0, 0, 0, \tau) &= 0,
\end{aligned}
\tag{4.13}
$$

where Œ is the exercise price (the remaining notation is explained earlier in the text).

For the purpose of finding an analytical formula for the value of a European call option on the maximum of three risky assets, it is not advisable to try to solve equation (4.12) subject to conditions (4.13) directly. Instead, it is better to use a simple and intuitive method, invented by *Johnson* (see [75]), to deduce the analytical formula.

Johnson's method uses the *Cox and Ross'* approach (see [42]) and the repeated use of change in numeraire. {The trick of using a change in numeraire was first employed by *Margrabe* (see [91]).} The joint effect of the Cox and Ross' approach and repeated use of Margrabe's trick is that it makes it possible to write down almost immediately the analytical formula for the value of a European call option on the maximum of three risky assets.

Johnson's method implies that the analytical formula for the value of a European call option on the maximum of three risky assets can be written as (see also [75])

$$H(x_1, x_2, x_3, T) =$$

$$x_1(T)\Phi_3[d_1(x_1, Œ, \sigma_1^2), d_1'(x_1, x_2, \sigma_{12}^2), d_1'(x_1, x_3, \sigma_{13}^2), \rho_{112}, \rho_{113}, \rho_{123}] +$$

$$x_2(T)\Phi_3[d_1(x_2, Œ, \sigma_2^2), d_1'(x_2, x_1, \sigma_{12}^2), d_1'(x_2, x_3, \sigma_{23}^2), \rho_{221}, \rho_{223}, \rho_{213}] +$$

$$x_3(T)\Phi_3[d_1(x_3, Œ, \sigma_3^2), d_1'(x_3, x_1, \sigma_{13}^2), d_1'(x_3, x_2, \sigma_{23}^2), \rho_{331}, \rho_{332}, \rho_{312}] -$$

$$Œe^{-rT}(1 - \Phi_3[-d_2(x_1, Œ, \sigma_1^2), -d_2(x_2, Œ, \sigma_2^2), -d_2(x_3, Œ, \sigma_3^2), \rho_{12}, \rho_{13}, \rho_{23}]), \quad (4.14)$$

where

$$\begin{aligned}
d_1(x_i(T), Œ, \sigma_i^2) &= \frac{ln(\frac{x_i(T)}{Œ}) + (r + \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}} && i = 1,2,3, \\
d_2(x_i(T), Œ, \sigma_i^2) &= d_1(x_i(T), Œ, \sigma_i^2) - \sigma_i\sqrt{T} && i = 1,2,3, \\
d_1'(x_i(T), x_j(T), \sigma_{ij}^2) &= \frac{ln(\frac{x_i(T)}{x_j(T)}) + \frac{1}{2}\sigma_{ij}^2 T}{\sigma_{ij}\sqrt{T}} && i = 1,2,3; j = 1,2,3; i \neq j, \\
\sigma_{ij}^2 &= \sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j + \sigma_j^2 && i = 1,2,3; j = 1,2,3; i \neq j, \\
\rho_{iij} &= \frac{\sigma_i - \rho_{ij}\sigma_j}{\sigma_{ij}} && i = 1,2,3; j = 1,2,3; i \neq j, \\
\rho_{ijk} &= \frac{\sigma_i^2 - \rho_{ij}\sigma_i\sigma_j - \rho_{ik}\sigma_i\sigma_k + \rho_{jk}\sigma_j\sigma_k}{\sigma_{ij}\sigma_{ik}} && i = 1,2,3; j = 1,2; \\
& && k = 2,3; i \neq j; i \neq k; j \neq k,
\end{aligned}$$

and $\Phi_3[\bullet]$ is the 3-variate standardized cumulative normal distribution function.

Although equation (4.14) is an analytical formula, a numerical method must be used in order to employ the formula for the computation of option values. This is due to the fact that values have to be computed from the 3-variate standardized cumulative normal distribution function four times when using equation (4.14). The method used in this paper for the evaluation of the 3-variate standardized cumulative normal distribution function is based on numerical integration and is developed in [119].

4.5.2 European call option on the minimum of three risky assets

The price of a European call option on the minimum of three risky assets must satisfy partial differential equation (4.12) subject to

$$
\begin{aligned}
H(x_1, x_2, x_3, 0) = max[min(x_1, x_2, x_3) - Œ, 0],\\
0 \leq H(x_1, x_2, x_3, \tau) \leq min(x_1, x_2, x_3),\\
H(x_1, x_2, 0, \tau) = 0,\\
H(x_1, 0, x_3, \tau) = 0,\\
H(0, x_2, x_3, \tau) = 0,\\
H(x_1, 0, 0, \tau) = 0,\\
H(0, x_2, 0, \tau) = 0,\\
H(0, 0, x_3, \tau) = 0,\\
H(0, 0, 0, \tau) = 0.
\end{aligned}
\tag{4.15}
$$

In this case also, Johnson's method should be used to deduce an analytical formula for the price of a European call option on the minimum of three risky assets, rather than trying to solve equation (4.12) subject to conditions (4.15) directly. This leads to that the analytical formula can be written as (see also[75])

$$
\begin{aligned}
H(x_1, x_2, x_3, T) = \\
x_1(T)\Phi_3[d_1(x_1, Œ, \sigma_1^2), -d_1'(x_1, x_2, \sigma_{12}^2), -d_1'(x_1, x_3, \sigma_{13}^2), -\rho_{112}, -\rho_{113}, \rho_{123}] + \\
x_2(T)\Phi_3[d_1(x_2, Œ, \sigma_2^2), -d_1'(x_2, x_1, \sigma_{12}^2), -d_1'(x_2, x_3, \sigma_{23}^2), -\rho_{221}, -\rho_{223}, \rho_{213}] + \\
x_3(T)\Phi_3[d_1(x_3, Œ, \sigma_3^2), -d_1'(x_3, x_1, \sigma_{13}^2), -d_1'(x_3, x_2, \sigma_{23}^2), -\rho_{331}, -\rho_{332}, \rho_{312}] - \\
Œe^{-rT}\Phi_3[d_2(x_1, Œ, \sigma_1^2), d_2(x_2, Œ, \sigma_2^2), d_2(x_3, Œ, \sigma_3^2), \rho_{12}, \rho_{13}, \rho_{23}],
\end{aligned}
\tag{4.16}
$$

where $d_1(x_i(T), Œ, \sigma_i^2), d_1'(x_i(T), x_j(T), \sigma_{ij}^2), d_2(x_i(T), Œ, \sigma_i^2), \sigma_{ij}^2, \rho_{iij}, \rho_{ijk}$ and $\Phi_3[\bullet]$ are defined in equation (4.14).

4.6 A transformation of the fundamental partial differential equation

This paper shows how to value contingent claims with three underlying state variables, by means of two different finite difference schemes. These methods are straightforward and therefore easy to understand, even for those who are not experts in finite difference methods.

A price must be paid, however, for the use of such simple finite difference schemes, and in some cases this is bad stability properties. Introductory evaluations indicated that these simple finite difference schemes have such bad stability properties that they are hardly useful when used directly on the fundamental partial differential equation (4.12).

The stability problems of the simple finite difference schemes in this paper clearly motivate a search for finite difference schemes that have better stability properties. There are several other possibilities to solve numerically three-dimensional partial differential equations. An excellent overview of such finite difference schemes is given in [90]. However, a finite difference scheme with good stability properties which could be used directly for the fundamental partial differential equation (4.12) would probably be rather complex, and therefore hard to understand for non-experts in numerical methods. It could, however, be worthwhile to try to find a more advanced finite difference scheme in a later investigation.

An approach that makes it possible to use the finite difference schemes in this paper is to use logarithmic transformations of the underlying state variables. The reason for using this type of transformations is that the type has been shown to be successful in the one-dimensional case (see e.g., [22] p.462, [54] or [69]).

Introduce the transformations

$$\begin{aligned} y_1 &= ln(x_1), \\ y_2 &= ln(x_2), \\ y_3 &= ln(x_3), \end{aligned} \tag{4.17}$$

and the relation

$$H(x_1, x_2, x_3, \tau) = G(y_1, y_2, y_3, \tau). \tag{4.18}$$

With equations (4.17) and (4.18) the partial derivatives become

$$\begin{aligned}
H_{x_i} &= G_{y_i}\frac{1}{x_i} && i = 1,2,3, \\
H_{x_i x_j} &= G_{y_i y_j}\frac{1}{x_i x_j} && i = 1,2,3; j = 1;2,3; i \neq j, \\
H_{x_i x_i} &= \frac{1}{x_i^2}(G_{y_i y_i} - G_{y_i}) && i = 1,2,3, \\
H_\tau &= G_\tau.
\end{aligned} \tag{4.19}$$

Insert equations (4.19) into the fundamental partial differential equation (4.12) and the result is

$$\begin{aligned}
&\frac{1}{2}G_{y_1y_1}\sigma_1^2 + \frac{1}{2}G_{y_2y_2}\sigma_2^2 + \frac{1}{2}G_{y_3y_3}\sigma_3^2 + \\
&G_{y_1y_2}\sigma_1\sigma_2\rho_{12} + G_{y_1y_3}\sigma_1\sigma_3\rho_{13} + G_{y_2y_3}\sigma_2\sigma_3\rho_{23} + \\
&(r - \frac{\sigma_1^2}{2})G_{y_1} + (r - \frac{\sigma_2^2}{2})G_{y_2} + (r - \frac{\sigma_3^2}{2})G_{y_3} - G_\tau = rG.
\end{aligned} \tag{4.20}$$

The value of a contingent claim with three underlying state variables must satisfy partial differential equation (4.20) subject to the relevant side conditions. Naturally, these side conditions are the same side conditions as the fundamental partial differential equation (4.12) is subject to, but with logarithmic transformed state variables.

4.7 The finite difference schemes

In this paper, the standard assumption has been made that the dynamics for the prices of the underlying assets follow geometric Brownian motions. That is, it is assumed that the price dynamics are

$$dS_i = \mu_i S_i dt + \sigma_i S_i dz_i \qquad i = 1,2,3, \tag{4.21}$$

where the notation in equation (4.21) is explained in assumption **A.7** in section 4.4. The solution to equation (4.21) is[9]

$$S_i(t) = S_i(t_0)e^{[(\mu_i - \frac{\sigma_i^2}{2})(t-t_0)+\sigma_i z_i]} \qquad i = 1, 2, 3. \tag{4.23}$$

Equation (4.23) shows two facts. Firstly, $S_i(t) > 0$ for all $t > t_0$ if $S_i(t_0) > 0$. [Equation (4.23) also shows that if $S_i(t_0) = 0$ then $S_i(t) = 0$ for all $t > t_0$.] Secondly, equation (4.23) shows that there are no finite upper limits for the underlying assets' prices (see also [35], [42] and [89]). These two facts imply that $S_i(t) \in \,]0, \infty[\;\; i = 1, 2, 3; t > t_0$. [It is assumed that $S_i(t_0) > 0 \;\; i = 1, 2, 3$.] Thus, since $S_i(t) = x_i(\tau)$ it must be concluded that

$$\begin{aligned} x_1(\tau) &\in \;]0, \infty[, \\ x_2(\tau) &\in \;]0, \infty[, \\ x_3(\tau) &\in \;]0, \infty[, \\ \tau &\in \;[0, T]. \end{aligned} \tag{4.24}$$

In this paper, however, the finite difference schemes are applied to partial differential equation (4.20). This equation is derived with the natural logarithms of the underlying assets' prices as state variables. This means that region (4.24) implies that the partial differential equation (4.20) is defined in the following region:

$$\begin{aligned} y_1(\tau) &\in \;]-\infty, \infty[, \\ y_2(\tau) &\in \;]-\infty, \infty[, \\ y_3(\tau) &\in \;]-\infty, \infty[, \\ \tau &\in \;[0, T]. \end{aligned} \tag{4.25}$$

---

[9]For the purpose of confirming that equation (4.23) is the solution to equation (4.21) simply apply Itô's lemma to equation (4.23). This gives

$$dS_i = \frac{\partial S_i}{\partial t}dt + \frac{\partial S_i}{\partial z_i}dz_i + \frac{1}{2}\frac{\partial^2 S_i}{\partial z_i^2}dz_i^2 = S_i(t)[\mu_i dt + \sigma_i dz_i] \quad i = 1, 2, 3, \tag{4.22}$$

which is equation (4.21).

For the purpose of using the finite difference schemes in this paper, the unbounded region (4.25) is approximated with a bounded region. To accomplish this, upper limits for the natural logarithms of the underlying assets' prices, R_1^U, R_2^U and R_3^U, are set so high that the probabilities that the natural logarithms of the prices will exceed these limits are very small. Lower limits, R_1^L, R_2^L and R_3^L, are also set so low (i.e., negative with high absolute values) that the probabilities that the natural logarithms of the underlying assets' prices will fall below these limits are very small. In other words, if the upper and the lower limits are set sufficiently large and small respectively, then the true region (4.25) will be very well approximated by

$$\begin{aligned} y_1(\tau) &\in [R_1^L, R_1^U], \\ y_2(\tau) &\in [R_2^L, R_2^U], \\ y_3(\tau) &\in [R_3^L, R_3^U], \\ \tau &\in [0, T]. \end{aligned} \tag{4.26}$$

Furthermore, in order to be able to use the finite difference schemes in this paper, something has to be specified about the contingent claim values at the boundaries (approximated) for $y_1(\tau), y_2(\tau)$ and $y_3(\tau)$. The "simplest" approach is to use the limiting contingent claim values as the natural logarithms of the underlying assets' prices approach infinity and minus infinity respectively[10]. That is, at the boundary where $y_1 = R_1^U$ the limiting contingent claim value as $y_1 \longrightarrow \infty$ is used and so on.

It should be noted that for many contingent claims it is more straightforward to specify conditions on the first derivatives or the second derivatives of the contingent claim value with respect to the prices of the underlying assets at some or all boundaries. It is this approach that is used in the numerical evaluations later on in this paper. Conditions directly on the contingent claim value at the boundaries give, however, the simplest and thus the most comprehensible description of the finite difference schemes. First, the finite

[10]The simplicity of this approach is that boundary conditions imposed directly on the value of the contingent claim are arrived at. The limiting contingent claim values themselves are, of course, often not easy to derive.

difference schemes will be described, assuming that values of the contingent claim at all boundaries have been derived. Then when the schemes are well understood, it is easy to adjust the schemes to handle other types of boundary conditions.

An example of adjustments of the finite difference schemes is given in subsection 4.7.5, where the schemes with exactly the same type of derivative boundary conditions that are used in the numerical evaluations in subsections 4.8.2 and 4.8.3 are described.

To begin with, it is thus assumed that values of the contingent claim at the boundaries (approximated) for $y_1(\tau), y_2(\tau)$ and $y_3(\tau)$ are derived, possibly by the "simple" approach described above, when describing the finite difference schemes.

The following notation will be used in the description of the finite difference schemes:

$$
\begin{aligned}
&\Delta y_1 = h_1, \\
&\Delta y_2 = h_2, \\
&\Delta y_3 = h_3, \\
&\Delta \tau = k, \\
&R_1^U = R_1^L + M_1 h_1, \\
&R_2^U = R_2^L + M_2 h_2, \\
&R_3^U = R_3^L + M_3 h_3, \\
&T = Nk, \\
&G(R_1^L + ih_1, R_2^L + jh_2, R_3^L + lh_3, sk) = G(i,j,l,s) \qquad && i \in \{0,1,2,\ldots,M_1-1,M_1\}, \\
&G(i,j,l,s) \approx G_{i,j,l}^s && j \in \{0,1,2,\ldots,M_2-1,M_2\}, \\
& && l \in \{0,1,2,\ldots,M_3-1,M_3\}, \\
& && s \in \{0,1,2,\ldots,N-1,N\}.
\end{aligned}
\tag{4.27}
$$

In equation (4.27), $G(i,j,l,s)$ denotes the exact value and $G_{i,j,l}^s$ denotes the approximate value calculated with the help of a finite difference scheme at grid point i,j,l at time step s.

For the description of the finite difference schemes, the initial and boundary conditions to partial differential equation (4.20) are introduced as follows:

Initial condition:

$$G(y_1, y_2, y_3, 0) = f(y_1, y_2, y_3). \tag{4.28}$$

Denote the discrete version of the initial condition by

$$f(R_1^L + ih_1, R_2^L + jh_2, R_3^L + lh_3) = f_{i,j,l}. \tag{4.29}$$

The true regions of the underlying state variables have been approximated with bounded regions. Some numbers must also be set for the contingent claim values at the approximated boundaries for $y_1(\tau), y_2(\tau)$ and $y_3(\tau)$, when $0 < \tau \leq T$. Since these contingent claim values are almost always approximations, it would be more correct to use $\approx$ instead of $=$ in equations (4.30), (4.32) and (4.34). But for the purpose of getting a more straightforward description of the finite difference schemes, the notation will not be burdened with this fact.

Sides:

$$\begin{aligned}
G(y_1, y_2, R_3^L, \tau) &= s1(y_1, y_2, \tau), \\
G(y_1, R_2^L, y_3, \tau) &= s2(y_1, y_3, \tau), \\
G(R_1^U, y_2, y_3, \tau) &= s3(y_2, y_3, \tau), \\
G(y_1, R_2^U, y_3, \tau) &= s4(y_1, y_3, \tau), \\
G(R_1^L, y_2, y_3, \tau) &= s5(y_2, y_3, \tau), \\
G(y_1, y_2, R_3^U, \tau) &= s6(y_1, y_2, \tau).
\end{aligned} \tag{4.30}$$

Denote the discrete versions of the condititions for the sides in the following way:

$$\begin{aligned}
sI(R_1^L + ih_1, R_2^L + jh_2, sk) &= sI_{i,j}^s \qquad I = 1, 6, \\
sI(R_1^L + ih_1, R_3^L + lh_3, sk) &= sI_{i,l}^s \qquad I = 2, 4, \\
sI(R_2^L + jh_2, R_3^L + lh_3, sk) &= sI_{j,l}^s \qquad I = 3, 5.
\end{aligned} \tag{4.31}$$

Edges:

$$G(y_1, R_2^L, R_3^L, \tau) = k1(y_1, \tau),$$

$$
\begin{aligned}
G(R_1^U, y_2, R_3^L, \tau) &= k2(y_2, \tau), \\
G(y_1, R_2^U, R_3^L, \tau) &= k3(y_1, \tau), \\
G(R_1^L, y_2, R_3^L, \tau) &= k4(y_2, \tau), \\
G(R_1^L, R_2^L, y_3, \tau) &= k5(y_3, \tau), \\
G(R_1^U, R_2^L, y_3, \tau) &= k6(y_3, \tau), \\
G(R_1^U, R_2^U, y_3, \tau) &= k7(y_3, \tau), \\
G(R_1^L, R_2^U, y_3, \tau) &= k8(y_3, \tau), \\
G(y_1, R_2^L, R_3^U, \tau) &= k9(y_1, \tau), \\
G(R_1^U, y_2, R_3^U, \tau) &= k10(y_2, \tau), \\
G(y_1, R_2^U, R_3^U, \tau) &= k11(y_1, \tau), \\
G(R_1^L, y_2, R_3^U, \tau) &= k12(y_2, \tau).
\end{aligned} \tag{4.32}
$$

Denote the discrete versions of the condititions for the edges in the following way:

$$
\begin{aligned}
kI(R_1^L + ih_1, sk) &= kI_i^s \qquad I = 1, 3, 9, 11, \\
kI(R_2^L + jh_2, sk) &= kI_j^s \qquad I = 2, 4, 10, 12, \\
kI(R_3^L + lh_3, sk) &= kI_l^s \qquad I = 5, 6, 7, 8.
\end{aligned} \tag{4.33}
$$

Corners:

The values at the corners are never used by the finite difference schemes proposed in this paper[11]. Functions for the values at the corners are, in any event, introduced for the purpose of gaining a complete description of the values at the boundaries of the (approximate) region of the underlying state variables.

$$
\begin{aligned}
G(R_1^L, R_2^L, R_3^L, \tau) &= h1(\tau), \\
G(R_1^U, R_2^L, R_3^L, \tau) &= h2(\tau), \\
G(R_1^U, R_2^U, R_3^L, \tau) &= h3(\tau),
\end{aligned}
$$

[11]The values at the corners may, however, be necessary for other finite difference schemes.

$$
\begin{aligned}
G(R_1^L, R_2^U, R_3^L, \tau) &= h4(\tau), \qquad (4.34)\\
G(R_1^L, R_2^L, R_3^U, \tau) &= h5(\tau),\\
G(R_1^U, R_2^L, R_3^U, \tau) &= h6(\tau),\\
G(R_1^U, R_2^U, R_3^U, \tau) &= h7(\tau),\\
G(R_1^L, R_2^U, R_3^U, \tau) &= h8(\tau).
\end{aligned}
$$

Denote the discrete versions of the condititions for the corners by

$$hI(sk) = hI^s \qquad I = 1, 2, 3, 4, 5, 6, 7, 8. \qquad (4.35)$$

The boundary conditions for a given time step are graphically illustrated in figure 4.1.

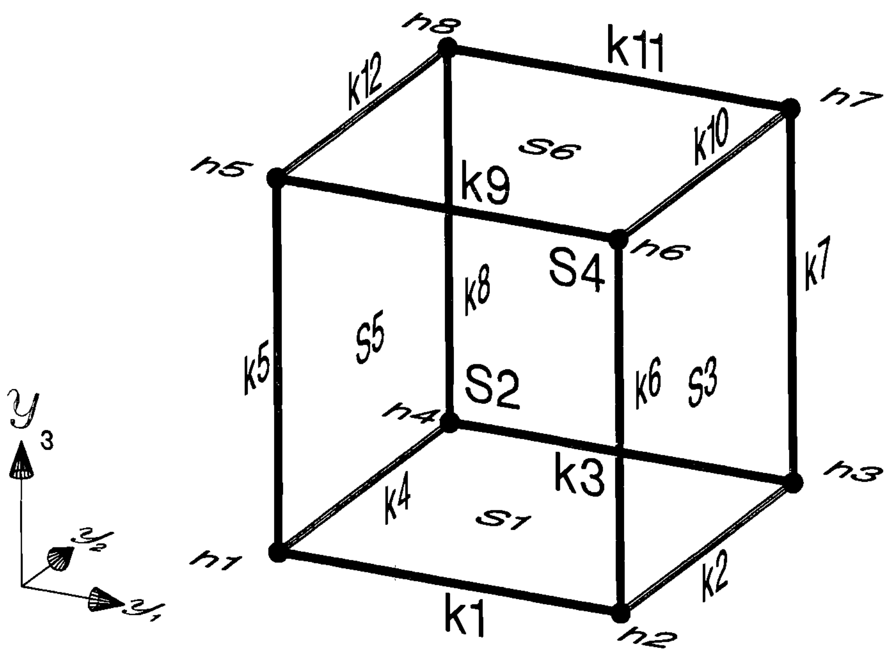

Figure 4.1: Graphical illustration of the boundary conditions.

4.7.1 Difference quotients and parameters

To obtain finite difference approximations to partial differential equation (4.20), the following difference quotients are introduced[12]:

$$
\begin{aligned}
\partial_{y_1} G(i,j,l,s) &= \frac{[G(i+1,j,l,s)-G(i-1,j,l,s)]}{2h_1} = G_{y_1}(i,j,l,s)+O(h_1^2),\\
\partial_{y_2} G(i,j,l,s) &= \frac{[G(i,j+1,l,s)-G(i,j-1,l,s)]}{2h_2} = G_{y_2}(i,j,l,s)+O(h_2^2),\\
\partial_{y_3} G(i,j,l,s) &= \frac{[G(i,j,l+1,s)-G(i,j,l-1,s)]}{2h_3} = G_{y_3}(i,j,l,s)+O(h_3^2),\\
\partial^f_{y_1}\partial^b_{y_1} G(i,j,l,s) &= \frac{[G(i+1,j,l,s)-2G(i,j,l,s)+G(i-1,j,l,s)]}{h_1^2} =\\
&= G_{y_1y_1}(i,j,l,s)+O(h_1^2),\\
\partial^f_{y_2}\partial^b_{y_2} G(i,j,l,s) &= \frac{[G(i,j+1,l,s)-2G(i,j,l,s)+G(i,j-1,l,s)]}{h_2^2} =\\
&= G_{y_2y_2}(i,j,l,s)+O(h_2^2),\\
\partial^f_{y_3}\partial^b_{y_3} G(i,j,l,s) &= \frac{[G(i,j,l+1,s)-2G(i,j,l,s)+G(i,j,l-1,s)]}{h_3^2} =\\
&= G_{y_3y_3}(i,j,l,s)+O(h_3^2),
\end{aligned}
\tag{4.36}
$$

$$
\begin{aligned}
&\partial_{y_1}\partial_{y_2} G(i,j,l,s) =\\
&= \frac{[G(i+1,j+1,l,s)-G(i-1,j+1,l,s)-G(i+1,j-1,l,s)+G(i-1,j-1,l,s)]}{4h_1h_2} =\\
&= G_{y_1y_2}(i,j,l,s)+O(\frac{(max[h_1,h_2])^4}{h_1h_2}),\\
&\partial_{y_1}\partial_{y_3} G(i,j,l,s) =\\
&= \frac{[G(i+1,j,l+1,s)-G(i-1,j,l+1,s)-G(i+1,j,l-1,s)+G(i-1,j,l-1,s)]}{4h_1h_3} =\\
&= G_{y_1y_3}(i,j,l,s)+O(\frac{(max[h_1,h_3])^4}{h_1h_3}),\\
&\partial_{y_2}\partial_{y_3} G(i,j,l,s) =\\
&\frac{[G(i,j+1,l+1,s)-G(i,j-1,l+1,s)-G(i,j+1,l-1,s)+G(i,j-1,l-1,s)]}{4h_2h_3} =\\
&= G_{y_2y_3}(i,j,l,s)+O(\frac{(max[h_2,h_3])^4}{h_2h_3}),
\end{aligned}
$$

$$
\partial^b_\tau G(i,j,l,s) = \frac{[G(i,j,l,s)-G(i,j,l,s-1)]}{k} = G_\tau(i,j,l,s)+O(k),
$$

[12]It will be assumed that the operators $\partial_{y_1}, \partial_{y_2}, \ldots$, work exactly the same way at the discrete approximations $G^s_{i,j,l}$ as at the exact values $G(i,j,l,s)$, i.e., $\partial_{y_1} G^s_{i,j,l} = \frac{G^s_{i+1,j,l}-G^s_{i-1,j,l}}{2h_1}$ etc.

$$\partial^f_\tau G(i,j,l,s) = \frac{[G(i,j,l,s+1)-G(i,j,l,s)]}{k} = G_\tau(i,j,l,s)+O(k).$$

$\partial^b_\bullet$ denotes a backward difference approximation, $\partial^f_\bullet$ a forward difference approximation and $\partial_\bullet$ a central difference approximation. $O(\bullet)$ represents the order of the errors in the difference approximations.

Before starting with the descriptions of the finite difference schemes the following parameters are defined:

$$\begin{aligned}
a^{y_1} &= \frac{1}{2}[(r-\frac{\sigma_1^2}{2})\frac{1}{h_1}-\frac{\sigma_1^2}{h_1^2}],\\
b^{y_1} &= r+\frac{1}{k}+\frac{\sigma_1^2}{h_1^2},\\
c^{y_1} &= -\frac{1}{2}[(r-\frac{\sigma_1^2}{2})\frac{1}{h_1}+\frac{\sigma_1^2}{h_1^2}],\\
a^{y_2} &= \frac{1}{2}[(r-\frac{\sigma_2^2}{2})\frac{1}{h_2}-\frac{\sigma_2^2}{h_2^2}],\\
b^{y_2} &= r+\frac{1}{k}+\frac{\sigma_2^2}{h_2^2},\\
c^{y_2} &= -\frac{1}{2}[(r-\frac{\sigma_2^2}{2})\frac{1}{h_2}+\frac{\sigma_2^2}{h_2^2}],\\
a^{y_3} &= \frac{1}{2}[(r-\frac{\sigma_3^2}{2})\frac{1}{h_3}-\frac{\sigma_3^2}{h_3^2}],\\
b^{y_3} &= r+\frac{1}{k}+\frac{\sigma_3^2}{h_3^2},\\
c^{y_3} &= -\frac{1}{2}[(r-\frac{\sigma_3^2}{2})\frac{1}{h_3}+\frac{\sigma_3^2}{h_3^2}],\\
d &= \frac{1}{k}-\frac{\sigma_1^2}{h_1^2}-\frac{\sigma_2^2}{h_2^2}-\frac{\sigma_3^2}{h_3^2},\\
d^{y_1y_2} &= \frac{1}{k}-\frac{\sigma_1^2}{h_1^2}-\frac{\sigma_2^2}{h_2^2},\\
d^{y_1y_3} &= \frac{1}{k}-\frac{\sigma_1^2}{h_1^2}-\frac{\sigma_3^2}{h_3^2},\\
d^{y_2y_3} &= \frac{1}{k}-\frac{\sigma_2^2}{h_2^2}-\frac{\sigma_3^2}{h_3^2},\\
e^{y_1y_2} &= \frac{\sigma_1\sigma_2\rho_{12}}{4h_1h_2},\\
e^{y_1y_3} &= \frac{\sigma_1\sigma_3\rho_{13}}{4h_1h_3},\\
e^{y_2y_3} &= \frac{\sigma_2\sigma_3\rho_{23}}{4h_2h_3}.
\end{aligned} \tag{4.37}$$

4.7.2 The explicit finite difference scheme

With the help of the difference operators in equation (4.36), discrete approximations can be obtained to partial differential equation (4.20). The discrete approximation to use in the explicit finite difference scheme is the following:

$$\frac{1}{2}\partial^f_{y_1}\partial^b_{y_1}G^s_{i,j,l}\sigma_1^2 + \frac{1}{2}\partial^f_{y_2}\partial^b_{y_2}G^s_{i,j,l}\sigma_2^2 + \frac{1}{2}\partial^f_{y_3}\partial^b_{y_3}G^s_{i,j,l}\sigma_3^2 +$$
$$\partial_{y_1}\partial_{y_2}G^s_{i,j,l}\sigma_1\sigma_2\rho_{12} + \partial_{y_1}\partial_{y_3}G^s_{i,j,l}\sigma_1\sigma_3\rho_{13} + \partial_{y_2}\partial_{y_3}G^s_{i,j,l}\sigma_2\sigma_3\rho_{23} +$$
$$\partial_{y_1}G^s_{i,j,l}(r - \frac{\sigma_1^2}{2}) + \partial_{y_2}G^s_{i,j,l}(r - \frac{\sigma_2^2}{2}) + \partial_{y_3}G^s_{i,j,l}(r - \frac{\sigma_3^2}{2}) - rG^{s+1}_{i,j,l} = \partial^f_\tau G^s_{i,j,l}. \quad (4.38)$$

Equation (4.38) can be restated as

$$\begin{aligned}
&G^{s+1}_{i,j,l} = \\
&\frac{k}{1+rk}[-a^{y_1}G^s_{i-1,j,l} - c^{y_1}G^s_{i+1,j,l} - a^{y_2}G^s_{i,j-1,l} - \\
&c^{y_2}G^s_{i,j+1,l} - a^{y_3}G^s_{i,j,l-1} - c^{y_3}G^s_{i,j,l+1} + dG^s_{i,j,l} + \\
&e^{y_1y_2}(G^s_{i+1,j+1,l} - G^s_{i-1,j+1,l} - G^s_{i+1,j-1,l} + G^s_{i-1,j-1,l}) + \\
&e^{y_1y_3}(G^s_{i+1,j,l+1} - G^s_{i-1,j,l+1} - G^s_{i+1,j,l-1} + G^s_{i-1,j,l-1}) + \\
&e^{y_2y_3}(G^s_{i,j+1,l+1} - G^s_{i,j-1,l+1} - G^s_{i,j+1,l-1} + G^s_{i,j-1,l-1})],
\end{aligned} \quad (4.39)$$

where $a^{y_1}, c^{y_1}, a^{y_2}, c^{y_2}, a^{y_3}, c^{y_3}, d, e^{y_1y_2}, e^{y_1y_3}$ and $e^{y_2y_3}$ are defined in equation (4.37).

Equation (4.39) shows how to obtain a solution to the approximate contingent claim value $G^{s+1}_{i,j,l}$ in time step $s+1$ in terms of approximate contingent claim values in time step s. The solution methodology is now as follows:

The initial condition (4.29) gives the contingent claim values when $s = 0$. In equation (4.26), the region for the logarithms of the underlying assets' prices is approximated with a cube, for a given value of s. With the help of the initial condition, equation (4.39) can be used to solve all the approximate contingent claim values at the inner grid points of the cube, when $s = 1$. In order to obtain solutions to the approximate contingent claim values in the next time step, the values at the side and edges of the cube are also needed, when $s = 1$. These values are provided by equations (4.31) and (4.33). With the approximate contingent claim values at time step $s = 1$, equation (4.39) can be used to obtain solutions to the approximate contingent claim values at the inner grid points for time step $s = 2$, and so forth and so on. This continues until the approximate contingent claim values at time step $s = N$ are found. The desired contingent claim value is $G^N_{i_0,j_0,l_0}$ where $ln(x_1(T)) = R_1^L + i_0 h_1, ln(x_2(T)) = R_2^L + j_0 h_2$ and $ln(x_3(T)) = R_3^L + l_0 h_3$, and $x_1(T), x_2(T)$ and $x_3(T)$ are the current prices of the underlying assets.

Although equation (4.39) provides a fairly simple formula for solving all the approximate contingent claim values at the inner grid points in each iteration (through time), the method is computationally laborious. As an example, consider a relatively modest size of the grid, $M_1 = M_2 = M_3 = 201$, and relatively few time steps, $N = 100$. These modelling decisions require equation (4.39) to be solved no less than 8×10^8 times. (In addition to this, the initial condition has to be computed at every grid point when $s = 0$, and the sides and edges of the cube have to be filled up in each time step.)

All solutions to equation (4.39) for a given value of s can, however, be computed in parallel. Using a massively parallel computer like the CM-2000 Connection Machine, as many grid points as there are processors can thus be solved at the same time.

4.7.3 The discrete approximations when using the generalized ADI-method

The generalized ADI-method[13] works iteratively through time, exactly as the explicit finite difference scheme in subsection 4.7.2. Thus, given the (approximate) contingent claim values in time step s, the (approximate) contingent claim values in time step $s + 1$ can be found. The iterations through time start when $s = 0$, where the contingent claim values can be calculated with the help of the initial condition (4.29). When the cube is filled up with the help of the initial condition, time step by time step is iterated until the

[13]In naming this finite difference scheme "the generalized ADI-method" may not be that clever. This derives from the fact that there are other somewhat different schemes that are called ADI-methods. The generalized ADI-method is related to these schemes, and an attempt will be made to distinguish the finite difference scheme in this paper by the pre-name generalized.

current time is reached, i.e., until $s = N$. The desired contingent claim value is $G^N_{i_0,j_0,l_0}$ where $ln(x_1(T)) = R^L_1 + i_0h_1, ln(x_2(T)) = R^L_2 + j_0h_2$ and $ln(x_3(T)) = R^L_3 + l_0h_3$, and $x_1(T), x_2(T)$ and $x_3(T)$ are the current prices of the underlying assets.

In a given time step, the method is (partly)[14] implicit in one of the logarithmic transformed (logarithmic transformed will, henceforth, often be abbreviated to log) price dimensions. (What is meant by an implicit finite difference scheme is explained in [2], [53] or [126].) The log price dimension for which the method is implicit changes according to a rotating schedule. If the method is implicit in the y_1-dimension in time step s, then the method is implicit in the y_2-dimension in time step $s+1$ and implicit in the y_3-dimension in time step $s+2$, and so forth and so on. To be more precise, the rotating schedule is as follows:

- The method is implicit in the y_1-dimension if $\frac{(s-1)}{3} = integer$.
- The method is implicit in the y_2-dimension if $\frac{(s-2)}{3} = integer$.
- The method is implicit in the y_3-dimension if $\frac{s}{3} = integer$.

In order to find an approximate solution to equation (4.20) subject to conditions (4.28), (4.30), (4.32) and (4.34) with the generalized ADI-method, a discrete approximation to equation (4.20) must be found. The discrete approximation changes, however, according to the rotating schedule just described.

4.7.3.1 The discrete approximation when $\frac{(s-1)}{3} = integer$

With the help of the difference operators in equation (4.36), discrete approximations to equation (4.20) can be obtained. When the method is implicit in the y_1-dimension equation (4.20) is approximated in the following way:

$$\frac{1}{2}\partial^f_{y_1}\partial^b_{y_1}G^s_{i,j,l}\sigma_1^2 + \frac{1}{2}\partial^f_{y_2}\partial^b_{y_2}G^{s-1}_{i,j,l}\sigma_2^2 + \frac{1}{2}\partial^f_{y_3}\partial^b_{y_3}G^{s-1}_{i,j,l}\sigma_3^2 +$$
$$\partial_{y_1}\partial_{y_2}G^{s-1}_{i,j,l}\sigma_1\sigma_2\rho_{12} + \partial_{y_1}\partial_{y_3}G^{s-1}_{i,j,l}\sigma_1\sigma_3\rho_{13} + \partial_{y_2}\partial_{y_3}G^{s-1}_{i,j,l}\sigma_2\sigma_3\rho_{23} +$$

[14]The generalized ADI-method is only partly implicit in the current implicit dimension since the method is always explicit in the cross derivatives.

$$\partial_{y_1} G^s_{i,j,l}(r - \frac{\sigma_1^2}{2}) + \partial_{y_2} G^{s-1}_{i,j,l}(r - \frac{\sigma_2^2}{2}) + \partial_{y_3} G^{s-1}_{i,j,l}(r - \frac{\sigma_3^2}{2}) - rG^s_{i,j,l} = \partial^b_\tau G^s_{i,j,l}. \quad (4.40)$$

With the help of equation (4.36), equation (4.40) can be restated as

$$\begin{aligned}
&a^{y_1} G^s_{i-1,j,l} + b^{y_1} G^s_{i,j,l} + c^{y_1} G^s_{i+1,j,l} = \\
&-a^{y_2} G^{s-1}_{i,j-1,l} - c^{y_2} G^{s-1}_{i,j+1,l} - a^{y_3} G^{s-1}_{i,j,l-1} - c^{y_3} G^{s-1}_{i,j,l+1} + d^{y_2 y_3} G^{s-1}_{i,j,l} + \\
&e^{y_1 y_2}(G^{s-1}_{i+1,j+1,l} - G^{s-1}_{i-1,j+1,l} - G^{s-1}_{i+1,j-1,l} + G^{s-1}_{i-1,j-1,l}) + \\
&e^{y_1 y_3}(G^{s-1}_{i+1,j,l+1} - G^{s-1}_{i-1,j,l+1} - G^{s-1}_{i+1,j,l-1} + G^{s-1}_{i-1,j,l-1}) + \\
&e^{y_2 y_3}(G^{s-1}_{i,j+1,l+1} - G^{s-1}_{i,j-1,l+1} - G^{s-1}_{i,j+1,l-1} + G^{s-1}_{i,j-1,l-1}), \quad (4.41)
\end{aligned}$$

where $a^{y_1}, b^{y_1}, c^{y_1}, a^{y_2}, c^{y_2}, a^{y_3}, c^{y_3}, d^{y_2 y_3}, e^{y_1 y_2}, e^{y_1 y_3}$ and $e^{y_2 y_3}$ are defined in equation (4.37).

4.7.3.2 The discrete approximation when $\frac{(s-2)}{3} = integer$

When the generalized ADI-method is implicit in the y_2-dimension the discrete approximation to equation (4.20) is obtained by

$$\begin{aligned}
&\frac{1}{2}\partial^f_{y_1}\partial^b_{y_1} G^{s-1}_{i,j,l}\sigma_1^2 + \frac{1}{2}\partial^f_{y_2}\partial^b_{y_2} G^s_{i,j,l}\sigma_2^2 + \frac{1}{2}\partial^f_{y_3}\partial^b_{y_3} G^{s-1}_{i,j,l}\sigma_3^2 + \\
&\partial_{y_1}\partial_{y_2} G^{s-1}_{i,j,l}\sigma_1\sigma_2\rho_{12} + \partial_{y_1}\partial_{y_3} G^{s-1}_{i,j,l}\sigma_1\sigma_3\rho_{13} + \partial_{y_2}\partial_{y_3} G^{s-1}_{i,j,l}\sigma_2\sigma_3\rho_{23} + \\
&\partial_{y_1} G^{s-1}_{i,j,l}(r - \frac{\sigma_1^2}{2}) + \partial_{y_2} G^s_{i,j,l}(r - \frac{\sigma_2^2}{2}) + \partial_{y_3} G^{s-1}_{i,j,l}(r - \frac{\sigma_3^2}{2}) - rG^s_{i,j,l} = \partial^b_\tau G^s_{i,j,l}. \quad (4.42)
\end{aligned}$$

Equation (4.42) can be rewritten as

$$\begin{aligned}
&a^{y_2} G^s_{i,j-1,l} + b^{y_2} G^s_{i,j,l} + c^{y_2} G^s_{i,j+1,l} = \\
&-a^{y_1} G^{s-1}_{i-1,j,l} - c^{y_1} G^{s-1}_{i+1,j,l} - a^{y_3} G^{s-1}_{i,j,l-1} - c^{y_3} G^{s-1}_{i,j,l+1} + d^{y_1 y_3} G^{s-1}_{i,j,l} + \\
&e^{y_1 y_2}(G^{s-1}_{i+1,j+1,l} - G^{s-1}_{i-1,j+1,l} - G^{s-1}_{i+1,j-1,l} + G^{s-1}_{i-1,j-1,l}) + \\
&e^{y_1 y_3}(G^{s-1}_{i+1,j,l+1} - G^{s-1}_{i-1,j,l+1} - G^{s-1}_{i+1,j,l-1} + G^{s-1}_{i-1,j,l-1}) + \\
&e^{y_2 y_3}(G^{s-1}_{i,j+1,l+1} - G^{s-1}_{i,j-1,l+1} - G^{s-1}_{i,j+1,l-1} + G^{s-1}_{i,j-1,l-1}), \quad (4.43)
\end{aligned}$$

where $a^{y_1}, c^{y_1}, a^{y_2}, b^{y_2}, c^{y_2}, a^{y_3}, c^{y_3}, d^{y_1 y_3}, e^{y_1 y_2}, e^{y_1 y_3}$ and $e^{y_2 y_3}$ are defined in equation (4.37).

4.7.3.3 The discrete approximation when $\frac{s}{3} = integer$

In order to find a discrete approximation to equation (4.20) when the method is implicit in the y_3-dimension, the derivatives are approximated with difference quotients in the following way:

$$
\begin{aligned}
&\frac{1}{2}\partial_{y_1}^f\partial_{y_1}^b G_{i,j,l}^{s-1}\sigma_1^2 + \frac{1}{2}\partial_{y_2}^f\partial_{y_2}^b G_{i,j,l}^{s-1}\sigma_2^2 + \frac{1}{2}\partial_{y_3}^f\partial_{y_3}^b G_{i,j,l}^{s}\sigma_3^2 + \\
&\partial_{y_1}\partial_{y_2} G_{i,j,l}^{s-1}\sigma_1\sigma_2\rho_{12} + \partial_{y_1}\partial_{y_3} G_{i,j,l}^{s-1}\sigma_1\sigma_3\rho_{13} + \partial_{y_2}\partial_{y_3} G_{i,j,l}^{s-1}\sigma_2\sigma_3\rho_{23} + \\
&\partial_{y_1} G_{i,j,l}^{s-1}(r-\frac{\sigma_1^2}{2}) + \partial_{y_2} G_{i,j,l}^{s-1}(r-\frac{\sigma_2^2}{2}) + \partial_{y_3} G_{i,j,l}^{s}(r-\frac{\sigma_3^2}{2}) - rG_{i,j,l}^{s} = \partial_\tau^b G_{i,j,l}^{s}.
\end{aligned} \tag{4.44}
$$

With the help of equation (4.36), equation (4.44) can be rewritten as

$$
\begin{aligned}
&a^{y_3} G_{i,j,l-1}^{s} + b^{y_3} G_{i,j,l}^{s} + c^{y_3} G_{i,j,l+1}^{s} = \\
&-a^{y_1} G_{i-1,j,l}^{s-1} - c^{y_1} G_{i+1,j,l}^{s-1} - a^{y_2} G_{i,j-1,l}^{s-1} - c^{y_2} G_{i,j+1,l}^{s-1} + d^{y_1 y_2} G_{i,j,l}^{s-1} + \\
&e^{y_1 y_2}(G_{i+1,j+1,l}^{s-1} - G_{i-1,j+1,l}^{s-1} - G_{i+1,j-1,l}^{s-1} + G_{i-1,j-1,l}^{s-1}) + \\
&e^{y_1 y_3}(G_{i+1,j,l+1}^{s-1} - G_{i-1,j,l+1}^{s-1} - G_{i+1,j,l-1}^{s-1} + G_{i-1,j,l-1}^{s-1}) + \\
&e^{y_2 y_3}(G_{i,j+1,l+1}^{s-1} - G_{i,j-1,l+1}^{s-1} - G_{i,j+1,l-1}^{s-1} + G_{i,j-1,l-1}^{s-1}),
\end{aligned} \tag{4.45}
$$

where $a^{y_1}, c^{y_1}, a^{y_2}, c^{y_2}, a^{y_3}, b^{y_3}, c^{y_3}, d^{y_1 y_2}, e^{y_1 y_2}, e^{y_1 y_3}$ and $e^{y_2 y_3}$ are defined in equation (4.37).

To give a full understanding of the generalized ADI-method it is suitable to describe the method in matrix notation. Such a description is given in the next subsection.

4.7.4 The generalized ADI-method in matrix notation

As previously mentioned, the generalized ADI-method works iteratively through time. It starts with the initial condition, equation (4.29), when $s = 0$, and stops with the approximate values for G when $s = N$. Given the approximate values for G in time step $s - 1$, one of the equations (4.41), (4.43) or (4.45), dependent on which dimension is the implicit one, offers a way of finding the approximate values for G in time step s.

The method outlined involves solving $M_2 - 1 \times M_3 - 1$ tridiagonal equation systems with $M_1 - 1$ unknown variables per time step when the method is implicit in the y_1-dimension, $M_1 - 1 \times M_3 - 1$ tridiagonal equation systems with $M_2 - 1$ unknown variables when the method is implicit in the y_2-dimension and $M_1 - 1 \times M_2 - 1$ tridiagonal equation systems with $M_3 - 1$ unknown variables when the method is implicit in the y_3-dimension. This feature makes the generalized ADI-method computationally laborious. As an example, assume a relatively modest size of the grid, $M_1 = M_2 = M_3 = 201$, and relatively few time steps, $N = 100$. These modelling decisions require the solution of 4×10^6 tridiagonal equation systems, each with 200 unknown variables.

There are, however, only three different tridiagonal matrices to consider. More precisely, the tridiagonal matrices are identical for all equation systems for a given time step, and they are also identical for different time steps for which the method is implicit in the same dimension. The last fact follows from equation (4.37), where the matrix elements can be seen to be independent of time. Thus, the "solution coefficients" to the tridiagonal matrices can be calculated once and for all.

When the method is implicit in the y_1-dimension the tridiagonal matrix has the following appearance:

$$\mathbf{A_{y_1}} = \begin{bmatrix} b^{y_1} & c^{y_1} & 0 & 0 & 0 & \cdots & 0 \\ a^{y_1} & b^{y_1} & c^{y_1} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_1} & b^{y_1} & c^{y_1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_1} & b^{y_1} & c^{y_1} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_1} & b^{y_1} \end{bmatrix}. \tag{4.46}$$

The tridiagonal matrix appears as

$$\mathbf{A_{y_2}} = \begin{bmatrix} b^{y_2} & c^{y_2} & 0 & 0 & 0 & \cdots & 0 \\ a^{y_2} & b^{y_2} & c^{y_2} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_2} & b^{y_2} & c^{y_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_2} & b^{y_2} & c^{y_2} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_2} & b^{y_2} \end{bmatrix}, \tag{4.47}$$

when the method is implicit in the y_2-dimension.

Finally, when the method is implicit in the y_3-dimension the tridiagonal matrix has the following appearance:

$$\mathbf{A_{y_3}} = \begin{bmatrix} b^{y_3} & c^{y_3} & 0 & 0 & 0 & \cdots & 0 \\ a^{y_3} & b^{y_3} & c^{y_3} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_3} & b^{y_3} & c^{y_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_3} & b^{y_3} & c^{y_3} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_3} & b^{y_3} \end{bmatrix}. \tag{4.48}$$

Although the tridiagonal matrices are identical for every equation system in each implicit dimension, the method is computationally intensive. In fact, the method requires so much computing power that it can hardly be executed on an ordinary PC. Many of these equation systems can, however, be solved in parallel. To be more exact, all equation systems in a given time step can be solved in parallel. The generalized ADI-method, as well as the explicit scheme in subsection 4.7.2, is therefore very well suited to be executed on a massively parallel computer, like the CM-2000 Connection Machine.

The discrete approximation of equation (4.20) is somewhat different depending on which dimension is implicit. Thus, the equation systems when using the generalized ADI-method are also somewhat different dependent on which dimension is implicit. The following description of the equation systems consists of one sub-subsection for each implicit dimension. Sub-subsection 4.7.4.1 gives a description of the equation systems when the method is implicit in the y_1-dimension. Sub-subsection 4.7.4.2 gives a description of the equation systems when the method is implicit in the y_2-dimension, and sub-subsection 4.7.4.3 gives a description of the equation systems when the method is implicit in the y_3-dimension.

4.7.4.1 The equation systems when $\frac{(s-1)}{3} = integer$ with matrix notation

In this sub-subsection, the equation systems will be described when the method is implicit in the y_1-dimension. For the purpose of getting a more tractable description, the following matrices are introduced:

$$\mathbf{G}^{sy_1}_{j,l} = \begin{bmatrix} G^s_{1,j,l} \\ G^s_{2,j,l} \\ G^s_{3,j,l} \\ \vdots \\ G^s_{M_1-2,j,l} \\ G^s_{M_1-1,j,l} \end{bmatrix}, \tag{4.49}$$

$$\mathbf{sy1}^s_{j,l} = \begin{bmatrix} a^{y_1} s5^s_{j,l} \\ 0 \\ 0 \\ \vdots \\ 0 \\ c^{y_1} s3^s_{j,l} \end{bmatrix}. \tag{4.50}$$

$\mathbf{G}^{sy_1}_{j,l}$ and $\mathbf{sy1}^s_{j,l}$ above are $M_1 - 1 \times 1$ matrices.

The equation systems when $s = 1$

As previously mentioned, the generalized ADI-method iterates through time. In order to be able to start up this iterative procedure, the values of G must be known at some time point. The initial values, equation (4.29), give these required values when $s = 0$ (i.e., when $\tau = 0$).

When $s = 1$ the equation systems are as follows:

$$\mathbf{A_{y_1}} \mathbf{G}^{1y_1}_{j,l} = \mathbf{R}^{1y_1}_{j,l}, \tag{4.51}$$

where

$$\begin{aligned}
&\mathbf{R}^{1y_1}_{j,l} = \\
&-a^{y_2}\mathbf{f}_{j-1,l} - c^{y_2}\mathbf{f}_{j+1,l} - a^{y_3}\mathbf{f}_{j,l-1} - c^{y_3}\mathbf{f}_{j,l+1} + d^{y_2 y_3}\mathbf{f}_{j,l} \\
&+e^{y_1 y_2}(\mathbf{f^{+1}}_{j+1,l} - \mathbf{f^{-1}}_{j+1,l} - \mathbf{f^{+1}}_{j-1,l} + \mathbf{f^{-1}}_{j-1,l}) \\
&+e^{y_1 y_3}(\mathbf{f^{+1}}_{j,l+1} - \mathbf{f^{-1}}_{j,l+1} - \mathbf{f^{+1}}_{j,l-1} + \mathbf{f^{-1}}_{j,l-1}) \\
&+e^{y_2 y_3}(\mathbf{f}_{j+1,l+1} - \mathbf{f}_{j-1,l+1} - \mathbf{f}_{j+1,l-1} + \mathbf{f}_{j-1,l-1}) - \mathbf{sy1}^1_{j,l},
\end{aligned}$$

$$\mathbf{f}_{j,l} = \begin{bmatrix} f_{1,j,l} \\ f_{2,j,l} \\ f_{3,j,l} \\ \vdots \\ f_{M_1-2,j,l} \\ f_{M_1-1,j,l} \end{bmatrix},$$

$$\mathbf{f}^{-1}{}_{j,l} = \begin{bmatrix} f_{0,j,l} \\ f_{1,j,l} \\ f_{2,j,l} \\ \vdots \\ f_{M_1-3,j,l} \\ f_{M_1-2,j,l} \end{bmatrix},$$

$$\mathbf{f}^{+1}{}_{j,l} = \begin{bmatrix} f_{2,j,l} \\ f_{3,j,l} \\ f_{4,j,l} \\ \vdots \\ f_{M_1-1,j,l} \\ f_{M_1,j,l} \end{bmatrix}.$$

The equation systems in equation (4.51) hold for $j \in \{1, \ldots, M_2-1\}$ and $l \in \{1, \ldots, M_3 - 1\}$, i.e., for $M_2 - 1 \times M_3 - 1$ tridiagonal equation systems.

From equation (4.26) it is known that $G(y_1, y_2, y_3, \tau)$ is approximated to be defined on a cube in the log price dimensions for a given τ. The solution to equation (4.51) gives the (approximate) solutions for the contingent claim value at all inner grid points of this cube for $\tau = k$. In order to solve the approximate values in the next time step, the values on the sides and edges of the cube also have to be known. These values are given by equations (4.31) and (4.33). That is, when all the approximate values at the inner grid points have been solved with the help of equation (4.51), the sides and edges of the cube are filled up with the help of equations (4.31) and (4.33).

The equation systems when $s > 1$ and $\frac{(s-1)}{3} = integer$

When $s > 1$, $\frac{(s-1)}{3} = integer$, $j \in \{1, \ldots, M_2-1\}$ and $l \in \{1, \ldots, M_3-1\}$ the unknown values $G^s_{1,j,l}, \ldots, G^s_{M_1-1,j,l}$ can be obtained by solving the following $M_2-1 \times M_3-1$ equation systems:

$$\mathbf{A_{y_1}} \mathbf{G}^{sy_1}_{j,l} = \mathbf{R}^{sy_1}_{j,l}, \tag{4.52}$$

where

$$\begin{aligned}
&\mathbf{R}^{sy_1}_{j,l} = \\
&-a^{y_2}\mathbf{G}^{(s-1)y_1}_{j-1,l} - c^{y_2}\mathbf{G}^{(s-1)y_1}_{j+1,l} - a^{y_3}\mathbf{G}^{(s-1)y_1}_{j,l-1} - c^{y_3}\mathbf{G}^{(s-1)y_1}_{j,l+1} + d^{y_2y_3}\mathbf{G}^{(s-1)y_1}_{j,l} \\
&+e^{y_1y_2}(\mathbf{G+1}^{(s-1)y_1}_{j+1,l} - \mathbf{G-1}^{(s-1)y_1}_{j+1,l} - \mathbf{G+1}^{(s-1)y_1}_{j-1,l} + \mathbf{G-1}^{(s-1)y_1}_{j-1,l}) \\
&+e^{y_1y_3}(\mathbf{G+1}^{(s-1)y_1}_{j,l+1} - \mathbf{G-1}^{(s-1)y_1}_{j,l+1} - \mathbf{G+1}^{(s-1)y_1}_{j,l-1} + \mathbf{G-1}^{(s-1)y_1}_{j,l-1}) \\
&+e^{y_2y_3}(\mathbf{G}^{(s-1)y_1}_{j+1,l+1} - \mathbf{G}^{(s-1)y_1}_{j-1,l+1} - \mathbf{G}^{(s-1)y_1}_{j+1,l-1} + \mathbf{G}^{(s-1)y_1}_{j-1,l-1}) - \mathbf{sy1}^s_{j,l},
\end{aligned}$$

$$\mathbf{G-1}^{sy_1}_{j,l} = \begin{bmatrix} G^s_{0,j,l} \\ G^s_{1,j,l} \\ G^s_{2,j,l} \\ \vdots \\ G^s_{M_1-3,j,l} \\ G^s_{M_1-2,j,l} \end{bmatrix}, \tag{4.53}$$

$$\mathbf{G+1}^{sy_1}_{j,l} = \begin{bmatrix} G^s_{2,j,l} \\ G^s_{3,j,l} \\ G^s_{4,j,l} \\ \vdots \\ G^s_{M_1-1,j,l} \\ G^s_{M_1,j,l} \end{bmatrix}. \tag{4.54}$$

When all the (approximate) values at the inner grid points have been solved with the help of equation (4.52), the sides and edges of the cube are filled up with the help of equations (4.31) and (4.33).

4.7.4.2 The equation systems when $\frac{(s-2)}{3} = integer$ with matrix notation

In this sub-subsection the equation systems will be described when the method is implicit in the y_2-dimension. In order to make the description easier to follow, the following matrices are introduced:

$$\mathbf{G}^{sy_2}_{i,l} = \begin{bmatrix} G^s_{i,1,l} \\ G^s_{i,2,l} \\ G^s_{i,3,l} \\ \vdots \\ G^s_{i,M_2-2,l} \\ G^s_{i,M_2-1,l} \end{bmatrix}, \tag{4.55}$$

$$\mathbf{G^{-1}}^{sy_2}_{i,l} = \begin{bmatrix} G^s_{i,0,l} \\ G^s_{i,1,l} \\ G^s_{i,2,l} \\ \vdots \\ G^s_{i,M_2-3,l} \\ G^s_{i,M_2-2,l} \end{bmatrix}, \tag{4.56}$$

$$\mathbf{G^{+1}}^{sy_2}_{i,l} = \begin{bmatrix} G^s_{i,2,l} \\ G^s_{i,3,l} \\ G^s_{i,4,l} \\ \vdots \\ G^s_{i,M_2-1,l} \\ G^s_{i,M_2,l} \end{bmatrix}, \tag{4.57}$$

$$\mathbf{sy2}^s_{i,l} = \begin{bmatrix} a^{y_2} s2^s_{i,l} \\ 0 \\ 0 \\ \vdots \\ 0 \\ c^{y_2} s4^s_{i,l} \end{bmatrix}, \tag{4.58}$$

where $\mathbf{G}^{sy_2}_{i,l}, \mathbf{G}^{-1}{}^{sy_2}_{i,l}, \mathbf{G}^{+1}{}^{sy_2}_{i,l}$ and $\mathbf{sy2}^s_{i,l}$ are $M_2 - 1 \times 1$ matrices.

When $s > 1$, $\frac{(s-2)}{3} = integer$, $i \in \{1, \ldots, M_1 - 1\}$ and $l \in \{1, \ldots, M_3 - 1\}$, the unknown values $G^s_{i,1,l}, \ldots, G^s_{i,M_2-1,l}$ can be obtained by solving following $M_1 - 1 \times M_3 - 1$ equation systems:

$$\mathbf{A_{y_2}} \mathbf{G}^{sy_2}_{i,l} = \mathbf{R}^{sy_2}_{i,l}, \tag{4.59}$$

where

$$\begin{aligned}
&\mathbf{R}^{sy_2}_{i,l} = \\
&-a^{y_1}\mathbf{G}^{(s-1)y_2}_{i-1,l} - c^{y_1}\mathbf{G}^{(s-1)y_2}_{i+1,l} - a^{y_3}\mathbf{G}^{(s-1)y_2}_{i,l-1} - c^{y_3}\mathbf{G}^{(s-1)y_2}_{i,l+1} + d^{y_1y_3}\mathbf{G}^{(s-1)y_2}_{i,l} \\
&+e^{y_1y_2}(\mathbf{G}^{+1}{}^{(s-1)y_2}_{i+1,l} - \mathbf{G}^{+1}{}^{(s-1)y_2}_{i-1,l} - \mathbf{G}^{-1}{}^{(s-1)y_2}_{i+1,l} + \mathbf{G}^{-1}{}^{(s-1)y_2}_{i-1,l}) \\
&+e^{y_1y_3}(\mathbf{G}^{(s-1)y_2}_{i+1,l+1} - \mathbf{G}^{(s-1)y_2}_{i-1,l+1} - \mathbf{G}^{(s-1)y_2}_{i+1,l-1} + \mathbf{G}^{(s-1)y_2}_{i-1,l-1}) \\
&+e^{y_2y_3}(\mathbf{G}^{+1}{}^{(s-1)y_2}_{i,l+1} - \mathbf{G}^{-1}{}^{(s-1)y_2}_{i,l+1} - \mathbf{G}^{+1}{}^{(s-1)y_2}_{i,l-1} + \mathbf{G}^{-1}{}^{(s-1)y_2}_{i,l-1}) - \mathbf{sy2}^s_{i,l}.
\end{aligned}$$

When equation (4.59) has been used to solve the values of the inner grid points, the sides and edges are filled up with the help of equations (4.31) and (4.33).

4.7.4.3 The equation systems when $\frac{s}{3} = integer$ with matrix notation

In this sub-subsection the equation systems will be described, when the method is implicit in the y_3-dimension. In order to make the description more tractable, the following matrices are introduced:

$$\mathbf{G}^{sy_3}_{i,j} = \begin{bmatrix} G^s_{i,j,1} \\ G^s_{i,j,2} \\ G^s_{i,j,3} \\ \vdots \\ G^s_{i,j,M_3-2} \\ G^s_{i,j,M_3-1} \end{bmatrix}, \tag{4.60}$$

$$\mathbf{G^{-1}}_{i,j}^{sy_3} = \begin{bmatrix} G^s_{i,j,0} \\ G^s_{i,j,1} \\ G^s_{i,j,2} \\ \vdots \\ G^s_{i,j,M_3-3} \\ G^s_{i,j,M_3-2} \end{bmatrix}, \tag{4.61}$$

$$\mathbf{G^{+1}}_{i,j}^{sy_3} = \begin{bmatrix} G^s_{i,j,2} \\ G^s_{i,j,3} \\ G^s_{i,j,4} \\ \vdots \\ G^s_{i,j,M_3-1} \\ G^s_{i,j,M_3} \end{bmatrix}, \tag{4.62}$$

$$\mathbf{sy_3}_{i,j}^{s} = \begin{bmatrix} a^{y_3} s1^s_{i,j} \\ 0 \\ 0 \\ \vdots \\ 0 \\ c^{y_3} s6^s_{i,j} \end{bmatrix}. \tag{4.63}$$

Above, $\mathbf{G}_{i,j}^{sy_3}, \mathbf{G^{-1}}_{i,j}^{sy_3}, \mathbf{G^{+1}}_{i,j}^{sy_3}$ and $\mathbf{sy_3}_{i,j}^{s}$ are $M_3 - 1 \times 1$ matrices.

When $\frac{s}{3} = integer$, $s > 1$, $i \in \{1, \ldots, M_1 - 1\}$ and $j \in \{1, \ldots, M_2 - 1\}$ the unknown values $G^s_{i,j,1}, \ldots, G^s_{i,j,M_3-1}$ can be obtained by solving following $M_1 - 1 \times M_2 - 1$ equation systems:

$$\mathbf{A_{y_3}} \mathbf{G}_{i,j}^{sy_3} = \mathbf{R}_{i,j}^{sy_3}, \tag{4.64}$$

where

$$\begin{aligned} &\mathbf{R}_{i,j}^{sy_3} = \\ &-a^{y_1}\mathbf{G}_{i-1,j}^{(s-1)y_3} - c^{y_1}\mathbf{G}_{i+1,j}^{(s-1)y_3} - a^{y_2}\mathbf{G}_{i,j-1}^{(s-1)y_3} - c^{y_2}\mathbf{G}_{i,j+1}^{(s-1)y_3} + d^{y_1y_2}\mathbf{G}_{i,j}^{(s-1)y_3} \\ &+e^{y_1y_2}(\mathbf{G}_{i+1,j+1}^{(s-1)y_3} - \mathbf{G}_{i-1,j+1}^{(s-1)y_3} - \mathbf{G}_{i+1,j-1}^{(s-1)y_3} + \mathbf{G}_{i-1,j-1}^{(s-1)y_3}) \end{aligned}$$

$$+e^{y_1 y_3}(\mathbf{G+1}^{(s-1)y_3}_{i+1,j} - \mathbf{G+1}^{(s-1)y_3}_{i-1,j} - \mathbf{G-1}^{(s-1)y_3}_{i+1,j} + \mathbf{G-1}^{(s-1)y_3}_{i-1,j})$$
$$+e^{y_2 y_3}(\mathbf{G+1}^{(s-1)y_3}_{i,j+1} - \mathbf{G+1}^{(s-1)y_3}_{i,j-1} - \mathbf{G-1}^{(s-1)y_3}_{i,j+1} + \mathbf{G-1}^{(s-1)y_3}_{i,j-1}) - \mathbf{sy3}^{s}_{i,j}.$$

When the (approximate) values at the inner grid points have been solved with the help of equation (4.64), the sides and edges of the cube are filled up with the help of equations (4.31) and (4.33).

4.7.5 The difference schemes with derivative boundary conditions

In the preceding descriptions of the finite difference schemes, it was assumed that conditions directly on the contingent claim values at the boundaries of the approximated region for the natural logarithms of the underlying assets' prices could be derived. For many contingent claims, however, it is easier to say something about the value of the first or second derivative of the contingent claim value with respect to an underlying asset's price as the price becomes very high or very low.

One trick is to assume that the second derivative of the contingent claim value, with respect to an underlying asset's price, is equal to zero for both very high and very low prices of this asset. The trick is equivalent to assuming that the first derivative of the contingent claim value, with respect to an underlying asset's price, is constant (with respect to this price) for both very high and very low prices of the asset. The constant value of the first derivative has, however, not to be specified.

This trick is not directly transferable to the case where we use the natural logarithms of the prices of the underlying assets as state variables. This follows from equation (4.19) where it can be seen that

$$\lim_{x_i \to 0} H_{x_i x_i} = 0 \Longrightarrow \lim_{y_i \to -\infty} (G_{y_i y_i} - G_{y_i}) = 0, \tag{4.65}$$
$$\lim_{x_i \to \infty} H_{x_i x_i} = 0 \not\Longleftrightarrow \lim_{y_i \to \infty} (G_{y_i y_i} - G_{y_i}) = 0. \tag{4.66}$$

In other words, the conclusion from equation (4.65) is that the trick to set the second derivative of the contingent claim value, with respect to the price of an underlying asset, equal to zero at the lower boundary for this price gives a derivative condition that can also be used at the lower boundary (i.e., negative with high absolute value) for the logarithm of this asset's price. Correspondingly, the conclusion from equation (4.66) is that the trick to set the second derivative of the contingent claim value with respect to the price of an underlying asset equal to zero at the upper boundary for this price, does not give a derivative condition that can be used at the upper boundary for the logarithm of the price of this asset[15].

Instead, in the description of the finite difference schemes with derivative boundary conditions it will be assumed that the first derivatives of the contingent claim value, with respect to the absolute prices of the underlying assets, approach constants as the absolute prices become very high[16]. For the boundaries where the natural logarithms of the underlying assets' prices are very small, the trick is used to set the second derivatives with respect to the absolute prices equal to zero. It is, thus, assumed that

$$\begin{aligned}
\lim_{x_1 \to \infty} H_{x_1}(x_1, x_2, x_3, \tau) &= D_1 \qquad x_2, x_3 \in]0, \infty[, \\
\lim_{x_2 \to \infty} H_{x_2}(x_1, x_2, x_3, \tau) &= D_2 \qquad x_1, x_3 \in]0, \infty[, \\
\lim_{x_3 \to \infty} H_{x_3}(x_1, x_2, x_3, \tau) &= D_3 \qquad x_1, x_2 \in]0, \infty[, \qquad (4.67)
\end{aligned}$$

[15]It can, of course, sometimes be plausible to assume that the second derivative of the contingent claim value, with respect to the natural logarithms of the prices of the underlying assets, is equal to zero for very high and very low values of the natural logarithms of the prices. It is, however, much easier to find out what happens intuitively to the contingent claim values as the absolute prices become very high or very low.

Equations (4.65) and (4.66) show that derivative conditions with the absolute prices as underlying state variables cannot be used directly as derivative conditions when the log prices are seen as underlying state variables.

[16]As previously mentioned, the assumption that the first derivatives of the contingent claim value, with respect to the absolute prices of the underlying assets, approach constants as the absolute prices become very high, is very close to assuming that the second derivatives with respect to the absolute prices approach zero as the absolute prices become very high. One difference is, however, that the assumption that the first derivatives of the contingent claim value approach constants ($\neq 0$) gives derivative conditions that can be used when the log prices are used as underlying state variables.

$$
\begin{aligned}
\lim_{x_1 \to 0} H_{x_1x_1}(x_1,x_2,x_3,\tau) &= 0 \qquad && x_2,x_3 \in]0,\infty[, \\
\lim_{x_2 \to 0} H_{x_2x_2}(x_1,x_2,x_3,\tau) &= 0 \qquad && x_1,x_3 \in]0,\infty[, \\
\lim_{x_3 \to 0} H_{x_3x_3}(x_1,x_2,x_3,\tau) &= 0 \qquad && x_1,x_2 \in]0,\infty[.
\end{aligned}
$$

D_1, D_2 and D_3 are constants in equation (4.67). Equation (4.67) gives derivative conditions for the case where the absolute prices of the underlying assets are used as state variables. If equation (4.67) is translated with the help of equations (4.17), (4.18), (4.19) and (4.65) to the case where the log prices of the underlying assets are seen as state variables, the following is derived:

$$
\begin{aligned}
\lim_{y_1 \to \infty} \frac{G_{y_1}(y_1,y_2,y_3,\tau)}{e^{y_1}} &= D_1 \qquad && y_2,y_3 \in]-\infty,\infty[, \\
\lim_{y_2 \to \infty} \frac{G_{y_2}(y_1,y_2,y_3,\tau)}{e^{y_2}} &= D_2 \qquad && y_1,y_3 \in]-\infty,\infty[, \\
\lim_{y_3 \to \infty} \frac{G_{y_3}(y_1,y_2,y_3,\tau)}{e^{y_3}} &= D_3 \qquad && y_1,y_2 \in]-\infty,\infty[, \\
\lim_{y_1 \to -\infty} [G_{y_1y_1}(y_1,y_2,y_3,\tau) - G_{y_1}(y_1,y_2,y_3,\tau)] &= 0 \qquad && y_2,y_3 \in]-\infty,\infty[, \\
\lim_{y_2 \to -\infty} [G_{y_2y_2}(y_1,y_2,y_3,\tau) - G_{y_2}(y_1,y_2,y_3,\tau)] &= 0 \qquad && y_1,y_3 \in]-\infty,\infty[, \\
\lim_{y_3 \to -\infty} [G_{y_3y_3}(y_1,y_2,y_3,\tau) - G_{y_3}(y_1,y_2,y_3,\tau)] &= 0 \qquad && y_1,y_2 \in]-\infty,\infty[.
\end{aligned} \tag{4.68}
$$

To be able to use the finite difference schemes in this paper, the true region (4.25) of the log prices of the underlying assets is approximated with region (4.26). Once this is done the conditions in equation (4.68) are converted to conditions applicable to the approximate region. This gives[17]

$$
\begin{aligned}
G_{y_1}(R_1^U,y_2,y_3,\tau) &= e^{R_1^U} D_1 \quad && y_2 \in [R_2^L,R_2^U], y_3 \in [R_3^L,R_3^U], \\
G_{y_2}(y_1,R_2^U,y_3,\tau) &= e^{R_2^U} D_2 \quad && y_1 \in [R_1^L,R_1^U], y_3 \in [R_3^L,R_3^U], \\
G_{y_3}(y_1,y_2,R_3^U,\tau) &= e^{R_3^U} D_3 \quad && y_1 \in [R_1^L,R_1^U], y_2 \in [R_2^L,R_2^U], \\
G_{y_1y_1}(R_1^L,y_2,y_3,\tau) - G_{y_1}(R_1^L,y_2,y_3,\tau) &= 0 \quad && y_2 \in [R_2^L,R_2^U], y_3 \in [R_3^L,R_3^U], \\
G_{y_2y_2}(y_1,R_2^L,y_3,\tau) - G_{y_2}(y_1,R_2^L,y_3,\tau) &= 0 \quad && y_1 \in [R_1^L,R_1^U], y_3 \in [R_3^L,R_3^U], \\
G_{y_3y_3}(y_1,y_2,R_3^L,\tau) - G_{y_3}(y_1,y_2,R_3^L,\tau) &= 0 \quad && y_1 \in [R_1^L,R_1^U], y_2 \in [R_2^L,R_2^U].
\end{aligned} \tag{4.69}
$$

[17]Equation (4.69) is, of course, an approximation of equation (4.68). Sometimes the approximation is good and sometimes it is bad.

However, if conditions (4.69) should be used in the finite difference schemes, discrete versions of them have to be derived. This can be done with the help of the difference quotients in equation (4.36). By doing this the following conditions are derived[18]:

$$
\begin{array}{rcll}
G^s_{M_1,j,l} &=& G^s_{M_1-1,j,l} + h_1 D_1 e^{R_1^U} & j \in \{1,2,\ldots,M_2-1\}; l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{i,M_2,l} &=& G^s_{i,M_2-1,l} + h_2 D_2 e^{R_2^U} & i \in \{1,2,\ldots,M_1-1\}; l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{i,j,M_3} &=& G^s_{i,j,M_3-1} + h_3 D_3 e^{R_3^U} & i \in \{1,2,\ldots,M_1-1\}; j \in \{1,2,\ldots,M_2-1\}, \\
G^s_{0,j,l} &=& \frac{1}{1+\frac{h_1}{2}}[2G^s_{1,j,l} - (1-\frac{h_1}{2})G^s_{2,j,l}] & j \in \{1,2,\ldots,M_2-1\}; l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{i,0,l} &=& \frac{1}{1+\frac{h_2}{2}}[2G^s_{i,1,l} - (1-\frac{h_2}{2})G^s_{i,2,l}] & i \in \{1,2,\ldots,M_1-1\}; l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{i,j,0} &=& \frac{1}{1+\frac{h_3}{2}}[2G^s_{i,j,1} - (1-\frac{h_3}{2})G^s_{i,j,2}] & i \in \{1,2,\ldots,M_1-1\}; j \in \{1,2,\ldots,M_2-1\}, \\
G^s_{M_1,M_2,l} &=& G^s_{M_1-1,M_2-1,l} + h_1 D_1 e^{R_1^U} + h_2 D_2 e^{R_2^U} & l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{M_1,j,M_3} &=& G^s_{M_1-1,j,M_3-1} + h_1 D_1 e^{R_1^U} + h_3 D_3 e^{R_3^U} & j \in \{1,2,\ldots,M_2-1\}, \\
G^s_{i,M_2,M_3} &=& G^s_{i,M_2-1,M_3-1} + h_2 D_2 e^{R_2^U} + h_3 D_3 e^{R_3^U} & i \in \{1,2,\ldots,M_1-1\}, \\
G^s_{0,M_2,l} &=& \frac{1}{1+\frac{h_1}{2}}[2G^s_{1,M_2-1,l} - (1-\frac{h_1}{2})G^s_{2,M_2-1,l}] + h_2 D_2 e^{R_2^U} & l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{0,j,M_3} &=& \frac{1}{1+\frac{h_1}{2}}[2G^s_{1,j,M_3-1} - (1-\frac{h_1}{2})G^s_{2,j,M_3-1}] + h_3 D_3 e^{R_3^U} & j \in \{1,2,\ldots,M_2-1\}, \\
G^s_{M_1,0,l} &=& \frac{1}{1+\frac{h_2}{2}}[2G^s_{M_1-1,1,l} - (1-\frac{h_2}{2})G^s_{M_1-1,2,l}] + h_1 D_1 e^{R_1^U} & l \in \{1,2,\ldots,M_3-1\}, \\
G^s_{M_1,j,0} &=& \frac{1}{1+\frac{h_3}{2}}[2G^s_{M_1-1,j,1} - (1-\frac{h_3}{2})G^s_{M_1-1,j,2}] + h_1 D_1 e^{R_1^U} & j \in \{1,2,\ldots,M_2-1\}, \\
G^s_{i,0,M_3} &=& \frac{1}{1+\frac{h_2}{2}}[2G^s_{i,1,M_3-1} - (1-\frac{h_2}{2})G^s_{i,2,M_3-1}] + h_3 D_3 e^{R_3^U} & i \in \{1,2,\ldots,M_1-1\}, \\
G^s_{i,M_2,0} &=& \frac{1}{1+\frac{h_3}{2}}[2G^s_{i,M_2-1,1} - (1-\frac{h_3}{2})G^s_{i,M_2-1,2}] + h_2 D_2 e^{R_2^U} & i \in \{1,2,\ldots,M_1-1\}, \\
G^s_{0,0,l} &=& \frac{1}{(1+\frac{h_1}{2})(1+\frac{h_2}{2})}[4G^s_{1,1,l} - 2(1-\frac{h_2}{2})G^s_{1,2,l} - &
\end{array}
\tag{4.70}
$$

[18]It should, however, be noted that the first derivatives for the boundaries where $y_1 = R_1^U, y_2 = R_2^U$ or $y_3 = R_3^U$ are modelled as backward difference approximations in equation (4.70), and not as central difference approximations as previously. To avoid confusion, backward difference approximations are used to approximate first order derivatives, with respect to a state variable at the upper boundary (e.g., $i = M_1$) for this variable. Central difference approximations are still used to approximate first order derivatives with respect to a state variable at the lower boundary (e.g., $j = 0$) for this variable.

$$
\begin{aligned}
& 2(1-\frac{h_1}{2})G^s_{2,1,l}+(1-\frac{h_1}{2})(1-\frac{h_2}{2})G^s_{2,2,l}] && l\in\{1,2,\ldots,M_3-1\},\\
G^s_{0,j,0} &= \frac{1}{(1+\frac{h_1}{2})(1+\frac{h_3}{2})}[4G^s_{1,j,1}-2(1-\frac{h_3}{2})G^s_{1,j,2}- \\
& 2(1-\frac{h_1}{2})G^s_{2,j,1}+(1-\frac{h_1}{2})(1-\frac{h_3}{2})G^s_{2,j,2}] && j\in\{1,2,\ldots,M_2-1\},\\
G^s_{i,0,0} &= \frac{1}{(1+\frac{h_2}{2})(1+\frac{h_3}{2})}[4G^s_{i,1,1}-2(1-\frac{h_3}{2})G^s_{i,1,2}- \\
& 2(1-\frac{h_2}{2})G^s_{i,2,1}+(1-\frac{h_2}{2})(1-\frac{h_3}{2})G^s_{i,2,2}] && i\in\{1,2,\ldots,M_1-1\}.
\end{aligned}
$$

4.7.5.1 The explicit finite difference scheme with derivative boundary conditions

There is only one difference between the explicit finite difference scheme, with the derivative boundary conditions described in subsection 4.7.5, and the explicit finite difference scheme with boundary conditions directly on the contingent claim value. The difference is how the values for the sides and edges of the cube are set. When derivative boundary conditions are employed, equation (4.70) is used. When boundary conditions directly on the contingent claim value are employed, equations (4.31) and (4.33) are used to set the values at the sides and edges of the cube. The solution methodology with derivative boundary conditions is therefore as follows:

The initial condition (4.29) gives the contingent claim values, when $s = 0$. Equation (4.26) establishes that the region for the natural logarithms of the underlying assets' prices has been approximated with a cube, for a given value of s. With the help of the initial condition, equation (4.39) can be used to establish all the approximate contingent claim values at the inner grid points of the cube, when $s = 1$. In order to find a solution to the approximate contingent claim values in the next time step, the values at the side and edges of the cube when $s = 1$ are required. These values are provided by equation (4.70). Once the approximate contingent claim values at time step $s = 1$ have been established, equation (4.39) can be used to find solutions to the approximate contingent claim values at the inner grid points for time step $s = 2$, and so forth and so on. This continues until the approximate contingent claim values at time step $s = N$ have been found. The desired contingent claim value is $G^N_{i_0,j_0,l_0}$ where $ln(x_1(T)) = R^L_1 + i_0h_1, ln(x_2(T)) = R^L_2 + j_0h_2$ and $ln(x_3(T)) = R^L_3 + l_0h_3$, and $x_1(T), x_2(T)$ and $x_3(T)$ are the current prices of the underlying assets.

4.7.5.2 The generalized ADI-method with derivative boundary conditions

The generalized ADI-method with derivative boundary conditions is mainly the same as with boundary conditions directly on the contingent claim value. This means that the generalized ADI-method with derivative boundary conditions iterates through time. It starts with the initial condition, equation (4.29), when $s = 0$, and stops with the approximate values for G when $s = N$. Given the approximate values for G in time step $s-1$, one of equations (4.41), (4.43) or (4.45), depending on which dimension is the implicit one, can be used to establish the approximate values for G in time step s.

The method outlined involves solving $M_2-1 \times M_3-1$ tridiagonal equation systems with M_1-1 unknown variables per time step when the method is implicit in the y_1-dimension, $M_1-1 \times M_3-1$ tridiagonal equation systems with M_2-1 unknown variables when the method is implicit in the y_2-dimension, and $M_1-1 \times M_2-1$ tridiagonal equation systems with M_3-1 unknown variables when the method is implicit in the y_3-dimension.

There are, however, two things that distinguish the scheme with derivative boundary conditions from the scheme with boundary conditions directly on the contingent claim value. Firstly, the equation systems to be solved are slightly modified. Secondly, we set the values at the sides and edges of the cube with the help of equation (4.70) instead of with the help of equations (4.31) and (4.33).

When the method is implicit in the y_1-dimension the tridiagonal matrix $\mathbf{A_{y_1}}$ in equation (4.46) is changed into

$$\mathbf{A_{y_1}} = \begin{bmatrix} (\frac{2a^{y_1}}{1+\frac{h_1}{2}} + b^{y_1}) & (c^{y_1} - a^{y_1}\frac{1-\frac{h_1}{2}}{1+\frac{h_1}{2}}) & 0 & 0 & 0 & \cdots & 0 \\ a^{y_1} & b^{y_1} & c^{y_1} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_1} & b^{y_1} & c^{y_1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_1} & b^{y_1} & c^{y_1} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_1} & (b^{y_1} + c^{y_1}) \end{bmatrix}. \tag{4.71}$$

The tridiagonal matrix when the method is implicit in the y_2-dimension is as follows:

$$\mathbf{A_{y_2}} = \begin{bmatrix} (\frac{2a^{y_2}}{1+\frac{h_2}{2}} + b^{y_2}) & (c^{y_2} - a^{y_2}\frac{1-\frac{h_2}{2}}{1+\frac{h_2}{2}}) & 0 & 0 & 0 & \cdots & 0 \\ a^{y_2} & b^{y_2} & c^{y_2} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_2} & b^{y_2} & c^{y_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_2} & b^{y_2} & c^{y_2} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_2} & (b^{y_2} + c^{y_2}) \end{bmatrix}. \tag{4.72}$$

Finally, when the method is implicit in the y_3-dimension, the tridiagonal matrix is as follows:

$$\mathbf{A_{y_3}} = \begin{bmatrix} (\frac{2a^{y_3}}{1+\frac{h_3}{2}} + b^{y_3}) & (c^{y_3} - a^{y_3}\frac{1-\frac{h_3}{2}}{1+\frac{h_3}{2}}) & 0 & 0 & 0 & \cdots & 0 \\ a^{y_3} & b^{y_3} & c^{y_3} & 0 & 0 & \cdots & 0 \\ 0 & a^{y_3} & b^{y_3} & c^{y_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & a^{y_3} & b^{y_3} & c^{y_3} \\ 0 & \cdots & \cdots & 0 & 0 & a^{y_3} & (b^{y_3} + c^{y_3}) \end{bmatrix}. \tag{4.73}$$

The remainder of the description of the generalized ADI-method with derivative boundary conditions in matrix notation will only deal with the case where $\frac{(s-1)}{3} = integer$ and $s > 1$, i.e., when the method is implicit in the y_1-dimension. All the other cases can easily be derived from extensions of subsection 4.7.4 in the same manner.

The equation systems when $\frac{(s-1)}{3} = integer$ and $s > 1$

When $s > 1$, $\frac{(s-1)}{3} = integer$, $j \in \{1, \ldots, M_2 - 1\}$ and $l \in \{1, \ldots, M_3 - 1\}$ the unknown values $G^s_{1,j,l}, \ldots, G^s_{M_1-1,j,l}$ can be obtained by solving the following $M_2 - 1 \times M_3 - 1$ equation systems:

$$\mathbf{A_{y_1}} \mathbf{G}^{sy_1}_{j,l} = \mathbf{R}^{sy_1}_{j,l}, \tag{4.74}$$

where

$$\begin{aligned}
&\mathbf{R}^{sy_1}_{j,l} = \\
&-a^{y_2}\mathbf{G}^{(s-1)y_1}_{j-1,l} - c^{y_2}\mathbf{G}^{(s-1)y_1}_{j+1,l} - a^{y_3}\mathbf{G}^{(s-1)y_1}_{j,l-1} - c^{y_3}\mathbf{G}^{(s-1)y_1}_{j,l+1} + d^{y_2y_3}\mathbf{G}^{(s-1)y_1}_{j,l} \\
&+e^{y_1y_2}(\mathbf{G^{+1}}^{(s-1)y_1}_{j+1,l} - \mathbf{G^{-1}}^{(s-1)y_1}_{j+1,l} - \mathbf{G^{+1}}^{(s-1)y_1}_{j-1,l} + \mathbf{G^{-1}}^{(s-1)y_1}_{j-1,l}) \\
&+e^{y_1y_3}(\mathbf{G^{+1}}^{(s-1)y_1}_{j,l+1} - \mathbf{G^{-1}}^{(s-1)y_1}_{j,l+1} - \mathbf{G^{+1}}^{(s-1)y_1}_{j,l-1} + \mathbf{G^{-1}}^{(s-1)y_1}_{j,l-1}) \\
&+e^{y_2y_3}(\mathbf{G}^{(s-1)y_1}_{j+1,l+1} - \mathbf{G}^{(s-1)y_1}_{j-1,l+1} - \mathbf{G}^{(s-1)y_1}_{j+1,l-1} + \mathbf{G}^{(s-1)y_1}_{j-1,l-1}) - c^{y_1}\mathbf{D_{y1}},
\end{aligned}$$

$\mathbf{G}^{sy_1}_{j,l}$ is defined in equation (4.49), $\mathbf{G^{-1}}^{sy_1}_{j,l}$ is defined in equation (4.53), $\mathbf{G^{+1}}^{sy_1}_{j,l}$ is defined in equation (4.54), and

$$\mathbf{D_{y1}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ h_1 D_1 e^{R_1^U} \end{bmatrix}.$$

From equation (4.26) it is known that $G(y_1, y_2, y_3, \tau)$ is approximated to be defined on a cube in the dimensions of the natural logarithms of the prices of the underlying assets for a given τ. The solution of equation (4.74) gives the (approximate) values for all inner grid points. In order to find solutions to the approximate values in the next time step the values on the sides and the edges of the cube also have to be known. These values are calculated by using equation (4.70).

4.8 Numerical evaluations

As mentioned earlier, numerical valuations of contingent claims with three underlying state variables are computationally laborious. The computations in this paper are performed on a massively parallel supercomputer, of type Connection Machine Model CM-2000.

The specific CM-2000 machine used for the computations in this paper is, however, a small one. It has "only" 4096 processors working in parallel. (The smallest Connection Machine Model CM-2000 has 4096 parallel processors, and the largest has 65536 parallel processors.) The small size of the machine requires the use of a smaller grid in the dimensions of the natural logarithms of the underlying assets' prices than desired. The grid size in the log price dimensions is $256 \times 256 \times 128$. (The computer ran out of memory when a larger grid size was used.) This grid size is, of course, too small to give the finite difference schemes a very high accuracy.

Preliminary evaluations showed that neither the generalized ADI-method nor the explicit finite difference scheme are stable for all choices of discretization intervals in the time dimension and log price dimensions. For given choices of interval lengths in the log price dimensions ($\Delta y_1, \Delta y_2$ and Δy_3), the length of the interval in the time dimension ($\Delta\tau$) must be sufficiently small (and also smaller for the explicit scheme than for the generalized ADI scheme) for the methods to be stable. The stability properties are, not surprisingly, dependent on the specific set of parameter values used. It is also hardly surprising that the generalized ADI-method has much better stability properties than the explicit finite difference scheme.

With the specification of the grid in this paper, the choice of 100 time steps is sufficient to make the generalized ADI-method stable for all evaluations performed in this paper. This can be compared to the 300 time steps that are required to make the explicit finite difference scheme stable in all evaluations in this paper.

Usually, many more time steps can be used for the same computing time, when using an explicit finite difference scheme, than when using an implicit finite difference scheme.

When using the Connection Machine Model CM-2000, however, one time step with the explicit finite difference scheme is faster, but not much faster, than one time step with the generalized ADI-method ($\approx$ 25 % faster). Thus, the explicit finite difference scheme is far from fast enough to make up for the extra time steps which must be used due to its bad stability properties.

4.8.1 Parameter values in the numerical evaluations

When valuing contingent claims with three underlying state variables there are many parameters. This means that there are many different strategies for how one can change the parameter values between different valuations. The strategy in this paper is as follows:

In all computations the riskfree rate, the time to maturity, the exercise price and the current prices of the underlying assets remain unchanged. Thus, the parameters to vary are the instantaneous standard deviations and correlation coefficients. But, in each valuation will $\sigma_1 = \sigma_2 = \sigma_3$ and $\rho_{12} = \rho_{13} = \rho_{23}$. To be more precise:

- Parameters to remain unchanged:
  - riskfree rate = r = 0.1
  - time to maturity = T = 1.0
  - exercise price = Œ = 10.0
  - current prices = $x_1(T) = x_2(T) = x_3(T) = 10.0$
- Parameters to vary:
  - instantaneous standard deviations = $\sigma_1 = \sigma_2 = \sigma_3$
  - instantaneous correlation coefficients = $\rho_{12} = \rho_{13} = \rho_{23}$

The specification of the grid has already been discussed earlier in section 4.8. But, in order to include the specifications of all option parameters and all grid parameters in the same subsection, the specification of the grid parameters is repeated below.

The size of the grid in the log price dimensions is $256 \times 256 \times 128$. That is, there are 256 grid points in the y_1-dimension and the y_2-dimension respectively, whereas there are only 128 grid points in the y_3-dimension. The reason for having only 128 grid points in the y_3-dimension is that the computer memory was exhausted when a larger grid size was used.

The number of time steps differs between the computations with the explicit finite difference scheme and the generalized ADI-method. This is due to the difference between the schemes' stability properties. Yet, the number of time steps are the same in all valuations for both schemes respectively, and are

$$N_{ADI} = 100,$$

$$N_{EXP} = 300.$$

4.8.2 Numerical evaluations of European call options on the maximum of three risky assets

Subsection 4.5.1 gives a description of the analytical formula and the initial condition for a European call option on the maximum of three risky assets. The initial condition in subsection 4.5.1 is continuous and has the absolute prices of the underlying assets as state variables. In order to be able to use the finite difference schemes in this paper, however, discrete versions of the initial condition have to be modelled when the log prices are viewed as state variables. With the help of equations (4.13), (4.17) and (4.18), the initial condition with the log prices of the underlying assets as state variables can be written as

$$G(y_1, y_2, y_3, 0) = max[max(e^{y_1}, e^{y_2}, e^{y_3}) - \text{Œ}, 0]. \tag{4.75}$$

Discretization of equation (4.75) gives

$$f_{i,j,l} = G^0_{i,j,l} = max[max(e^{R_1^L + ih_1}, e^{R_2^L + jh_2}, e^{R_3^L + lh_3}) - \text{Œ}, 0]. \tag{4.76}$$

The boundaries are modelled as derivative boundary conditions, exactly in the same manner as was described in subsection 4.7.5. The constants D_1, D_2, and D_3 are set equal

$r = 0.1$; $x_1(T) = x_2(T) = x_3(T) = \text{Œ} = 10.0$; $T = 1.0$; Grid size = 256 x 256 x 128				
$\sigma_1 = \sigma_2 = \sigma_3$	$\rho_{12} = \rho_{13} = \rho_{23}$	ANALYTICAL	EXPL($N_{EXP} = 300$)	ADI($N_{ADI} = 100$)
0.20	0.50	2.267	2.278	2.277
0.30	0.50	3.018	3.022	3.020
0.40	0.50	3.801	3.802	3.800
0.20	0.00	2.712	2.713	2.712
0.20	0.25	2.503	2.506	2.505

Table 4.1: European call option on the maximum of three risky assets.

to unity. That is, equation (4.70) is used with $D_1 = D_2 = D_3 = 1.0$ for the purpose of modelling the discrete boundary conditions, in the case of a European call option on the maximum of three risky assets. The motivation for setting $D_1 = D_2 = D_3 = 1.0$ for a European call option on the maximum of three risky assets is as follows:

In subsection 4.7.5, the first derivatives of the option value with respect to the absolute prices were approximated with the constants D_1, D_2 and D_3 at the upper boundaries for the absolute prices. For example, the value of the first derivative of the option value with respect to x_1 (the absolute price of the first underlying asset) is approximated with D_1 at the upper boundary for x_1. The question is what D_1 should reasonably be for a European call option on the maximum of three risky assets. The probability of exercising the option by choosing to buy the first asset is high when x_1 is high (this probability is, of course, close to unity only when x_1 is very high and also much higher than x_2 and x_3). Furthermore, if the probability of exercising the option by choosing to buy the first asset is high, then an increase in x_1 will increase the value of the option by almost the same amount. Thus, a plausible choice is to set D_1 to unity (this is, at least, plausible for most of the upper boundary for x_1). The same type of reasoning suggests, of course, that $D_2 = 1.0$ and $D_3 = 1.0$ are also reasonable choices.

Valuations are performed with different sets of parameter values in the way described in subsection 4.8.1. The results from these valuations are presented in table 4.1.

From table 4.1 it can be seen that both the explicit scheme and the generalized ADI-method are accurate in all valuations[19]. (This is especially true if we think of the small size of the grid in the dimensions of the log prices of the underlying assets.) The accuracy varies, however, for different sets of paramater values. The cases with high instantaneous correlation coefficients and low instantaneous variances seem to be the cases with the lowest accuracy.

4.8.3 Numerical evaluations of European call options on the minimum of three risky assets

The analytical formula and the initial condition for a European call option on the minimum of three risky assets, with the absolute prices of the underlying assets as state variables, are given in subsection 4.5.2. When using the finite difference schemes in this paper, a discrete initial condition derived with the log prices of the underlying assets as state variables is required.

[19]Equation (4.14) and equation (4.16) in section 4.5 show that evaluations of the analytical formula for the value of a European call option on the maximum of three risky assets, as well as the analytical formula for the value of a European call option on the minimum of three risky assets, requires the evaluation of the 3-variate standardized cumulative normal distribution function no less than 4 times.

As mentioned in subsection 4.5.1, the method for evaluation of the 3-variate standardized cumulative normal distribution function is based on numerical integration. This method guarantees a result with an error smaller than 1×10^{-4}. (In most cases the error is, of course, much smaller.)

Thus, the analytical formula for the value a European call option on the maximum of three risky assets [equation (4.14)], as well as the analytical formula for the value a European call option on the minimum of three risky assets [equation (4.16)], indicates that if every evaluation of the 3-variate standardized cumulative normal distribution function gives the maximal error, and all 4 errors work in the same direction when the "analytical" option price is calculated, a *very* high upper bound for the error in the analytical option price is given by

$$bound_{up} = \sum_{i=1}^{3} x_i(T)(1 \times 10^{-4}) + \text{Œ}e^{-rT}(1 \times 10^{-4}) \approx 3.9 \times 10^{-3}. \tag{4.77}$$

The figure given in equation (4.77) is, indeed, the *worst* possible case. Most often when the numerical method is used to evaluate the analytical formulas the third decimal in the option price will be correct, and in nearly all cases the error in the option price is no larger then $\pm 1 \times 10^{-3}$.

Using equations (4.15), (4.17) and (4.18) to find the continuous initial condition with the log prices of the underlying assets as state variables, it can be established that

$$G(y_1, y_2, y_3, 0) = max[min(e^{y_1}, e^{y_2}, e^{y_3}) - Œ, 0]. \tag{4.78}$$

Discretization of equation (4.78) gives

$$f_{i,j,l} = G^0_{i,j,l} = max[min(e^{R_1^L+ih_1}, e^{R_2^L+jh_2}, e^{R_3^L+lh_3}) - Œ, 0]. \tag{4.79}$$

The boundary conditions are modelled in the same way as described in subsection 4.7.5. In this case, the constants D_1, D_2 and D_3 are set equal to zero. Thus, equation (4.70) is used with $D_1 = D_2 = D_3 = 0.0$ when the discrete boundary conditions for a European call option on the minimum of three risky assets are modelled[20]. The motivation for setting $D_1 = D_2 = D_3 = 0.0$ for a European call option on the minimum of three risky assets is as follows:

In subsection 4.7.5, the first derivatives of the option value with respect to the absolute prices were approximated with the constants D_1, D_2 and D_3 at the upper boundaries for the absolute prices. For example, the value of the first derivative of the option value with respect to x_1 (the absolute price of the first underlying asset) is approximated with D_1 at the upper boundary for x_1. The question now is, what D_1 should reasonably be for a European call option on the minimum of three risky assets. The probability of exercising the option by choosing to buy the first asset is low when x_1 is high (this probability is, of

[20]A problem arises when $D_1 = D_2 = D_3 = 0.0$. This problem is of the same type as the problem to model the second derivatives of the contingent claim value with respect to the absolute prices of the underlying assets equal to zero for very high absolute prices, as discussed in subsection 4.7.5. This follows from equation (4.19) which establishes that

$$\lim_{x_i \to \infty} H_{x_i} = 0 \not\Rightarrow \lim_{y_i \to \infty} G_{y_i} = 0 \qquad i = 1, 2, 3. \tag{4.80}$$

In other words, a first derivative of the contingent claim value with respect to an absolute price that is close to zero for very high values of this absolute price, does not (it is, of course, not ruled out either) imply that the first derivative of the contingent claim value with respect to the logarithm of this price is also close to zero for high values of the log price. From equation (4.69) it can be seen, however, that this condition is actually enforced at the upper boundaries for the log prices by setting $D_1 = D_2 = D_3 = 0.0$.

$r = 0.1$; $x_1(T) = x_2(T) = x_3(T) = Œ = 10.0$; $T = 1.0$; Grid size = 256 x 256 x 128				
$\sigma_1 = \sigma_2 = \sigma_3$	$\rho_{12} = \rho_{13} = \rho_{23}$	ANALYTICAL	EXPL($N_{EXP} = 300$)	ADI($N_{ADI} = 100$)
0.20	0.50	0.525	0.518	0.519
0.30	0.50	0.571	0.562	0.563
0.40	0.50	0.623	0.617	0.618
0.20	0.00	0.235	0.232	0.232
0.20	0.25	0.364	0.360	0.360

Table 4.2: European call option on the minimum of three risky assets.

course, only close to zero when x_1 also is much higher than x_2 and x_3). Furthermore, if the probability of exercising the option by buying the first asset is low, then an increase in x_1 has almost no influence on the value of the option. Thus, a plausible choice is to set D_1 close to zero (this is, in any event, plausible for most of the upper boundary for x_1). The same type of reasoning leads to the conclusion that $D_2 = 0.0$ and $D_3 = 0.0$ are reasonable choices as well.

Valuations are made with exactly the same choices of parameter values as were used in the valuations of a European call option on the maximum of three risky assets. The results from these valuations are compiled in table 4.2.

Table 4.2 shows that both the explicit finite difference scheme and the generalized ADI-method are accurate in all valuations. (The accuracy seems to be slightly better for a European call option on the maximum of three risky assets than for a European call option on the minimum of three risky assets. This is possibly due to the problem of modelling the boundary conditions for a European call option on the minimum of three risky assets, as discussed in a footnote in this subsection.)

Moreover, precisely as for a European call option on the maximimum of three risky assets, the cases with high instantaneous correlation coefficients and low instantaneous variances seem to be the cases with the lowest accuracy.

4.9 Conclusions

Valuation of contingent claims with three underlying state variables by means of the "simple" finite difference schemes in this paper is a story of mixed success. The story is, however, mainly a success story.

Let us first consider the negative aspects. Both the explicit finite difference scheme and the generalized ADI-method have very bad stability properties, when applied to the partial differential equation derived with the prices of the underlying assets as state variables. However, when the schemes are used on the partial differential equation derived with the natural logarithms of the prices of the underlying assets as state variables, the stability properties become relatively good, for both the explicit finite difference scheme and the generalized ADI-method.

The stability problems for the simple schemes of this paper clearly motivate the search for finite difference schemes with better stability properties. More advanced finite difference schemes, however, rapidly become rather complex, and thereby hard to understand for non-experts of numerical methods. It might still be worthwhile to try to find a finite difference scheme with better stability properties when used for the valuation of contingent claims with three underlying state variables in a investigation in the future.

The positive aspects then remain. Both the explicit finite difference scheme and the generalized ADI-method proved accurate in every valuation performed. This is especially true if we think of the small size of the grid in the dimensions of the log prices that we were forced to use. The accuracy varies, however, with different sets of parameter values. The cases with high instantaneous correlation coefficients and low instantaneous variances seem to be the cases with the lowest accuracy.

A brief comparison between the explicit finite difference scheme and the generalized ADI-method indicates that the generalized ADI-method has much better stability properties than the explicit finite difference scheme. This means that the explicit finite difference scheme requires many more time steps to achieve stability than the generalized ADI-method. This drawback for the explicit finite difference scheme is partly eliminated since it iterates faster through time than the generalized ADI-method. When using the Connection Machine Model CM-2000, however, the explicit scheme was far from fast enough to make up for the extra time steps required, due to its bad stability properties.

The better stability properties of the generalized ADI-method clearly speak for using that scheme. The explicit finite difference scheme is, however, easier to implement. Both schemes hence have advantages and drawbacks.

To summarize: Numerical valuation of multivariate contingent claims by means of finite difference methods on a massively parallel computer, like the CM-2000 machine, seems to be a useful approach. Moreover, that approach should be useful for academics as well as for practitioners.

Chapter 5

Paper D: A Lattice Approach for Pricing of Multivariate Contingent Claims

5.1 Introduction

The purpose of this paper is to develop a lattice (or tree) approach, which can be used for pricing of multivariate contingent claims. Before doing this, however, an introduction will be given.

Since the pioneering papers by *Black and Scholes* (see [12]), and *Merton* (see [99], [100] and [101]), option pricing has been one of the most research-intensive areas within the field of economics. From all of the thousands of papers that have been published, it can be concluded that the option pricing framework can be applied to a very wide range of problems in financial economics.

Deregulations and the very rapid progress within information technology have made the financial markets increasingly sophisticated in the last decade. On these markets, corporations (and their financial advisors) have invented more and more exotic methods for the financing of their activities. These methods of innovative financing include the issuing of complex contingent claims, the payoffs of which can depend on several underlying state variables and involve embedded packages of options.

Analyses of some of these complex contingent claims have been published in academic journals. Only in rare cases, however, have researchers been able to derive closed-form solutions. Indeed, for contingent claims with more than one underlying state variable, closed-form solutions have been very scarce. (Closed-form solutions for some of these last cases are given in [29], [30], [75], [91] and [129].)

Since an overwhelming majority of complex contingent claims lack closed-form solutions, numerical methods often have to be relied on in order to price them. For the case of only one underlying state variable, a wide variety of methods have been used. These methods include analytical approximation, compound option methods, series solutions, Monte Carlo simulation, numerical integration, lattice approaches and finite difference methods. There are also a vast number of pricing applications with only one underlying state variable in the literature.

When there are two underlying state variables, the number of published pricing applications is substantially decreased. Moreover, the variety of numerical methods used is reduced.

For the case of contingent claims with three underlying state variables, there is to the author's knowledge only one example of a valuation by numerical methods. The example is a valuation by means of a lattice approach (see [16]).

When using numerical methods for option pricing, a trade off between accuracy and speed of computation must be made. The importance of finding effective numerical algorithms becomes more pronounced as the number of underlying state variables increases. This is due to the fact that the computational burden increases very rapidly as more state variables are added, for most numerical methods.

Lattice approaches are among the most popular numerical methods for option pricing. The reasons for this popularity are that most lattice approaches are intuitively simple, flexible and can "easily" handle American type options.

The method developed by *Cox, Ross and Rubinstein* (CRR) in 1979 (see [43]) is the most well-known, and presumably also the most commonly used, among the lattice approaches which have been suggested in the option pricing context. The CRR approach has later been extended and modified by *Bartter and Rendleman* (see [7]), *Boyle* (see [14] and [15]), *Boyle, Evnine and Gibbs* (BEG) (see [16]) and *Hull and White* (HW) (see [68] and [69][1]).

In the search for a lattice framework that can easily be extended to handle more than one underlying state variable, hitherto the BEG approach (see [16]) and the HW90 approach (see [69]) are perhaps the most promising.

The BEG approach is a multivariate extension of the CRR method. BEG approximate the system of underlying processes (all of which are assumed to follow geometric Brownian motions) with a multivariate binomial lattice. BEG first set the jump sizes, and then determine the jump probabilities in a fashion that assures convergence between the discrete approximation and the system of continuous processes as the step size in the time dimension approaches zero.

Advantages of the BEG approach are that it is intuitive, easy to understand and that the number of branches emanating from each node in the lattice increases fairly slowly when the number of state variables increases (since a multivariate binomial lattice is used). Drawbacks of the BEG approach are that it has a rather slow convergence[2] (not efficient), and it is not general either with respect to the processes of the underlying state variables

---

[1]In [69], Hull and White discuss option pricing with the help of the explicit finite difference method. But when they start to modify the explicit finite diffference method, it is more straightforward to interpret the method as a trinomial lattice approach. This is in correspondence to results found by *Brennan and Schwartz* , which show that in the case with one underlying state variable the explicit finite difference method in many respects is equivalent to a trinomial lattice approach (see [22]).

When Hull and White extend their method to two underlying state variables, the method becomes a bivariate trinomial lattice approach. That is, the method in [69] is *not* an explicit finite difference method when there is more than one underlying state variable. (At least not an explicit finite difference method where the usual finite difference quotients are used.)

[2]A lattice method is said to be convergent when the approximate solution established with the help of the method approaches the exact solution as the length of the step size in the time dimension approaches zero. The rate with which the approximate solution established with a lattice method approaches the exact solution as the number of time steps is increased is called the method's rate of convergence.

or the number of underlying state variables. This last drawback follows from the fact that the approach does not guarantee that all jump probabilities will be non-negative when the number of underlying state variables is larger than two. Moreover, the problem with possible negative jump probabilities increases when the number of underlying state variables increases.

The HW90 approach starts with transformations, in two steps, of the processes of the underlying state variables (if needed). The first step transforms all processes into processes with constant instantaneous standard deviations and correlation coefficients. The second step transforms these processes (transformed in the first step) into a system of uncorrelated processes. Then the system of uncorrelated processes is approximated with a multivariate trinomial lattice. The jump probabilities and the jump sizes for each transformed underlying state variable are determined with the help of an explicit finite difference formulation. Negative jump probabilities for a (transformed) underlying state variable are handled with an astute modification of the branching process for that particular variable.

Advantages of the HW90 approach are that it is general with respect to both the type of underlying processes and applications. Drawbacks of the HW90 approach are that it is rather complex (at least compared to the BEG approach) and that the number of branches emanating from each node in the lattice increases very fast as the number of state variables increases (since a multivariate trinomial lattice is used). The last fact makes the HW90 approach computationally extremely burdensome when the number of underlying state variables is greater than two [3]. From a practical point of view, the HW90 approach can perhaps be used for the pricing of contingent claims with three underlying state variables (with the help of a powerful computer). However, to use the approach to price contingent claims with more than three underlying state variables is, in the author's view, impracticable, even with the help of the most powerful of supercomputers.

The lattice approach developed in this paper can be viewed as an improvement of

[3] In [69], HW only exemplifies the approach for contingent claims with one and two underlying state variables.

the BEG approach. The idea of performing two steps of introductory transformations of the processes of the underlying state variables is, however, borrowed from the HW90 approach.

The first step transforms (the risk-neutralized) processes of the underlying state variables into Itô processes with constant instantaneous drifts, standard deviations and correlation coefficients. The second step transforms the processes already transformed in the first step into a system of uncorrelated processes. Then the system of uncorrelated processes is approximated with a multivariate binomial lattice, as in the BEG approach. In opposition to the BEG approach, however, all jump probabilities are first set (all probabilities equal), and then the jump sizes are determined in a fashion that assures convergence between the discrete approximation and the system of uncorrelated processes as the step size in the time dimension approaches zero[4].

There are several advantages in setting all jump probabilities equal, as compared to setting the jump sizes and then deriving the probabilities. For example, a more efficient method (faster convergence) and a method that is easier to implement is achieved, and the problem with possible negative jump probabilities is avoided. Moreover, the lattice approach developed in this paper:

- can easily be extended to cases with an arbitrary number of underlying state variables (of course, the computational burden will be large, but far from as large as with the HW90 approach, for cases with many state variables).
- can handle some variety of stochastic processes (through the transformation procedure).
- is simple and easy to understand.
- gives very simple formulas for the jump sizes and the jump probabilities.

[4]The approach to first set all jump probabilities equal and then choose the jump sizes to assure convergence cannot be used in the BEG framework. The correlations between the processes of the underlying assets lead to inconsistent equations.

- makes the number of branches emanating from each node in the lattice increase fairly slowly as the number of state variables increases.

The rest of the paper is organized as follows:

Section 5.2 describes the construction of the lattice. That section also includes an outline of the approach. Numerical evaluations are performed in section 5.3. Since the lattice approach derived in this paper can be viewed as an improvement of the BEG approach, the numerical results from the approach developed in this paper are compared to numerical results from the BEG approach. Conclusions are given in section 5.4.

5.2 Construction of the lattice

The purpose of this paper is to find a numerical method which can be used for pricing of multivariate contingent claims. In the standard option pricing setting, this pricing problem is continuous in time and space dimensions (a discretization of the problem is performed when the lattice approach is derived).

The following assumptions underlie the continuous problem to be solved numerically (see [100] and [101])[5]:

A.1 There are no transaction costs, taxes or problems with indivisibilities of assets.

A.2 There are a sufficient number of investors with comparable wealth levels so that each investor believes that he/she can buy and sell as much of an asset as he/she wants at the market price.

A.3 An exchange market exists for borrowing and lending at the same rate of interest.

A.4 Short-sales of all assets, with full use of the proceeds, is allowed.

A.5 Trading in assets takes place continuously in time.

[5]It is not necessary for all of the assumptions mentioned below to be fulfilled for the option pricing framework to hold. Indeed, the framework holds even if many of these assumptions are substantially weakened. See [100] for a discussion of this.

A.6 There is a riskless asset whose rate of return per unit time is known and constant. Denote this rate of return by r.

A.7 The underlying state variables are prices of traded[6] securities. The dynamics for the prices of the underlying securities are described by the following stochastic differential equations[7]:

$$dS_i = \mu_i S_i dt + \sigma_i S_i dz_i \qquad i = 1, \ldots, n, \tag{5.1}$$

where

- t = calendar time.
- μ_i = S_i's instantaneous expected rate of return per unit time.
- σ_i^2 = the instantaneous variance per unit time of S_i's instantaneous rate of return. σ_i^2 $\forall i$ are assumed to be constants.
- $z_i(t)$ = a Wiener process. That is, $E(dz_i) = 0$, $Var(dz_i) = dt$ and $dz_i \sim N(0, dt)$ where $N(\bullet, \bullet)$ indicates the normal distribution. It is also assumed that $cov(dz_i(t), dz_j(t)) = \rho_{ij} dt$ where $\rho_{ij} = constant$ $i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$ and that $cov(dz_i(t), dz_j(s)) = 0$ $i, j = 1, \ldots, n; s \neq t$.

A.8 It is assumed that investors prefer more to less. It is also assumed that investors agree upon σ_i $i = 1, \ldots, n$, ρ_{ij} $i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$ and $cov(dz_i(t), dz_j(s)) = 0$ $i, j = 1, \ldots, n; s \neq t$, but it is *not* assumed that they necessarily agree upon μ_i $i = 1, \ldots, n$.

---

[6]The underlying state variables do not have to be prices of traded securities for the continuous model, or for the lattice method in this paper, to hold. If one of the underlying state variables is not the price of a traded asset it can be included in the option pricing framework with the help of a measure called the the market price of risk for that variable (see [42] and [61]).

[7]The lattice approach in this paper does not require the processes of the underlying state variables to be of the type in equation (5.1). The approach is directly applicable in all cases where the risk-neutralized processes can be transformed into Itô processes with constant instantaneous drifts, standard deviations and correlation coefficients.

A.9 The price of the contingent claim is a function of the prices of the underlying securities and time. Denote the price of the contingent claim by $W(S_1,\ldots,S_n,t)$. $W(S_1,\ldots,S_n,t)$ is a real valued continuous non-random function such that $\frac{\partial W}{\partial t}$, $\frac{\partial W}{S_i}$ $\forall i$ and $\frac{\partial^2 W}{\partial S_i \partial S_j}$ $\forall i,j$ are continuous.

One common approach for pricing contingent claims is to derive a partial differential equation that the price of the contingent claim must satisfy, with the help of Itô's lemma and some arbitrage arguments. The price of the contingent claim is given by the solution to this partial differential equation. If no analytical solution can be found, as is often the case, then the partial differential equation can be numerically solved with finite difference methods. When finite difference methods are used, the partial differential equation is approximated with a difference equation. That is, finite difference methods use a discrete approximation of the partial differential equation.

Another approach, for the purpose of numerically solving the contingent claim price, is to approximate the continuous processes of the underlying state variables with discrete versions[8]. It is this latter approach that is used when the contingent claim price is achieved with the help of a lattice method.

Lattice approaches are based on risk-neutral pricing (see [42] and [61]). The arguments behind risk-neutral pricing can, very briefly and superficially, be described in the following manner:

With assumptions **A1-A9** the contingent claim price can, in principle, be derived by pure arbitrage arguments. Since the contingent claim price can be determined by the absence of arbitrage alone, this price does not directly depend on the investors' risk preferences (it is only required that investors prefer more to less). Investors' risk preferences are, of course, important when the processes of the underlying state variables and the riskless asset are determined in some equilibrium. These processes are, however, perceived as given entities when pricing the contingent claim. Thus, for the purpose of pricing the

[8]The approach of using a discrete approximation of the partial differential equation and the approach of using discrete approximations of the underlying processes are equivalent in some senses (e.g., see [22]).

contingent claim, it does not matter what the investors' risk preferences are. As long as they determine the same *relevant* parameter values, they will value the contingent claim identically.

In the cases where all the underlying state variables are traded assets with price processes satisfying equation (5.1), it is possible to show that the contingent claim price does not depend on μ_i $i = 1, \ldots, n$ (e.g., see [42]). The only relevant parameters for the pricing problem are r, σ_i $i = 1, \ldots, n$ and ρ_{ij} $i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$[9]. That is, when pricing the contingent claim it is "only" necessary to find the equilibrium price processes in some world where preferences are given and consistent with the specified values of r, σ_i $i = 1, \ldots, n$ and ρ_{ij} $i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$.

A convenient choice of preferences for many problems is risk-neutrality. In a risk-neutral world, equilibrium requires that the expected rate of return on both the underlying assets and the contingent claim must equal the riskless rate of return. To obtain the risk-neutralized processes of the prices of the underlying assets, their instantaneous expected rate of returns are replaced by the riskless rate of return. That is, the risk-neutralized processes become

$$dS_i = rS_i dt + \sigma_i S_i dz_i \qquad i = 1, \ldots, n. \tag{5.2}$$

The construction of the lattice approach in this paper starts with two steps of transformations of the underlying assets' price processes. In the first step, the processes in equation (5.2) are transformed into Itô processes with constant instantaneous drifts, standard deviations and correlation coefficients. The second step transforms the processes transformed in the first step into a system of uncorrelated processes. These uncorrelated processes are also Itô processes with constant instantaneous drifts and standard deviations.

At this moment, a relevant question is: "Are the state variables after the two steps of transformations sufficient to determine uniquely the value of the contingent claim?". The

[9]If some of the underlying state variables are not prices of traded assets, then the risk-neutral valuation arguments must be extended (see [40]).

answer to this question is: "Yes, they are sufficient", and the reason is the following:

Assumption **A.9** establishes that the value of the contingent claim is a real valued function (although unknown) of the prices of the underlying securities and time. Thus, given the prices of the underlying securities at a given point in time, there is a unique real number that determines the value of the contingent claim. Moreover, both steps of introductory transformations are based on one-to-one functions. Since the contingent claim value is a function of the prices of the underlying securities and time, and there are one-to-one mappings between the transformed state variables and the prices of the underlying securities, there also exists a rule that determines one and only one contingent claim value for a given set of values of the transformed state variables at a given point in time. This means that the value of the contingent claim by definition (e.g., see [59] p.5) is a function (this function is, of course, also unknown) of the transformed state variables and time. Hence, if the values of the transformed state variables are known at any point in time, it is possible to deduce (in principle) the price of the contingent claim at that time point. The problem is, of course, that the functional form of the pricing function is not known.

Once the transformations have been performed, the system of uncorrelated processes is approximated with a multivariate binomial lattice. Before this approximation is dicussed, however, a multivariate binomial lattice will be described shortly.

A multivariate binomial lattice, with n underlying state variables, can be interpreted as a $(n+1)$ dimensional tree, one time dimension and one dimension for each state variable. From each node in the tree, 2^n branches emanate. This is due to the fact that each transformed state variable can jump either up or down. Since there are n state variables, this gives 2^n possible combinations. Thus, from a given node in the tree there are 2^n states of the world that can occur during the next time period. If instead the perspective is from standing at the root of the tree (i.e., the single node at current time) and looking s time periods ahead, there would then be $(s+1)^n$ possible states of the world that can occur in time step s (each possible state of the world is represented by a node in the tree). Over s time periods each transformed state variable can have between 0 and s

up jumps[10]. Since there are n state variables this gives $(s+1)^n$ possible combinations.

The reasoning behind the approximation of the system of uncorrelated processes with a multivariate binomial lattice is as follows:

First, divide the time to maturity for the contingent claim into N equally long time periods. For a time period, the continuous joint end of period distribution of the transformed state variables (conditional on the values of the transformed state variables at the beginning of the time period) is then approximated with a 2^n-jump distribution.

To derive the 2^n-jump distribution that should approximate the continuous distribution, all jump probabilities are first set to be equal (i.e., all probabilities $= \frac{1}{2^n} = \frac{1}{m}$). The jump sizes are then determined in a fashion that ensures convergence between the 2^n-jump distribution and the continuous conditional joint end of period distribution for the transformed state variables as N approaches infinity (i.e., as the length of the time period approaches 0).

When a consistent lattice for the approximation of the processes of the transformed state variables is constructed, the contingent claim can be priced by using risk-neutral pricing and working recursively backward through time. That is, the terminal condition gives the price of the contingent claim at each of the $(N+1)^n$ nodes at the expiration date. The price of the contingent claim one time period earlier can be calculated by discounting the expected[11] price of the contingent claim at the riskless rate, at each of the N^n nodes. With these prices, the contingent claim prices at each of the $(N-1)^n$ nodes two time periods before the expiration date can be calculated, and so forth and so on. Thus, the contingent claim price at current time, which is the desired price, can be obtained by

[10]When the number of up jumps of a state variable are known, the number of down jumps for this variable is given by (*s - number of up jumps*). Hence, it is only necessary to keep track of the number of up jumps for each state variable in order to be able to completely specify the position in the tree.

[11]This is a conditional expectation. It is conditional on the information available at the specific node in question.

working recursively backward through time with the following formula[12]:

$$W^s_{i_1,i_2,\ldots,i_n} = e^{-rk}\frac{1}{m}(W^{s+1}_{i_1+1,i_2+1,\ldots,i_n+1} + W^{s+1}_{i_1+1,i_2+1,\ldots,i_n} + \ldots + W^{s+1}_{i_1,i_2,\ldots,i_n}) \tag{5.3}$$

$$0 \le s < N; 0 \le i_1 \le s; 0 \le i_2 \le s; \ldots; 0 \le i_n \le s,$$

where[13]

[12]Equation (5.3), perhaps, looks somewhat confusing. It can be clarified by writing down the formula for the cases with two and three underlying state variables. In the case of two underlying state variables, the formula becomes

$$W^s_{i,j} = e^{-rk}\frac{1}{4}(W^{s+1}_{i+1,j+1} + W^{s+1}_{i+1,j} + W^{s+1}_{i,j+1} + W^{s+1}_{i,j})$$

$$0 \le s < N; 0 \le i \le s; 0 \le j \le s,$$

and in the case of three underlying state variables the formula becomes

$$W^s_{i,j,l} = e^{-rk}\frac{1}{8}(W^{s+1}_{i+1,j+1,l+1} + W^{s+1}_{i+1,j+1,l} +$$

$$W^{s+1}_{i+1,j,l+1} + W^{s+1}_{i,j+1,l+1} + W^{s+1}_{i+1,j,l} + W^{s+1}_{i,j+1,l} + W^{s+1}_{i,j,l+1} + W^{s+1}_{i,j,l})$$

$$0 \le s < N; 0 \le i \le s; 0 \le j \le s; 0 \le l \le s.$$

[13]Once the formula for use in the backward recursion has been derived [i.e., equation (5.3)], it can be interesting to make a brief comparison between the multivariate binomial lattice, the explicit finite difference method and the multivariate trinomial lattice.

From equation (5.3) it is clear that the multivariate binomial lattice uses a linear combination of 2^n (n is the number of underlying state variables) different contingent claim prices from the previous time step in order to calculate one contingent claim price in the subsequent time step. The explicit finite difference method (with correlations between the state variables and the usual finite difference quotients) uses a linear combination of $(4\frac{n(n-1)}{2}+2n+1 = 2n^2+1)$ different contingent claim prices in the previous time step to calculate one contingent claim price in the subsequent time step. Finally, the multivariate trinomial lattice uses a linear combination of 3^n old contingent claim prices to calculate one new contingent claim price.

Thus, for $1 \le n \le 2$ the explicit finite difference method and the multivariate trinomial lattice include equal numbers of old prices in the linear combination. Both methods include more old prices than the multivariate binomial lattice.

When $n \ge 3$ the multivariate trinomial lattice always includes more old prices in the linear combination than both of the other methods. Moreover, the number of old prices to include in the linear combination

- $W^s_{i_1,i_2,\ldots,i_n}$ = the (approximate) contingent claim price s time steps from current time when underlying state variable 1 has jumped up i_1 times, underlying state variable 2 has jumped up i_2 times, ... and underlying state variable n has jumped up i_n times.

- T = the contingent claim's time to maturity from $t = t_0$, where t_0 is current time. [That is, time to maturity from an arbitrary t ($t_0 \leq t \leq t_0+T$) is equal to $T-(t-t_0)$.]

- N = the number of time steps that T is divided into.

- $k = T/N$ = the length of a time step.

- $m = 2^n$ = the number of branches that emanate from each node in the lattice.

- $1/m$ = the equal probability for each branch emanating from a node.

The lattice framework described above is directly applicable to European type contingent claims. American type contingent claims, however, can easily be handled by checking the early exercise conditions at the nodes in the lattice. Dividend payments can also be handled in the manner that is usually done when a lattice approach is used (e.g., see [68]).

As previously mentioned, the BEG approach (see [16]) and the HW90 approach (see [69]) are hitherto perhaps the two most promising lattice approaches for the pricing of contingent claims with more than one underlying state variable. Although general, the HW90 approach is computationally very burdensome. This means that the HW90 approach is not very useful for contingent claims with more than two underlying state variables.

The BEG approach also has shortcomings. Its major problem is the possibility of negative jump probabilities when the number of underlying state variables exceeds two. Moreover, as can be seen from the resulting formula for the jump probabilities in BEG's

---

increases *much* faster for the multivariate trinomial lattice than for any of the other two methods.

Furthermore, when $1 \leq n \leq 6$, the multivariate binomial lattice has the smallest number of old prices in the linear combination. When $n \geq 7$, it is, however, the explicit finite difference method that has the smallest number of old prices in the linear combination.

paper {see equation (15) in [16][14]}, the problem with possible negative jump probabilities increases as the number of underlying state variables increases. Another problem for the BEG approach is that it has slow convergence.

The lattice approach developed in this paper should be viewed as an improvement of the BEG approach. Both of the problems previously discussed concerning the BEG approach are handled in the approach in this paper.

The problem of possible negative jump probabilities is avoided by first setting all probabilities equal and then deriving the jump sizes. This policy cannot be used in the BEG framework, since it leads to inconsistent equations. However, by performing introductory transformations of the processes of the underlying state variables the policy can be used[15]. Furthermore, the formulas for the resulting jump sizes become extremely simple, at least compared to the formulas for the resulting jump probabilities in the BEG approach {see equation (15) in [16]}.

The problem of slow convergence in the BEG approach is also handled. This is done, in part, by setting all jump probabilities equal. Having all jump probabilities equal usually leads to a efficient lattice. This fact has been mentioned by more than one author. For

[14]There is a misprint in the resulting formula for the jump probabilities in BEG's paper. Equation (15) in [16] should be

$$p_j = \frac{1}{M}[1 + \sum_{k=1}^{n-1} \sum_{m=k+1}^{n} \delta_{km}(j)\rho_{km} + \sqrt{h}\sum_{k=1}^{n} \delta_k(j)\frac{\mu_k}{\sigma_k}] \qquad j = 1, \ldots, M,$$

and not

$$p_j = \frac{1}{M}[\sum_{k=1}^{n-1} \sum_{m=k+1}^{n} \delta_{km}(j)\rho_{km} + \sqrt{h}\sum_{k=1}^{n} \delta_k(j)\frac{\mu_k}{\sigma_k}] \qquad j = 1, \ldots, M$$

as in BEG's paper.

[15]When using the introductory transformations of the processes of the state variables proposed in this paper, BEG's policy can be used without the problem of possible negative jump probabilities. That is, after the transformations, the jump sizes can be set and then the jump probabilities derived, exactly as in [16], without having the possibility of getting negative jump probabilities (More precisely, negative jump probabilities can always be avoided by chosing a sufficiently small length of the time step.). There are, however, other advantages of having all jump probabilities equal. These advantages are discussed later in this section. Thus, the policy of having all jump probabilities equal is preferred, in any case.

example, when empirically testing a trinomial lattice with one underlying state variable *Boyle* concludes "Best results were obtained when the probabilities were roughly equal." (see [15] p.5). As another example, when discussing the construction of an efficient lattice *Hull and White* conclude "Generally, a satisfactory lattice is one in which the transition probabilities are not regularly close to either 0.0 or 1.0." (which certainly is true when all jump probabilities are equal) (see [68] p.241).

Another advantage that results from having all jump probabilities equal is that the method becomes easy to implement. With all jump probabilities equal there is only one probability to keep track of, instead of 2^n probabilities which is the case when all jump probabilities are different.

In order to be able to use the lattice approach outlined above, two questions have to be answered. These are "How to perform the transformations?" and "How to choose the appropriate jump sizes?". The purpose of the next two subsections is to provide the answers to these questions.

5.2.1 The transformations

In this subsection, the two introductory steps of transformations will be described. The first step of transformations should transform the risk-neutralized processes of the underlying state variables into Itô processes with constant instantaneous drifts, standard deviations and correlation coefficients.

In this paper it is assumed that the risk-neutralized processes of the underlying state variables are described by equation (5.2), i.e., they are geometric Brownian motions. In the case of geometric Brownian motions, it is a well-known fact that Itô processes with constant drifts and standard deviations can be obtained by simple logarithmic transformations. Therefore, make the transformations

$$y_i \equiv ln(S_i) \qquad\qquad i = 1, 2, \ldots, n. \tag{5.4}$$

With the help of Itô's lemma (e.g., see [89]), it is easy to verify that the stochastic differentials of $y_i \;\; i = 1, \ldots, n$ are (e.g., see [67] pp.82-83)

$$dy_i = (r - \frac{\sigma_i^2}{2})dt + \sigma_i dz_i \qquad\qquad i = 1, \ldots, n. \tag{5.5}$$

Since r and $\sigma_i \;\; i = 1, \ldots, n$ are assumed to be constants, it can be concluded that equation (5.5) represents Itô processes with constant instantaneous drifts and standard deviations. Further, from assumption **A.7** in section 5.2, it is known that $cov(dz_i(t), dz_j(t)) = \rho_{ij}dt$ where $\rho_{ij} \;\; i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$ are constants. Thus, the instantaneuos correlation coefficients between the processes in equation (5.5) are also constants. That is, equation (5.5) is exactly the desired accomplishment with the first step of transformations.

With the second step of transformations, the aim is to transform the processes in equation (5.5) into a system of uncorrelated processes. This is easily accomplished with the help of Cholesky factorization.

Introduce the vector

$$\mathbf{dy} = \begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_n \end{bmatrix}. \tag{5.6}$$

From assumption **A.7** it is known that $z_i \;\; i = 1, \ldots, n$ are Wiener processes. Thus, from equation (5.5) and equation (5.6) it can be seen that $[N_n(\bullet, \bullet)$ denotes the n-variate normal distribution]

$$\mathbf{dy} \sim N_n([\mathbf{r} - \frac{\boldsymbol{\sigma}^2}{\mathbf{2}}]dt, \boldsymbol{\Omega} dt), \tag{5.7}$$

where

$$(\mathbf{r} - \frac{\boldsymbol{\sigma}^2}{\mathbf{2}}) = \begin{bmatrix} (r - \frac{\sigma_1^2}{2}) \\ (r - \frac{\sigma_2^2}{2}) \\ \vdots \\ (r - \frac{\sigma_n^2}{2}) \end{bmatrix},$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \cdots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \sigma_2\sigma_3\rho_{23} & \cdots & \sigma_2\sigma_n\rho_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \cdots & \cdots & \sigma_n^2 \end{bmatrix}.$$

The fact that a linear combination of normal distributed variables itself is normally distributed (e.g., see [88] p.472) implies that[16]

$$\mathbf{du} \equiv \mathbf{Kdy}, \tag{5.8}$$

where $\mathbf{K}$ is a $(n \times n)$-dimensional constant matrix, is distributed as

$$\mathbf{du} \sim N_n(\mathbf{K}(\mathbf{r} - \frac{\boldsymbol{\sigma}^2}{\mathbf{2}})dt, \mathbf{K\Omega K}^{\mathbf{T}}dt) \tag{5.9}$$

($\mathbf{A}^{\mathbf{T}}$ denotes the transpose of the matrix $\mathbf{A}$). If, with *Merton's* terminology, the further non-restrictive assumption is made that all of the original underlying assets are distinct[17] (see [98] p.874), then $\boldsymbol{\Omega}$ will be a symmetric and positive definite matrix.

Since $\boldsymbol{\Omega}$ is a symmetric and positive definite matrix, it can be Cholesky factorized (e.g., see [55] p.108). Thus, $\boldsymbol{\Omega}$ can be rewritten as

$$\boldsymbol{\Omega} = \mathbf{LL}^{\mathbf{T}} = \mathbf{LIL}^{\mathbf{T}}, \tag{5.10}$$

where $\mathbf{I}$ is the $(n \times n)$-dimensional identity matrix and

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & \cdots & 0 \\ l_{21} & l_{22} & 0 & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{n1} & l_{n2} & \cdots & \cdots & l_{n(n-1)} & l_{nn} \end{bmatrix}.$$

Furthermore, since $\boldsymbol{\Omega}$ is a positive definite matrix, it can be inverted, and because $\boldsymbol{\Omega}^{-1} = (\mathbf{LL}^{\mathbf{T}})^{-1} = (\mathbf{L}^{\mathbf{T}})^{-1}\mathbf{L}^{-1} = (\mathbf{L}^{-1})^{\mathbf{T}}\mathbf{L}^{-1}$, the matrix $\mathbf{L}$ also can be inverted. Moreover, since $\mathbf{L}$ is a lower tringular matrix, it is easy to invert.

---

[16]In addition to equation (5.8), it is, of course, assumed that $\mathbf{u}(t_0) = \mathbf{Ky}(t_0)$. That is, the processes u_i $i = 1, \ldots, n$ are given fixed starting points.

[17]With distinct assets, Merton means that none of the assets' returns can be written as an instantaneous linear combination of the other assets' returns.

If one asset's return can be written as an instantaneous linear combination of the other assets' returns, then this asset is redundant and can, without any effect, be excluded from the pricing relation.

If equations (5.9) and (5.10) are observed, it can be seen that setting $\mathbf{K} = \mathbf{L}^{-1}$ is a clever choice. With this choice of $\mathbf{K}$, it is possible to establish that

$$
\begin{aligned}
E(\mathbf{du}) &= \mathbf{K}(\mathbf{r} - \frac{\sigma^2}{2})dt = \mathbf{L}^{-1}(\mathbf{r} - \frac{\sigma^2}{2})dt, \\
Var(\mathbf{du}) &= \mathbf{K\Omega K^T}dt = \mathbf{L}^{-1}(\mathbf{LL^T})(\mathbf{L}^{-1})^{\mathbf{T}}dt = \\
&\qquad (\mathbf{L}^{-1}\mathbf{L})[\mathbf{L^T}(\mathbf{L^T})^{-1}]dt = \mathbf{I}dt.
\end{aligned}
\tag{5.11}
$$

With the help of equations (5.5), (5.8) and (5.11), the stochastic differential of $\mathbf{u}$ can be written as

$$
\mathbf{du} = \nu dt + \mathbf{d}\xi = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix} dt + \begin{bmatrix} d\xi_1 \\ d\xi_2 \\ \vdots \\ d\xi_n \end{bmatrix}, \tag{5.12}
$$

where

$$
\begin{aligned}
\nu &= \mathbf{L}^{-1}(\mathbf{r} - \frac{\sigma^2}{2}), \\
\mathbf{d}\xi &\sim N_n(\mathbf{0}, \mathbf{I}dt).
\end{aligned}
$$

Thus, equation (5.12) gives the desired system of uncorrelated processes.

Previously in section 5.2, it was claimed that the state variables after the two steps of transformations are sufficient to uniquely determine the value of the contingent claim. This can now be shown in a more stringent fashion. With the choice of setting $\mathbf{K} = \mathbf{L}^{-1}$, it is also assured that the matrix $\mathbf{K}$ can be inverted (of course, $\mathbf{K}^{-1} = \mathbf{L}$). Since $\mathbf{K}$ is non-singular, the transformation in equation (5.8) is based on a one-to-one mapping. Furthermore, equation (5.4) shows that the first step of introductory transformations is also based on a one-to-one mapping. From these facts it is clear that there is a one-to-one mapping between the (final) transformed state variables (whose processes are approximated with a multivariate binomial lattice), and the original state variables (the prices of the underlying assets). Thus, there exists an invertible function from the original state variables into the transformed state variables. Denote this function by

$$
\mathbf{u} = g(\mathbf{S}),
$$

$$\Longleftrightarrow \tag{5.13}$$

$$\mathbf{S} = g^{-1}(\mathbf{u}),$$

where $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}^T$, $\mathbf{S} = \begin{bmatrix} S_1 & S_2 & \dots & S_n \end{bmatrix}^T$ and T denotes the transpose[18].

From assumption **A.9** in section 5.2 it is known that $W(\mathbf{S}, t)$, i.e., the price of the contingent claim is a function of $\mathbf{S}$ and time. Equation (5.13) shows that it also is possible to express the price of the contingent claim as a function of the transformed state variables and time. That is

$$H(\mathbf{u}, t) = W(g^{-1}(\mathbf{u}), t) = W(\mathbf{S}, t), \tag{5.16}$$

where $H(\mathbf{u}, t)$ gives the contingent claim price as a function of the transformed state

[18]To be slightly more concrete, with the help of equations (5.4) and (5.8) it is easy to find a rule that gives the unique prices of the underlying securities given the values of the transformed state variables at a given point in time. This rule is

$$\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} = \begin{bmatrix} e^{y_1} \\ e^{y_2} \\ \vdots \\ e^{y_n} \end{bmatrix}, \tag{5.14}$$

where

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{K}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}. \tag{5.15}$$

The rule given by equations (5.14) and (5.15) is essential for starting the solution algorithm, when using the lattice approach developed in this paper. This is due to the fact that the terminal condition that gives the price of the contingent claim at maturity is normally expressed as a function of the prices of the underlying assets. With the help of equations (5.14) and (5.15), the values of the transformed state variables at each node at maturity can be converted into prices of the underlying assets, and the value of the contingent claim at each node at maturity can thus be calculated.

The rule given by equations (5.14) and (5.15) is often also necessary when valuing American type contingent claims. This follows from the fact that the early exercise conditions are usually expressed in terms of the prices of the underlying assets and time. It is therefore needed to convert the values of the transformed state variables into prices of the underlying assets at the nodes where the early exercise condition must be checked.

variables and time.

The description of the transformations perhaps looks somewhat confusing. In reality it is, however, only the first step of the transformations that can be tricky (if that step is possible). (The first step was very simple in this case, since the risk-neutralized processes of the original underlying state variables were assumed to follow geometric Brownian motions.)

The second step is only a matter of plugging in the variance-covariance matrix of the processes transformed in the first step. This follows from the fact that software for Cholesky factorization and inversion of lower triangular matrices are available in many statistical and mathematical packages[19].

Equation (5.12) gives the system of uncorrelated transformed state variables. It is this system that it is desired to approximate with a multivariate binomial lattice, for the purpose of numerically pricing the contingent claim. The next subsection shows how to find the jump sizes in the approximation of equation (5.12) with a multivariate binomial lattice.

5.2.2 Derivation of the jump sizes

In the approach developed in this paper, all jump probabilities, for each branch at each node, are set to be equal. This means that the jump sizes are the only additional information required to completely specify the multivariate binomial lattice, which is used to approximate system (5.12). The principle for the derivation of the jump sizes is as follows:

The time to maturity for the contingent claim has been divided into N equally long time periods. Consider one time period. Approximate the continuous end of period distribution of system (5.12) (conditional on the system variables at the beginning of the period)

[19]The algorithm for Cholesky factorization (e.g., see [55] p.108 ff.) and the algorithm for inversion of lower triangular matrices (e.g., see [116] pp.45-46) are not complex. With some programming experience, it is hence easy to write the programmes oneself.

with a 2^n-jump distribution. This 2^n-jump distribution is to converge to the conditional end of period distribution as the length of the time period approaches zero. Since the jump probabilities have already been specified as equal, only the appropriate values of the jump sizes have to be derived in order to make the convergence become true.

BEG (see [16]) show that the characteristic function can, for some distributions, be a very useful tool when assuring the convergence of one distribution towards another distribution. The characteristic function will be used in the subsequent derivation of the jump sizes.

Consider a time period, and introduce the variable

$$R_i = u_i - \bar{u}_i \qquad\qquad i = 1, \ldots, n,$$

where $u_i \;\; i = 1, \ldots, n$ is the stochastic outcome for transformed variable i at the end of the time period, and $\bar{u}_i \;\; i = 1, \ldots, n$ is its given outcome (more precisely, everything in the derivation is conditional on $\bar{u}_i \;\; i = 1, \ldots, n$) at the beginning of the time period.

From equation (5.12) it follows directly that:

$$\mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} \sim N_n(\nu k, \mathbf{I}k),$$

where $k = T/N$ is the length of the time period, ν is defined in equation (5.12) and $\mathbf{I}$ is the $(n \times n)$-dimensional identity matrix.

Introduce the characteristic function for the random variable $\mathbf{R}$ (e.g., see [88] p.151)

$$\Psi(\iota\Theta_1, \ldots, \iota\Theta_n) = E(e^{(\sum_{j=1}^{n} \iota\Theta_j R_j)}), \tag{5.17}$$

where ι is the imaginary unit, $\sqrt{-1}$, and $\Theta_j \in \Re \;\; j = 1, \ldots, n$.

The characteristic function has a simple analytical form in the case of a jointly normal distributed random variable (e.g., see [88] p.472). This simple analytical form indicates that the characteristic function for $\mathbf{R}$ can be rewritten as

$$\Psi_c(\iota\Theta_1, \ldots, \iota\Theta_n) = e^{(\iota k \sum_{j=1}^{n} \Theta_j \nu_j - \frac{k}{2} \sum_{j=1}^{n} \Theta_j^2)}. \tag{5.18}$$

In order to derive equation (5.18), two facts were used. Firstly, all $R_i \quad i = 1, \ldots, n$ are uncorrelated. Secondly, all $R_i \; i = 1, \ldots, n$ have variances equal to k.

If the length of the time period approaches zero, equation (5.18) can be approximated with its first order Maclaurin expansion. That is,

$$\Psi_c(\iota\Theta_1, \ldots, \iota\Theta_n) = 1 + \iota k \sum_{j=1}^{n} \Theta_j \nu_j - \frac{k}{2} \sum_{j=1}^{n} \Theta_j^2 + O(k), \tag{5.19}$$

where the $O(k)$ term can be ignored as k approaches zero[20].

For the discrete 2^n-jump distribution, the value of each of the transformed state variables can either jump up or down[21]. It is the sizes of these jumps which must be derived. A clever functional form[22] for the jump sizes, as will be shown shortly, is to require that the size of the up jump in the value of state variable i satisfies

$$up_i = b_i a_i \sqrt{k} + a_i k \qquad i = 1, \ldots, n, \tag{5.20}$$

and that the size of a down jump in the value of state variable i satisfies

$$do_i = -b_i a_i \sqrt{k} + a_i k \qquad i = 1, \ldots, n. \tag{5.21}$$

With the structure imposed by equations (5.20) and (5.21), the problem of finding the jump sizes is reduced to the problem of finding the appropriate expressions for the parameters $a_i, b_i \in \Re \; i = 1, \ldots, n$.

---

[20] $O(k)$ is an asymptotic order symbol, which can roughly be said to include all terms with a factor $k^\alpha \;\; \alpha > 1$.

[21] It might be misleading to call the jumps up and down. Actually, both jumps can be up, both jumps can be down or one jump can be up and the other down. The up jump is, however, always more up or less down than the down jump. That is, if both jumps are up, then the up jump is more up than the down jump and if both jumps are down then the up jump is less down than the down jump.

[22] The choice of functional form for the jump sizes in this paper is related to the choice of jump sizes in the classical paper by CRR (see [43]). CRR require, in their single state variable model, that

$$up = \sigma\sqrt{k},$$
$$down = -\sigma\sqrt{k},$$

where σ is the instantaneous standard deviation of the rate of return on the underlying stock (the stock price is assumed to follow a geometric Brownian motion).

As a help for writing down the characteristic function for the discrete approximation, the following function is introduced:

$$\delta_i(l) = \begin{cases} +1 & \text{if state variable } i \text{ jumps upwards at branch } l \\ -1 & \text{if state variable } i \text{ jumps downwards at branch } l \end{cases} \tag{5.22}$$
$$i = 1, \ldots, n; l = 1, \ldots, m.$$

Equations (5.17), (5.20), (5.21) and (5.22) imply that the characteristic function for the 2^n-jump distribution can be written as

$$\Psi_d(\iota\Theta_1, \ldots, \iota\Theta_n) = \frac{1}{m}\sum_{l=1}^{m} e^{(\sum_{j=1}^{n}[\delta_j(l)a_j b_j\sqrt{k}+a_j k]\Theta_j \iota)}. \tag{5.23}$$

In equation (5.23), the fact that all branches have the same probability has been used, i.e., the probability for each branch is equal to $\frac{1}{m} = \frac{1}{2^n}$.

Equation (5.23) can also be approximated with a Maclaurin expansion as the length of the time step approaches zero. In this case, however, a second order expansion is used. Thus,

$$\begin{aligned}
&\Psi_d(\iota\Theta_1, \ldots, \iota\Theta_n) = \\
&\frac{1}{m}\sum_{l=1}^{m}\{1 + \sum_{j=1}^{n}[\delta_j(l)a_j b_j\sqrt{k} + a_j k]\Theta_j\iota - \\
&\frac{1}{2}(\sum_{j=1}^{n}[\delta_j(l)a_j b_j\sqrt{k} + a_j k]\Theta_j)^2\} + O(k) = \\
\frac{1}{m}\sum_{l=1}^{m}\{1 + \sum_{j=1}^{n}&[\delta_j(l)a_j b_j\sqrt{k} + a_j k]\Theta_j\iota - \frac{k}{2}[\sum_{j=1}^{n}\delta_j(l)a_j b_j\Theta_j]^2\} + O(k) = \\
1 + \frac{1}{m}\sum_{l=1}^{m}\sum_{j=1}^{n}&[\delta_j(l)a_j b_j\sqrt{k} + a_j k]\Theta_j\iota - \frac{k}{2m}\sum_{l=1}^{m}\sum_{j=1}^{n}a_j^2 b_j^2\Theta_j^2 - \\
&\frac{k}{m}\sum_{l=1}^{m}\sum_{j=1}^{n-1}\sum_{s=j+1}^{n}\delta_{js}(l)a_j b_j a_s b_s\Theta_j\Theta_s + O(k),
\end{aligned} \tag{5.24}$$

where

$$\delta_{js}(l) = \begin{cases} +1 & \text{if state variables } j \text{ and } s \text{ jump in the same direction at branch } l \\ -1 & \text{if state variables } j \text{ and } s \text{ jump in opposite directions at branch } l \end{cases}$$
$$j = 1, \ldots, n-1; s = 2, \ldots, n; j < s; l = 1, \ldots, m,$$

and the $O(k)$ term can be ignored as k approaches zero.

It is a well-known fact that there is a one-to-one correspondence between distribution functions and characteristic functions (e.g., see [114] p.12). Thus, to ensure that the 2^n-jump distribution converges to the continuous distribution of $\mathbf{R}$ as the length of the time period approaches zero, equation (5.19) is set equal to equation (5.24) [the $O(k)$ terms can be ignored as $k \longrightarrow 0$]. If this is done and like coefficients are equated, the result is as follows:

- terms with $\Theta_j \iota$:

$$\frac{1}{m}\sum_{l=1}^{m}[\delta_j(l)a_jb_j\sqrt{k} + a_jk] = \nu_jk \qquad j = 1,\ldots,n. \tag{5.25}$$

- terms with $\frac{k}{2}\Theta_j^2$:

$$\frac{1}{m}\sum_{l=1}^{m} a_j^2b_j^2 = 1 \qquad j = 1,\ldots,n. \tag{5.26}$$

- terms with $\Theta_j\Theta_sk$:

$$\frac{1}{m}\sum_{l=1}^{m} \delta_{js}(l)a_jb_ja_sb_s = 0 \qquad j = 1,\ldots,n-1; s = 2,\ldots,n; j < s. \tag{5.27}$$

To attain more useful formulas, equations (5.25), (5.26) and (5.27) will be rewritten. Thus, equation (5.25) can be rewritten as

$$\begin{aligned} \frac{a_jb_j\sqrt{k}}{m}\sum_{l=1}^{m}\delta_j(l) + \frac{a_jk}{m}\sum_{l=1}^{m}1 &= \nu_jk \qquad j = 1,\ldots,n, \\ &\Longleftrightarrow \\ a_j &= \nu_j \qquad j = 1,\ldots,n. \end{aligned} \tag{5.28}$$

The implication in equation (5.28) follows from the fact that $\sum_{l=1}^{m}\delta_j(l) = 0$ $j = 1,\ldots,n$, since each state variable jumps up at an equal number of branches as it jumps down.

Equation (5.26) can be rewritten as

$$\begin{aligned} \frac{a_j^2b_j^2}{m}\sum_{l=1}^{m}1 &= 1 \qquad j = 1,\ldots,n, \\ &\Longleftrightarrow \end{aligned}$$

$$
\begin{aligned}
a_j^2 b_j^2 &= 1 \qquad\qquad j = 1, \ldots, n, \\
&\Longleftrightarrow \\
b_j &= \pm\frac{1}{a_j} \qquad\qquad j = 1, \ldots, n.
\end{aligned} \tag{5.29}
$$

Finally, equation (5.27) can be rewritten as

$$
\begin{aligned}
\frac{a_j b_j a_s b_s}{m} \sum_{l=1}^{m} \delta_{js}(l) &= 0 \qquad j = 1, \ldots, n-1; s = 2, \ldots, n; j < s, \\
&\Longleftrightarrow \\
0 &= 0 \qquad j = 1, \ldots, n-1; s = 2, \ldots, n; j < s.
\end{aligned} \tag{5.30}
$$

The implication in equation (5.30) follows from the fact that $\sum_{l=1}^{m} \delta_{js}(l) = 0 \;\; j = 1, \ldots, n-1; s = 2, \ldots, n; j < s$, since state variables j and s jump in the same direction at exactly the same number of branches as they jump in opposite directions. Equation (5.30) shows that equation (5.27) does not give any useful information in the search for a_j and $b_j \;\; j = 1, \ldots, n$. Equation (5.30), however, also shows that equation (5.27) is consistent.

From equation (5.28) and equation (5.29), it can be concluded that the parameters are determined by the very simple formulas

$$
\begin{aligned}
a_j &= \nu_j \qquad\qquad j = 1, \ldots, n, \\
b_j &= \frac{1}{\nu_j} \qquad\qquad j = 1, \ldots, n.
\end{aligned} \tag{5.31}
$$

In equation (5.31), the positive root in equation (5.29) is chosen. With this choice, it can be established that the size of the up jump will always be more up (or less down) than the size of the down jump.

To obtain the desired jump sizes, substitute equation (5.31) into equation (5.20) and equation (5.21). It then follows that with

$$
\begin{aligned}
up_j &= \sqrt{k} + \nu_j k \qquad\qquad j = 1, \ldots, n, \\
do_j &= -\sqrt{k} + \nu_j k \qquad\qquad j = 1, \ldots, n,
\end{aligned} \tag{5.32}
$$

the 2^n-jump distribution will converge towards the continuous distribution of $\mathbf{R}$ as the length of the time period approaches zero. Thus, the formulas for the appropriate jump sizes are *extremely* simple, at least compared to the resulting formulas for the jump probabilities in the BEG approach {see [16] equation (15)}.

5.3 Numerical evaluations

The numerical evaluations in this paper have two purposes. The first and main purpose is, of course, to give an understanding of the accuracy of the lattice approach developed. The second purpose is to compare the accuracy of the method developed in this paper to the accuracy of the BEG approach. This last purpose is important since it has been claimed that the approach developed in this paper can be viewed as an improvement of the BEG approach.

Both of the above purposes are achieved by applying the lattice approaches to some of the rare multivariate contingent claims that also have closed-form solutions. Thus, the numerical solutions from the method developed in this paper and from the BEG approach can be compared to each other but also to the analytical solution[23] for each set of parameter values and choice of number of time steps.

The multivariate contingent claims that the lattice approaches are applied to are European call options on the maximum of three and four risky assets, European call options on the minimum of three and four risky assets and European put options on the minimum of three and four risky assets. For all of these contingent claims, it is possible to derive closed-form solutions for their values.

[75] is referred to for the closed-form solutions for European call options on the maximum of several risky assets and for European call options on the minimum of several risky assets. The closed-form solution for European put options on the minimum of several risky assets has, to the author's knowledge, not appeared in the literature. This closed-form solution can, however, easily be derived by the help of the ingenious trick suggested by *Johnson* in [75]. Thus, use of Johnson's trick implies that the closed-form solution for the value of a European put option on the minimum of several risky assets can be written as

---

[23]Actually, the values from the closed-form solutions are also achieved by a numerical method. This means that when values from the lattice approaches are compared to values from the closed-form solutions, in reality values from different numerical methods are compared.

$$W(S_1(t),\ldots,S_n(t),t) =$$

$$Œe^{-r[T-(t-t_0)]}(1-\Phi_n[d_2(S_1),\ldots,d_2(S_n),\rho_{12},\ldots,\rho_{1n},\rho_{23},\ldots,\rho_{(n-1)n}]) - \tag{5.33}$$
$$\sum_{i=1}^{n} S_i\Phi_n[-d_1(S_i),d_2'(S_1,S_i),\ldots,d_2'(S_n,S_i),\rho_{ii1},\ldots,\rho_{iin},\rho_{i12},\ldots,\rho_{i(n-1)n}],$$

where

$$\begin{aligned}
d_1(S_i) &= \frac{ln(\frac{S_i(t)}{Œ})+(r+\frac{1}{2}\sigma_i^2)[T-(t-t_0)]}{\sigma_i\sqrt{T-(t-t_0)}} && i=1,\ldots,n,\\
d_2(S_i) &= d_1(S_i)-\sigma_i\sqrt{T-(t-t_0)} && i=1,\ldots,n,\\
d_2'(S_i,S_j) &= \frac{ln(\frac{S_i(t)}{S_j(t)})-\frac{1}{2}\sigma_{ij}^2[T-(t-t_0)]}{\sigma_{ij}\sqrt{T-(t-t_0)}} && i=1,\ldots,n; j=1,\ldots,n; i\neq j,\\
\sigma_{ij}^2 &= \sigma_i^2-2\rho_{ij}\sigma_i\sigma_j+\sigma_j^2 && i=1,\ldots,n; j=1,\ldots,n; i\neq j,\\
\rho_{iij} &= \frac{\sigma_i-\rho_{ij}\sigma_j}{\sigma_{ij}} && i=1,\ldots,n; j=1,\ldots,n; i\neq j,\\
\rho_{ijk} &= \frac{\sigma_i^2-\rho_{ij}\sigma_i\sigma_j-\rho_{ik}\sigma_i\sigma_k+\rho_{jk}\sigma_j\sigma_k}{\sigma_{ij}\sigma_{ik}} && i=1,\ldots,n; j=1,\ldots,n-1;\\
& && k=2,\ldots,n; i\neq j; i\neq k; j<k,
\end{aligned}$$

Œ is the exercise price, and $\Phi_n[\bullet]$ is the n-variate standardized cumulative normal distribution function. The remaining notation in equation (5.33) is explained previously in the text.

In addition to applying the lattice methods to the pricing of the contingent claims above, the method developed in this paper is also applied to the pricing of American put options on the minimum of three risky assets. In this case, no closed-form solution exists. It is hence not possible to get a hint of the accuracy of the lattice method by comparing the solutions from the lattice method to the analytical solutions. The aim of these valuations is instead to show that the lattice approach developed in this paper can easily handle American type contingent claims.

The efficiency of a numerical method is usually affected by the values of the parameters. That is, the same method can have fast convergence for one specific set of parameter values while it has slow convergence for another set. For this reason, valuations are performed for different sets of parameter values. The number of parameters is, however, large when valuing multivariate contingent claims. For the purpose of reducing the range of possible sets of parameter values, the following strategy is used:

In all valuations the riskfree rate of return, the time to maturity, the exercise price and the current prices of the underlying assets remain unchanged. Thus, the parameters to vary are the instantaneous variances and correlation coefficients. The instantaneous variances for each asset's rate of return are, however, equal in a given valuation. To be more precise:

- Parameters to remain unchanged:
  - riskfree rate = $r = 0.1$
  - time to maturity = $T = 1.0$
  - exercise price = Œ $= 10.0$
  - current prices = $S_i(t_0) = 10.0 \qquad i = 1, \ldots, n$
- Parameters to vary:
  - standard deviations = $\sigma_i \qquad i = 1, \ldots, n$
    (however, in each valuation are $\sigma_i = \sigma_j \;\; i = 1, \ldots, n; j = 1, \ldots, n$)
  - correlation coefficients = $\rho_{ij} \qquad i = 1, \ldots, n-1; j = 2, \ldots, n; i < j$

A lattice method's rate of convergence is used here to mean the rate with which the numerical solution approaches the accurate solution as the number of time steps used in the lattice method increases. To be able to compare this very important property between the method developed in this paper and the BEG approach, valuations for each set of parameter values are performed with several different number of time steps.

5.3.1 Numerical evaluations of contingent claims with three underlying state variables

The results from the evaluations of contingent claims with three underlying state variables are presented in table 5.1.

The discussion about the performance of the lattice method developed in this paper will begin without involving the BEG approach. In table 5.1, the option values computed with the lattice method developed in this paper are given in the column marked with NEK. (In the remainder of the paper the lattice method developed in this paper is called the NEK approach.)

By just glancing at table 5.1, it can immediately be concluded that the NEK approach works extremely well. Moreover, if the numerical results are observed in some detail, the good performance of the NEK approach becomes even more clear.

In this paper, three different types of options that also have closed-form solutions are valued. Each option type is valued with five different sets of parameter values. Thus, there are fifteen different cases for which the values from the NEK approach can be compared to the values from the closed-form solutions[24].

In table 5.1, it can be seen that the NEK approach gives results that satisfy most needs after as few as 10 time steps. To be more concrete, table 5.1 indicates that the

---

[24]As mentioned earlier, the values from the closed-form solutions are achieved by a numerical method. For options with three underlying risky assets, the 3-variate standardized cumulative normal distribution function has to be evaluated no less than four times.

In this paper, the 3-variate standardized cumulative normal distribution function is evaluated with the help of numerical integration (for more details see [119]). This algorithm for evaluating the multivariate normal distribution function guarantees a result with an error smaller than 1×10^{-4}. (In most cases, of course, the error is much smaller.)

Thus, the closed-form solutions for the options [see equation (5.33); the closed-form solutions for the price of a European call option on the maximum of several risky assets and the price of a European call option on the minimum of several risky assets are very similar to equation (5.33)] with three underlying assets indicate that if every evaluation of the 3-variate standardized cumulative normal distribution function gives the maximal error and all 4 errors work in the same direction when the "analytical" option price is calculated, a *very* high upper bound for the error in the analytical option prices is given by

$$bound_{up} = \sum_{i=1}^{3} S_i(1 \times 10^{-4}) + Œe^{-rT}(1 \times 10^{-4}) \approx 3.9 \times 10^{-3}. \tag{5.34}$$

The figure given in equation (5.34) is, indeed, the *worst* possible case. Most often when the numerical method is used for evaluating the analytical formulas, the third decimal in the option price will be correct, and in nearly all cases the error in the option price is no larger than $\pm 1 \times 10^{-3}$.

$r = 0.1$; $S_1(t_0) = S_2(t_0) = S_3(t_0) =$ Œ $= 10.0$; $T = 1.0$							
$\sigma_1 = \sigma_2 = \sigma_3 = 0.2$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.1$							
N	Eu. call max.		Eu. call min.		Eu. put min.		Am. put min.
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG	NEK
10	2.627	2.575	0.297	0.271	0.891	0.855	1.013
20	2.630	2.603	0.282	0.277	0.881	0.869	1.011
50	2.631	2.620	0.286	0.281	0.887	0.878	1.018
100	2.631	2.625	0.283	0.282	0.884	0.881	1.017
Analytical	2.631		0.284		0.884		
$\sigma_1 = \sigma_2 = \sigma_3 = 0.2$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.7$							
N	Eu. call max.		Eu. call min.		Eu. put min.		Am. put min.
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG	NEK
10	2.045	1.926	0.700	0.697	0.660	0.615	0.758
20	2.042	1.989	0.693	0.695	0.653	0.634	0.757
50	2.042	2.021	0.694	0.694	0.653	0.644	0.760
100	2.041	2.031	0.693	0.694	0.651	0.648	0.760
Analytical	2.041		0.694		0.652		
$\sigma_1 = \sigma_2 = \sigma_3 = 0.1$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$							
N	Eu. call max.		Eu. call min.		Eu. put min.		Am. put min.
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG	NEK
10	1.570	1.513	0.537	0.521	0.179	0.159	0.277
20	1.569	1.542	0.535	0.529	0.176	0.168	0.287
50	1.569	1.558	0.537	0.534	0.177	0.173	0.293
100	1.569	1.563	0.537	0.535	0.177	0.175	0.295
Analytical	1.568		0.537		0.177		
$\sigma_1 = \sigma_2 = \sigma_3 = 0.4$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$							
N	Eu. call max.		Eu. call min.		Eu. put min.		Am. put min.
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG	NEK
10	3.795	3.668	0.623	0.635	1.945	1.869	2.038
20	3.799	3.739	0.623	0.628	1.944	1.905	2.043
50	3.801	3.775	0.623	0.625	1.942	1.926	2.045
100	3.801	3.788	0.623	0.624	1.941	1.933	2.045
Analytical	3.801		0.623		1.940		
$\sigma_1 = \sigma_2 = \sigma_3 = 0.2$; $\rho_{12} = 0.8$; $\rho_{13} = 0.6$; $\rho_{23} = 0.3$							
N	Eu. call max.		Eu. call min.		Eu. put min.		Am. put min
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG	NEK
10	2.162	neg. pr.	0.587	neg. pr.	0.702	neg. pr.	0.805
20	2.159	neg. pr.	0.584	neg. pr.	0.696	neg. pr.	0.807
50	2.159	neg. pr.	0.584	neg. pr.	0.695	neg. pr.	0.809
100	2.158	neg. pr.	0.584	neg. pr.	0.694	neg. pr.	0.809
Analytical	2.158		0.584		0.694		

Table 5.1: Numerical evaluations of options with three underlying state variables.

Note: neg. pr. denotes that the BEG approach leads to negative jump probabilities.

NEK approach has the following accuracy with 10 time steps:

- The maximum error[25] is $= 13 \times 10^{-3}$.
- In 14 (out of 15) cases the error is $\leq 8 \times 10^{-3}$.
- In 11 cases the error is $\leq 6 \times 10^{-3}$.
- In 8 cases the error is $\leq 4 \times 10^{-3}$.

From table 5.1 it can be seen that with 20 time steps the NEK approach has the following accuracy:

- In all (15) cases the error is $\leq 4 \times 10^{-3}$.
- In 14 cases the error is $\leq 3 \times 10^{-3}$.
- In 13 cases the error is $\leq 2 \times 10^{-3}$.
- In 9 cases the error is $\leq 1 \times 10^{-3}$.

That is, with 20 time steps there is no case where the error is larger than 4×10^{-3}, and in $\approx 87\%$ of the cases the error is less than or equal to 2×10^{-3}. Thus, already after 20 time steps the results from the NEK approach are astonishingly accurate.

With 50 time steps, all evaluations with the NEK approach are very accurate. Table 5.1 indicates that:

- In all (15) cases the error is $\leq 3 \times 10^{-3}$.
- In 14 cases the error is $\leq 2 \times 10^{-3}$.
- In 12 cases the error is $\leq 1 \times 10^{-3}$.
- In 7 cases there is no error [down to (and including) the third decimal].

---

[25]The absolute value of the deviation between the value from a lattice approach and the value from the analytical formula is called error.

Finally, with 100 time steps the NEK approach gives (in principle) values that are correct down to (and including) the third decimal in all cases. Table 5.1 shows that there are some cases where the error is equal to 1×10^{-3}, but these errors can be due as well to rounding errors or errors in the analytical values.

The findings for the NEK approach can be summed up as follows:

The NEK approach shows a very good performance. The values from the approach are accurate enough for most purposes after as few as 10 time steps. With 20 time steps, the maximum error is 4×10^{-3}, and the error is less than or equal to 2×10^{-3} in approximate 87% of the cases. The results are further improved with 50 time steps, and with 100 time steps the NEK approach gives values that (in principle) are correct down to (and including) the third decimal.

Next, to the comparison between the NEK approach and the BEG approach. For this comparison there are only 12 cases available, since the BEG approach leads to negative jump probabilities for one set of parameter values.

The fact that the BEG approach can lead to negative jump probabilities is a big disadvantage for the method, when there are more than two underlying assets. It is not hard to find sets of parameter values which give negative jump probabilities when there are three underlying assets. Moreover, the problem increases as the number of underlying assets is further increased.

When a set of parameter values leads to negative jump probabilities and this fact is disregarded, often nothing dramatic happens. That is, the BEG approach will be approximately as efficient as when all jump probabilities are positive. There are, however, several reasons to be very cautious when there are negative jump probabilities. Consider the following aspects:

Negative jump probabilities are conceptually completely wrong. This means that the derivation of the BEG approach does not hold with negative jump probabilities. Thus, we cannot use BEG's derivation as a guarantee for convergence as the number of time steps increases in these cases.

There is also a problem of stability[26] related to convergence. In [22], *Brennan and Schwartz* show that the explicit finite difference method is equivalent to a trinomial lattice approach, in the case of one underlying state variable. To guarantee stability in the explicit finite difference method (with one underlying state variable), it can be shown that all three coefficients that relate known contingent claim values in the previous time step to the unknown contingent claim value in the subsequent time step (i.e., the jump probabilities in the trinomial lattice interpretation) must be non-negative (see [22] p.464 or [69] p.92). If this feature can be extended to the BEG approach, which is reasonable to believe (at least to some extent), pricing problems exist within CCA for which the BEG approach is unstable irrespectively of how small the time step, k, is made[27].

If table 5.1 is observed and the results from the NEK approach are compared to the results from the BEG approach, the following conclusions can be made:

- In 8 (out of 12) cases the value from the NEK approach with 10 time steps is at least as accurate (more accurate in 7 cases) as the value from the BEG approach with 100 time steps.

- In 11 (out of 12) cases the value from the NEK approach with 20 time steps is at least as accurate as the value from the BEG approach with 100 time steps.

- The remaining case is somewhat odd. It is the case of a European call option on the minimum of three risky assets with $\sigma_i = 0.2$ $i = 1,2,3$ and $\rho_{ij} = 0.7$ $i = 1,2; j = 2,3; i < j$. From the table it can be seen that the error in the value from the NEK approach is 1×10^{-3} already after 20 time steps, and that the error still is 1×10^{-3}

[26]In short, the problem of stability concerns the unstable growth or stable decay of the errors in the arithmetical operations needed when a numerical method is used. Each calculation carried out introduces a round-off error. Generally, a numerical method is stable when the cumulative effect of all the rounding errors introduced when the method is used is negligible (see [126] p.57).

[27]In many applications with finite difference methods (a lattice approach can be interpreted as a form of explicit finite difference method), the conditions for convergence and stability are the same (see [69] p.92). Thus, the convergence problem and the stability problem of the BEG approach may actually be the same problem in two different forms.

after 100 time steps. Thus, this is probably a case where the analytical value has a error of $+1 \times 10^{-3}$. However, both methods work very well in this case.

The comparison between the NEK approach and the BEG approach can be summed up in the following way:

The BEG approach has the disadvantage that it can lead to negative jump probabilities when there are more than two underlying assets. Moreover, this problem increases as further assets are added.

The NEK approach is, however, superior to the BEG approach, even if the possibility of negative jump probabilities is disregarded when using the BEG approach. This is due to the NEK approach's fast convergence. The NEK approach has a rate of convergence that is more than 10 times as fast as the rate of convergence for the BEG approach in a majority of the cases. In the remaining cases (with the exception of the odd case previously discussed), the rate of convergence for the NEK approach is at least 5 times as fast as the rate of convergence for the BEG approach.

In table 5.1, some valuations of American put options on the minimum of three risky assets with the NEK approach are presented. There is no closed-form solution for the price of this type of instrument. There are therefore no analytical values to compare the values from the NEK approach to. Instead, the purpose of doing these valuations is to show that the NEK approach can easily handle options of the American type.

5.3.2 Numerical evaluations of contingent claims with four underlying state variables

The results of evaluations of options with four underlying assets are given in table 5.2.

As in subsection 5.3.1, the discussion will begin with the performance of the NEK approach without involving the BEG approach. Table 5.2 confirms that the NEK approach also shows very good performance for options with four underlying assets.

Three different option types, with four underlying assets that have closed-form solutions are valued. Moreover, these options are valued with five different sets of parameter values[28]. Thus, there are 15 cases where the values from the NEK approach can be compared to analytical values[29].

Table 5.2 provides the following facts regarding the accuracy of the NEK approach with 10 time steps:

- The maximum error is $= 12 \times 10^{-3}$.
- In 14 (out of 15) cases the error is $\leq 9 \times 10^{-3}$.
- In 13 cases the error is $\leq 7 \times 10^{-3}$.

[28]Compare the fourth sets of parameter values (the sets with $\sigma_i = 0.4 \;\; i = 1, \ldots, n$) between table 5.1 and table 5.2. It may seem strange that all $\rho_{ij} = 0.5$ in table 5.1 whereas this is not the case in table 5.2. This is due to the fact that with all σ_i equal and all $\rho_{ij} = 0.5$, some of the jump probabilities become equal to zero in the BEG approach when there are four underlying assets. This is exactly the case for the third set of parameter values in table 5.2. Therefore, to avoid two sets of parameter values that give some of the jump probabilities equal to zero all ρ_{ij} are not equal to 0.5 in the fourth set in table 5.2.

[29]As was discussed in a footnote in subsection 5.3.1, the analytical values are achieved by a numerical method. When there are four underlying assets, the 4-variate standardized cumulative normal distribution function must be evaluated as many as five times. The algorithm used for evaluating the multivariate normal distribution function guarantees an error that is smaller than 1×10^{-4}. (In *most* cases the error is much smaller.)

Thus, the closed-form solutions for the options [see equation (5.33); the closed-form solutions for the price of a European call option on the maximum of several risky assets and the price of a European call option on the minimum of several risky assets are very similar to equation (5.33)] with four underlying assets indicate that if every evaluation of the 4-variate standardized cumulative normal distribution function gives the maximal error, and all 5 errors work in the same direction when the "analytical" option price is calculated, a *very* high upper bound for the error in the analytical option prices is given by

$$bound_{up} = \sum_{i=1}^{4} S_i(1 \times 10^{-4}) + \text{Œ} e^{-rT}(1 \times 10^{-4}) \approx 4.9 \times 10^{-3}. \tag{5.35}$$

The figure given in equation (5.35) is, indeed, the *worst* possible case. Most of the times when the numerical method is used for evaluating the analytical formulas, the third decimal in the option price will be correct. The possibility of getting an error in the third decimal is, of course, a little bit greater with four underlying assets than with three underlying assets.

$r = 0.1$; $S_1(t_0) = S_2(t_0) = S_3(t_0) = S_4(t_0) = \mathcal{E} = 10.0$; $T = 1.0$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2$; $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = \rho_{34} = 0.1$						
N	Eu. call max.		Eu. call min.		Eu. put min.	
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG
10	2.998	2.935	0.172	0.151	1.065	1.036
20	3.007	2.973	0.164	0.158	1.061	1.049
40	3.010	2.992	0.167	0.162	1.066	1.056
Analytical	3.010		0.165		1.064	

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2$; $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = \rho_{34} = 0.7$						
N	Eu. call max.		Eu. call min.		Eu. put min.	
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG
10	2.230	neg. pr.	0.592	neg. pr.	0.738	neg. pr.
20	2.228	neg. pr.	0.588	neg. pr.	0.732	neg. pr.
40	2.228	neg. pr.	0.589	neg. pr.	0.732	neg. pr.
Analytical	2.227		0.589		0.731	

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.1$; $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = \rho_{34} = 0.5$						
N	Eu. call max.		Eu. call min.		Eu. put min.	
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG
10	1.702	1.620	0.448	0.418	0.213	0.195
20	1.701	1.662	0.447	0.434	0.210	0.203
40	1.700	1.682	0.447	0.441	0.210	0.207
Analytical	1.700		0.449		0.211	

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.4$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.7$; $\rho_{14} = \rho_{24} = \rho_{34} = 0.2$						
N	Eu. call max.		Eu. call min.		Eu. put min.	
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG
10	4.363	4.189	0.391	0.391	2.217	2.127
20	4.367	4.283	0.390	0.390	2.213	2.169
40	4.368	4.326	0.390	0.390	2.210	2.188
Analytical	4.368		0.390		2.208	

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2$; $\rho_{12} = \rho_{13} = \rho_{23} = 0.7$; $\rho_{14} = \rho_{24} = \rho_{34} = 0.5$						
N	Eu. call max.		Eu. call min.		Eu. put min.	
(# time steps)	NEK	BEG	NEK	BEG	NEK	BEG
10	2.378	neg. pr.	0.492	neg. pr.	0.797	neg. pr.
20	2.375	neg. pr.	0.489	neg. pr.	0.792	neg. pr.
40	2.375	neg. pr.	0.490	neg. pr.	0.792	neg. pr.
Analytical	2.374		0.490		0.791	

Table 5.2: Numerical evaluations of options with four underlying state variables.

Note: neg. pr. denotes that the BEG approach leads to negative jump probabilities.

- In 9 cases the error is $\leq 4 \times 10^{-3}$.
- In 6 cases the error is $\leq 2 \times 10^{-3}$.

Thus, the results from the NEK approach are sufficiently accurate for most needs after as few as 10 time steps also when there are four underlying state variables.

If table 5.2 is considered, it can be seen that with 20 time steps the results from the NEK approach establish the following facts:

- In all (15) cases the error is $\leq 5 \times 10^{-3}$.
- In 14 cases the error is $\leq 3 \times 10^{-3}$.
- In 12 cases the error is $\leq 2 \times 10^{-3}$.
- In 11 cases the error is $\leq 1 \times 10^{-3}$.

That is, with 20 time steps the maximum error is equal to 5×10^{-3}, in 80% of the cases the error is $\leq 2 \times 10^{-3}$, and in $\approx$ 73% of the cases the error is $\leq 1 \times 10^{-3}$. Therefore, the values from the NEK approach are also astonishingly accurate with 20 time steps for options with four underlying assets.

From table 5.2 it can be seen that with 40 time steps the NEK approach has the following accuracy:

- In all (15) cases the error is $\leq 2 \times 10^{-3}$.
- In 11 cases the error is $\leq 1 \times 10^{-3}$.

Thus, the NEK approach is, indeed, *very* accurate with 40 time steps.

The findings regarding the NEK approach for options with four underlying risky assets can be summed up in the following way:

Already after 10 time steps, the values from the approach are accurate enough for most needs. With 20 time steps the maximum error is equal to 5×10^{-3}, in 80% of the cases the error is $\leq 2 \times 10^{-3}$ and in $\approx$ 73% of the cases the error is $\leq 1 \times 10^{-3}$. The error is not larger than 2×10^{-3} in any case, when the NEK approach is used with 40 time steps.

From the aforementioned facts, it is clear that the NEK approach performs well.

The next step is to compare the NEK approach to the BEG approach. There are only 9 cases available for this comparison. This is because the BEG approach leads to negative jump probabilities for two of the sets of parameter values. (Subsection 5.3.1 provides a discussion of the disadvantages of possible negative jump probabilities for the BEG approach.)

If table 5.2 is considered and the accuracy of the NEK approach is compared to the accuracy of the BEG approach, it can be seen that:

- In 7 (out of 9) cases the value from the NEK approach with 10 time steps is more accurate (in most of these cases substantially more accurate) than the value from the BEG approach with 40 time steps.

- In 1 case the value from the NEK approach with 20 time steps is better (without being better with 10 time steps) than the value from the BEG approach with 40 time steps. This case is a European call option on the minimum of four assets with $\sigma_i = 0.2 \;\; i = 1, \ldots, 4$ and $\rho_{ij} = 0.1 \;\; i = 1, 2, 3; j = 2, 3, 4; i < j$. As can be seen from the table, this is not due to bad performance of the NEK approach. Rather, this is one case where the BEG approach has good accuracy.

- The remaining case is a European call option on the minimum of four assets with $\sigma_i = 0.4 \;\; i = 1, \ldots, 4$. In this case, both methods have very good (extremely good by the BEG approach's standards) accuracy. Both methods give exactly the same value with all different number of time steps, and the error is never larger than 1×10^{-3}.

From the above facts, it should be clear that the NEK approach has substantially faster convergence than the BEG approach. In a large majority of the cases, the rate of convergence for the NEK approach is far more than 4 times as fast as the rate of convergence for the BEG approach. Actually, the results in tables 5.1 and 5.2 suggest that the NEK approach as compared to the BEG approach has, on average, at least as many (probably more) times faster convergence for options with four underlying assets as for options with three underlying assets.

5.4 Conclusions

The last decade's increased sophistication of financial instruments has also led to an increased demand for efficient valuation methods. In this paper, a lattice approach (the NEK approach) for the pricing of multivariate contingent claims is developed.

Hitherto, one of the most promising of the lattice approaches for valuation of multivariate contingent claims has been the BEG approach (see [16]). The BEG approach has, however, some problems, e.g., can lead to negative jump probabilities and has slow convergence. These problems are handled in the approach developed in this paper.

A large number of valuations of options with three and four underlying state variables are performed, with both the NEK approach and the BEG approach. It is shown that the NEK approach has a very fast convergence. The results are accurate enough for many needs with as few as 10 time steps, and the results are astonishingly accurate with 20 time steps.

If the results from the NEK approach are compared to the results from the BEG approach, the following conclusion can be made:

The NEK approach is superior to the BEG approach, even if the possibility of negative jump probabilities when using the BEG approach is disregarded. This is due to the superiority of the NEK approach's rate of convergence. A fast rate of convergence is especially important when valuing multivariate contingent claims, since the computational burden increases very fast as more time steps are used.

In the paper, valuations of American options with the help of the NEK approach are also performed. Thus, the NEK approach can easily handle contingent claims with early exercise conditions.

Moreover, the lattice approach developed in this paper is rather easy to implement.

This follows from the fact that all jump probabilities are equal and that the formulas for the jump sizes are *extremely* simple.

Bibliography

[1] Ackoff, R.L., *Scientific Method: Optimizing Applied Research Decisions.* New York, USA: John Wiley & Sons, 1962.

[2] Ames, W.F., *Numerical Methods for Partial Differential Equations.* New York and San Francisco, USA: Academic Press, 1977.

[3] Ananthanarayanan, A.L., and E.S. Schwartz, "Retractable and Extendible Bonds: The Canadian Experience", *The Journal of Finance* 35: pp. 31-47, 1980.

[4] Anton, H., *Elementary linear algebra*, 5:th edition. New York, USA: John Wiley & Sons, 1987.

[5] Bailey, J., "Implementing Fine-grained Scientific Algorithms on the Connection Machine Super Computer", Thinking Machines Technical Report Series TR90-1, Thinking Machines Corporation, Cambridge, Massachusetts, USA, 1990.

[6] Bain, J.S., *Barriers to New Competition*, Cambridge, Massachusetts, USA: Harvard University Press, 1956.

[7] Bartter, B., and R. Rendleman, "The Pricing of Options on Debt Securities", *Journal of Financial and Quantitative Analysis* 15: pp. 11-24, 1980.

[8] Ben-Horim, M., and W.L. Silber, "Financial Innovation: A Linear Programming Approach", *Journal of Banking and Finance* 1: pp. 277-296, 1977.

[9] Bernstein, P.L., *Capital Ideas*, New York, USA: The Free Press, 1992.

[10] Black, D.G., "Success and Failure of Futures Contracts: Theory and Empirical Evidence", Monograph Series in Finance and Economics, 1986-1, Salomon Brothers Center for the Study of Financial Institutions, Graduate School of Business Administration, New York University, 1986.

[11] Black, D.G., and W.L. Silber, "Pioneering Products: Some Empirical Evidence from Futures Markets", Working Paper Series, No. 395, Salomon Brothers Center for the Study of Financial Institutions, Graduate School of Business Administration, New York University, 1986.

[12] Black, F., and M. Scholes, "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81: pp. 637-659, 1973.

[13] Boghosian, B.M., "Computational Physics on the Connection Machine, Massive Parallelism-a New Paradigm", *Computers in Physics* Jan./Feb., 1990.

[14] Boyle, P.P., "Option Valuation using a Three-Jump Process", *International Options Journal* 3: pp. 7-12, 1986.

[15] Boyle, P.P., "A Lattice Framework for Option Pricing with Two State Variables", *Journal of Financial and Quantitative Analysis* 23: pp. 1-12, 1988.

[16] Boyle, P.P., Evnine, E., and S. Gibbs, "Numerical Evaluation of Multivariate Contingent Claims", *The Review of Financial Studies* 2: pp. 241-250, 1989.

[17] Boyle, P.P., and E.S. Schwartz, "Equilibrium Prices of Guarantees Under Equity-Linked Contracts", *Journal of Risk and Insurance* 44: pp. 639-660, 1978.

[18] Brauer, G.A., and R. Ravichandran, "How Sweet is Silver?", *The Journal of Portfolio Management* Vol. 12 No. 4: pp. 33-42, 1986.

[19] Brennan, M.J., and E.S. Schwartz, "Savings Bonds, Retractable Bonds and Callable Bonds", *Journal of Financial Economics* 5: pp. 67-88, 1977.

[20] Brennan, M.J., and E.S. Schwartz, "The Valuation of American Put Options", *The Journal of Finance* 32: pp. 449-462, 1977.

[21] Brennan, M.J., and E.S. Schwartz, "Convertible Bonds: Valuation and Optimal Strategies for Call and Conversion", *The Journal of Finance* 32: pp. 1699-1715, 1977.

[22] Brennan, M.J., and E.S. Schwartz, "Finite Difference Methods and Jump Processes arising in the Pricing of Contingent Claims: A Synthesis", *Journal of Financial and Quantitative Analysis* 13: pp. 461-474, 1978.

[23] Brennan, M.J., and E.S. Schwartz, "Corporate Income Taxes, Valuation, and the Problem of Optimal Capital Structure", *Journal of Business* 51: pp. 103-114, 1978.

[24] Brennan, M.J., and E.S. Schwartz, "Conditional Predictions of Bond Prices and Returns", *Journal of Finance* 35: pp. 405-417, 1979.

[25] Brennan, M.J., and E.S. Schwartz, "Analyzing Convertible Bonds", *Journal of Financial and Quantitative Analysis* 15: pp. 907-929, 1980.

[26] Brennan, M.J., and E.S. Schwartz, "An Equilibrium Model of Bond Pricing and a Test of Market Efficiency", *Journal of Financial and Quantitative Analysis* 17: pp. 301-329, 1982.

[27] Brenner, M., Courtadon, G., and M. Subrahmanyam, "Options on the Spot and Options on Futures", *The Journal of Finance* 60: pp. 1303-1317, 1985.

[28] Budd, N., "The Future of Commodity-Indexed Financing", *Harvard Business Review*, pp. 44-50, (July-Aug) 1983.

[29] Carr, P., "A Note on the Pricing of Commodity-Linked Bonds", *The Journal of Finance* 42: pp. 1071-1076, 1987.

[30] Carr, P., "The Valuation of Sequential Exchange Opportunities", *The Journal of Finance* 43: pp. 1235-1256, 1988.

[31] Carverhill, A., "Numerical Methods" in Stewart Hodges (editor), *Options: Recent Advances in Theory and Practice*: pp. 70-79. Manchester, UK: Manchester University Press, 1990.

[32] Carverhill, A., and N. Webber, "American Options: Theory and Numerical Analysis" in Stewart Hodges (editor), *Options: Recent Advances in Theory and Practice*: pp. 80-94. Manchester, UK: Manchester University Press, 1990.

[33] Chen, A.H., and J.W. Kensinger, "Uncommon Equity", *The Continental Bank: Journal of Applied Corporate Finance* Vol. 5 No. 1: pp. 36-43, 1992.

[34] Chew, D.H., and C.W. Smithson, "The Uses of Hybrid Debt in Managing Corporate Risk", *The Continental Bank: Journal of Applied Corporate Finance* Vol. 4 No. 4: pp. 79-89, 1992.

[35] Chung, K.L., and R.J. Williams, *Introduction to Stochastic Integration*, 2:nd edition. Boston, USA: Birkhäuser, 1990.

[36] Copeland, T.E., and J.F. Weston, *Financial Theory and Corporate Policy*, 3:rd edition. USA: Addison-Wesley Publishing Company, 1988.

[37] Coombs, R., Saviotti, P., and V. Walsh, *Economics and Technological Change*. London, UK: Macmillan Education, 1987.

[38] Courtadon, G., "The Pricing of Options on Default-Free Bonds", *Journal of Financial and Quantitative Analysis* 17: pp. 75-100, 1982.

[39] Courtadon, G., "A More Accurate Finite Difference Approximation for the Valuation of Options", *Journal of Financial and Quantitative Analysis* 17: pp. 697-703, 1982.

[40] Cox, J.C., Ingersoll, J.E., and S.A. Ross, "An Intertemporal General Equilibrium Model of Asset Prices", *Econometrica* 53: pp. 363-384, 1985.

[41] Cox, J.C., Ingersoll, J.E., and S.A. Ross, "A Theory of the Term Structure of Interest Rates", *Econometrica* 53: pp. 385-407, 1985.

[42] Cox, J.C., and S.A. Ross, "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics* 3: pp. 145-166, 1976.

[43] Cox, J.C., Ross, S.A., and M. Rubinstein, "Option Pricing: A Simplified Approach", *Journal of Financial Economics* 7: pp. 229-264, 1979.

[44] Donnely, T.G., "Algorithm 462 Bivariate Normal Distribution [S15]", *Communications of the ACM* 16: p. 638, 1973.

[45] Dufey, G., and I.H. Giddy, "Innovation in the International Financial Markets", *Journal of International Business Studies*: pp. 33-51, Fall 1981.

[46] Eckl, S., Robinson, J.N., and D.C. Thomas, *Financial Engineering. A Handbook of Derivative Products.* Cambridge, Massachusetts, USA: Basil Blackwell, 1991.

[47] "Economic and Financial Indicators 2", *The Economist* May 16th: p. 136, 1992.

[48] "Taming the Derivative Beast", *The Economist* May 23rd: pp. 81-82, 1992.

[49] Eldén, L., and L. Wittmeyer-Koch, *Numerisk analys-en introduktion.* Lund, Sweden: Studentlitteratur, 1987.

[50] Evans, J.O., and R.E. Odeh, "Algorithm AS70. The Percentage Points of the Normal Distribution", *Applied Statistics* 23, pp. 96-97, 1974.

[51] Finnerty, J.D., "Financial Engineering in Corporate Finance: An Overview", *Financial Management* Vol.17 No.4: pp. 14-33, 1988.

[52] Finnerty, J.D., "An Overview of Corporate Securities Innovation", *The Continental Bank: Journal of Applied Corporate Finance* Vol.4 No.4: pp. 23-39, 1992.

[53] Forsythe, G.E., and W.R. Wasow, *Finite-difference Methods for Partial Differential Equations.* New York, London: John Wiley & sons, 1960.

[54] Geske, R., and K. Shastri, "Valuation by Approximation: A Comparison of Alternative Option Valuation Techniques", *Journal of Financial and Quantitative Analysis* 20: pp. 45-71, 1985.

[55] Gill, P.E., Murray, W., and M.H. Wright, *Numerical Linear Algebra and Optimization. Volume 1.* Redwood City, CA, USA: Addison-Wesley Publishing Company, 1991.

[56] Gladstein, M., Merton, R.C., and M. Scholes, "The Returns and Risk of Alternative Call Option Portfolio Investment Strategies", *Journal of Business* 51: pp. 183-242, 1978.

[57] Gladstein, M., Merton, R.C., and M. Scholes, "The Returns and Risk of Alternative Put Option Portfolio Investment Strategies", *Journal of Business* 55: pp. 1-55, 1982.

[58] Glazer, A., "The Advantages of Being First", *American Economic Review* 75: pp. 473-480, 1985.

[59] Haaser, N.B., and J.A. Sullivan, *Real Analysis.* New York, USA: Dover, 1991.

[60] Hadamard, J., *Lectures on Cauchy's Problem in Linear Partial Differential Equations.* New Haven, USA: Yale University Press, 1923.

[61] Harrison, J.M., and D.M. Kreps, "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Financial Economics* 20: pp. 381-408, 1979.

[62] Hawawini, G., "Financial Innovation and Incentive Schemes to Accelerate the Development of Capital Markets: The Case of France", Working Paper Series, No. 357, Salomon Brothers Center for the Study of Financial Institutions, Graduate School of Business Administration, New York University, 1985.

[63] Healy, M.J.R, "Algorithm AS7. Inversion of a Positive Semi-definite Symmetric Matrix", *Applied Statistics* 17, pp. 86-87, 1968.

[64] Herring, R.J., and A.M. Santomero, "Finanssektorns roll i samhällsekonomin" in R.J. Herring, A.M. Santomero, and S. Viotti, *Finanssektorn och välfärden*: pp. 11-90. Stockholm, Sweden: SNS Förlag, 1991.

[65] Hill, I.D., "Algorithm AS66. The Normal Integral", *Applied Statistics* 22, pp. 424-427, 1973.

[66] Hockney, R.W., and C.R. Jesshope, *Parallel Computers 2*, 2:nd edition. Bristol, UK: Adam Hilger, 1988.

[67] Hull, J., *Options, Futures, and other Derivative Securities.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[68] Hull, J., and A. White, "The Use of the Control Variate Technique in Option Pricing", *Journal of Financial and Quantitative Analysis* 23: pp. 237-251, 1988.

[69] Hull, J., and A. White, "Valuing Derivative Securities Using the Explicit Finite Difference Method", *Journal of Financial and Quantitative Analysis* 25: pp. 87-100, 1990.

[70] Hutchinson, J.M., and S.A. Zenios, "Financial Simulations on a Massively Parallel Connection Machine", *International Journal of Supercomputer Applications*, The MIT Press Vol. 5(2), Summer, 1991.

[71] Ingersoll, J., "An Examination of Corporate Call Policies on Convertible Securities", *The Journal of Finance* 32: pp. 463-478, 1977.

[72] Jennergren, L.P., and B. Näslund, "Gimo-fallet en gång till—som om Högsta Domstolen hade kunnat optionsteori", *BETA* 1/89: pp. 100-108, 1989.

[73] Jennergren, L.P., and B. Näslund, "Models for the Valuation of International Convertible Bonds", *Journal of International Financial Management and Accounting* 2:2 & 3: pp. 93-110, 1990.

[74] Jennergren, L.P., and T. Sörensson, "Choice of Model in Convertible Valuation-A Case Study", *Omega* 4: pp. 185-191, 1991.

[75] Johnson, H., "Options on the Maximum or the Minimum of Several Assets", *Journal of Financial and Quantitative Analysis* 22: pp. 277-283, 1987.

[76] Kamien, M.I., and N.L. Schwartz, *Dynamic Optimization. The Calculus of Variations and Optimal Control in Economics and Management.* New York, USA: North-Holland, 1981.

[77] Kamien, M.I., and N.L. Schwartz, *Market Structure and Innovation.* New York, USA: Cambridge University Press, 1982.

[78] Kane, E.J., "Good Intentions and Unintended Evil: The Case against Selective Credit Allocation", *Journal of Money, Credit and Banking* 9: pp. 55-69, 1977.

[79] Kane, E.J., "Accelerating Inflation, Technological Innovation and the Decreasing Effectiveness of Banking Regulation", *The Journal of Finance* 36: pp. 355-367, 1981.

[80] Kane, E.J., "Policy Implications of Structural Changes in Financial Markets", *American Economic Review* 73: pp. 96-100, 1983.

[81] Kane, E.J., "Microeconomic and Macroeconomic Origins of Financial Innovation" in Federal Reserve Bank of St. Louis (editor), *Financial Innovations*: pp. 3-20. Boston, USA: Kluwer-Nijhoff Publishing, 1984.

[82] Kemna, A.G.Z., *Options in Real and Financial Markets.* Ph.D. dissertation, Erasmus Universiteit Rotterdam, 1988.

[83] Kernigan, B.W., and D.M. Ritchie, *The C Programming Language,* 2:nd edition. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

[84] Kimberly, J.R., "Managerial Innovation" in P.C. Nystrom and W.H. Starbuck (editors), *Handbook of Organizational Design, Volume 1*: pp. 84-104. New York, USA: Oxford University Press, 1981.

[85] Kindleberger, C.P., "Bubbles" in J. Eatwell, M. Milgate, and P. Newman (editors), *The New Palgrave. A Dictionary of Economics. Volume 1*: pp. 281-282. London, UK: The Macmillan Press Limited, 1987.

[86] King, R.D., "The Effect of Convertible Bond Equity Values on Dilution and Leverage", *The Accounting Review* 59: pp. 419-431, 1984.

[87] Kyle, A.S., "Market Failures and the Regulation of Financial Markets", Paper presented at Conference on Operation and Regulation of Financial Intermediaries and Financial Markets, Stockholm School of Economics, Stockholm, Sweden, 20-22 August, 1992.

[88] Lindgren, B.W., *Statistical Theory*, 3:rd edition. New York, USA: Macmillan Publishing Co., 1976.

[89] Malliaris, A.G., and W.A. Brock, *Stochastic Methods in Economics and Finance.* Amsterdam, Holland: North-Holland, 1982.

[90] Marchuk, G.I., "Splitting and Alternating Direction Methods" in P.G. Ciarlet and J.L. Lions (editors), *Handbook of Numerical Analysis, Volume I*: pp. 197-462. Amsterdam, Holland: North Holland, 1990.

[91] Margrabe, W., "The Value of an Option to Exchange one Asset for Another", *The Journal of Finance* 33: pp. 177-186, 1978.

[92] Marki, S.R., *Derivative Financial Products.* USA: Harper Business (A division of Harper Collins Publishers), 1991.

[93] Marr, W., Trimble, J.L., and R. Varma, "Innovation in Global Financing: The Case of Euroequity Offerings", *The Continental Bank: Journal of Applied Corporate Finance* Vol. 5 No. 1: pp. 50-54, 1992.

[94] Mason, S.P., and R.C. Merton, "The Role of Contingent Claims Analysis in Corporate Finance" in E. I. Altman and M. G. Subrahmanyam (editors), *Recent Advances in Corporate Finance*: pp. 7-40. Homewood, USA: Irwin, 1985.

[95] McConnel, J.J., and E.S. Schwartz, "LYON Taming", *The Journal of Finance* 61: pp. 561-576, 1986.

[96] McKee, S., and A.R. Mitchell, "Alternating Direction Methods for Parabolic Equations in Two Space Dimensions with a Mixed Derivative", *The Computer Journal*: Vol. 13 No. 1: pp. 81-86, 1970.

[97] Merton, R.C., "Optimum Consumption and Portfolio Rules in a Continuous-Time Model", *Journal of Economic Theory* 3: pp. 373-413, 1971.

[98] Merton, R.C., "An Intertemporal Capital Asset Pricing Model", *Econometrica* 41: pp. 867-887, 1973.

[99] Merton, R.C., "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science* 4: pp. 141-183, 1973.

[100] Merton, R.C., "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *The Journal of Finance* 29: pp. 449-470, 1974.

[101] Merton, R.C., "On the Pricing of Contingent Claims and The Modigliani-Miller Theorem", *Journal of Financial Economics* 5: pp. 241-250, 1977.

[102] Merton, R.C., "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance", *The Geneva Papers on Risk and Insurance* 14: pp. 225-261, 1989.

[103] Merton, R.C., *Continuous-Time Finance.* Cambridge, Massachusetts, USA: Basil Blackwell, 1990.

[104] Merton, R.C., "The Financial System and Economic Performance", *Journal of Financial Services Research*: pp. 263-300, 1990.

[105] Merton, R.C., "Financial Innovation and Economic Performance", *The Continental Bank: Journal of Applied Corporate Finance* Vol. 4 No. 4: pp. 12-22, 1992.

[106] Merton, R.C., "Operation and Regulation in Financial Intermediation: A Functional Perspective", Paper presented at Conference on Operation and Regulation of Financial Intermediaries and Financial Markets, Stockholm School of Economics, Stockholm, Sweden, 20-22 August, 1992.

[107] Miller, M.H., "Financial Innovation: The Last Twenty Years and the Next", *Journal of Financial and Quantitative Analysis* 21: pp. 459-471, 1986.

[108] Miller, M.H., "Financial Innovation: Achievements and Prospects", *The Continental Bank: Journal of Applied Corporate Finance* Vol. 4 No. 4: pp. 4-11, 1992.

[109] Minsky, H.P., "Financial Innovations and Financial Instability: Observations and Theory" in Federal Reserve Bank of St. Louis (editor), *Financial Innovations*: pp. 21-41. Boston, USA: Kluwer-Nijhoff Publishing, 1984.

[110] Norris, F., "New Securities Are a Hard Sell", *The New York Times* May 19th, 1992.

[111] Näslund, B., "Financial Innovations: A Comparison with R&D in Physical Products", Research Paper 6309, EFI, Stockholm School of Economics, 1986.

[112] Näslund, B., "Some Views on the Current State of Financial Theory", Research Paper 6313, EFI, Stockholm School of Economics, 1986.

[113] Näslund, B., "Finansiella Produktegenskaper" in L. Berg and B. Näslund (editors), *Finansiella Innovationer*: pp. 144-153. Stockholm, Sweden: Norstedts, 1988.

[114] Parzen, E., *Stochastic Processes*. San Francisco, USA: Holden-Day, 1962.

[115] Priovolos, T., "Experience with Commodity-Linked Issues" in T. Priovolos and R.C. Duncan (editors), *Commodity Risk Management and Finance*: pp. 11-38, New York, USA: Oxford University Press, 1991.

[116] Press, W.H., Flannery B.P., Teukolsky S.A., and T.W. Vetterling, *Numerical Recipes in C. The Art of Scientific Computing*. New York, USA: Cambridge University Press, 1988.

[117] Rogers, E.M., *Diffusion of Innovations*, 3:rd edition. New York, USA: The Free Press, 1983.

[118] Ross, S.A., "Institutional Markets, Financial Marketing, and Financial Innovation", *The Journal of Finance* 44: pp. 541-556, 1989.

[119] Schervish, M.J., "Algorithm AS195. Multivariate Normal Probabilities with Error Bound", *Applied Statistics* 33, pp. 81-94, 1984.

[120] Schervish, M.J., "Corrections. Algorithm AS195. Multivariate Normal Probabilities with Error Bound", *Applied Statistics* 34, p. 103, 1985.

[121] Schmalensee, R., "Product Differentiation Advantages of Pioneering Brands", *American Economic Review* 72, pp. 349-365.

[122] Schwartz, E.S., "The Valuation of Warrants: Implementing a New Approach", *Journal of Financial Economics* 4: pp. 79-93, 1977.

[123] Silber, W.L., "The Process of Financial Innovation", *American Economic Review* 73: pp. 89-95, 1983.

[124] Smith, C.W., "Option Pricing: A Review", *Journal of Financial Economics* 3: pp. 3-51, 1976.

[125] Smith, C.W., Smithson, C.W., and D.S. Wilford, *Managing Financial Risk*. New York, USA: Harper & Row, 1989.

[126] Smith, G.D., *Numerical Solution of Partial Differential Equations*. Oxford, UK: Oxford University Press, 1965.

[127] Stewart, G.W., *Introduction to Matrix Computations*. Orlando, Florida, USA: Academic Press, Inc., 1973.

[128] Stoneman, P., *The Economic Analysis of Technological Change*. Oxford, UK: Oxford University Press, 1983.

[129] Stulz, R.M., "Options on the Minimum or the Maximum of Two Risky Assets", *Journal of Financial Economics* 10: pp. 161-185, 1982.

[130] Sundaresan, S., "Valuation of Swaps", Unpublished manuscript, Columbia University, 1989.

[131] Thomée, V., "Finite Difference Methods for Linear Parabolic Equations" in P.G. Ciarlet and J.L. Lions (editors), *Handbook of Numerical Analysis, Volume I*: pp. 5-196. Amsterdam, Holland: North Holland, 1990.

[132] Tirole, J., "Theories of Speculation" in S. Bhattacharya and G.M. Constantinides (editors), *Financial Markets and Incomplete Information. Frontiers of Modern Financial Theory, Volume 2*: pp. 351-361. Savage, Maryland, USA: Rowman & Littlefield, 1989.

[133] Van Horne, J.C., "Of Financial Innovations and Excesses", *The Journal of Finance* 40: pp. 621-631, 1985.

[134] Viotti, S., "Den svenska finanssektorns utveckling—ett samhällsekonomiskt perspektiv" in R.J. Herring, A.M. Santomero, and S. Viotti, *Finanssektorn och välfärden*: pp. 93-116. Stockholm, Sweden: SNS Förlag, 1991.

[135] Vasicek, O., "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics* 5: pp. 177-188, 1977.

[136] Wamsley, J., *The New Financial Instruments.* New York, USA: John Wiley & Sons, 1988.

[137] Worzel, K.J., and S.A. Zenios, "Parallel- and Super-Computing in The Financial Services Industry", The Wharton School of the University of Pennsylvania, Unpublished manuscript, 1991.

[138] Young, D., "The Numerical Solution of Elliptic and Parabolic Partial Differential Equations" in John Todd (editor), *Survey of Numerical Analysis*: pp: 380-438. New York, USA: McGraw-Hill, 1962.

[139] Zenios, S.A., "Parallel- and Super-Computing in The Practice of Management Science", The Wharton School of the University of Pennsylvania, Unpublished manuscript, 1991.

[140] Zenios, S.A., "Massively Parallel Computations for Financial Planning Under Uncertainty" in J. Mesirov (editor), *Very Large Scale Computing in the 21-st Century.* Philadelphia, USA: SIAM, 1991.

[141] Zenios, S.A., "Parallel computing: The introduction of novel computer architectures greatly alters technical landscape", *OR/MS TODAY*: pp. 44-49, August 1992.

EFI - reports since 1987

Published in the language indicated by the title

Andersson, T., Ternström, B., Private Foreign Investments and Welfare Effects: A comparative study of five countries in southeast Asia. Research report

Andersson, T., Ternström, B., External Capital and Social Welfare in South-East Asia. Research report

Benndorf, H., Marknadsföringsplanering och samordning mellan företag i industriella system. EFI/MTC

Bergman, L., Mäler, K-G., Ståhl, I., Överlåtelsebara utsläppsrätter. En studie av kolväteutsläpp i Göteborg. Research report

Björkegren, D., Mot en kognitiv organisationsteori. Research report

Claesson, K., Effektiviteten på Stockholms Fondbörs

Davidsson, P. Growth Willingness in Small Firms. Entrepreneurship - and after? Research report

Engshagen, I., Finansiella nyckeltal för koncern versus koncernbolag. En studie om finansiella måltal. Research report

Fredriksson, O., Holmlöv, PG., Julander, C-R., Distribution av varor och tjänster i informationssamhället.

Hagstedt, P., Sponsring - mer än marknadsföring. EFI/MTC

Järnhäll, B., On the Formulation and Estimation of Models of Open Economies. Research report

Kylén, B., Digitalkartans ekonomi. Samhällsekonomiska modeller för strategiska val. Delrapport 2 i "Nordisk Kvantif"

Lundgren, S., Elpriser: Principer och praktik. Research report

Schwarz, B., Hederstierna, A., Socialbidragens utveckling - orsaker eller samvariationer? Research report

Swartz, E., Begreppet federativ organisation belyst i ett organisationsteoretiskt, juridiskt och empiriskt perpektiv. Research report.

Westlund, A.H. Öhlén, S., Business Cycle Forecasting in Sweden; A problem analysis. Research report

1988

Andréasson, I-M., Costs of Controls on Farmers' Use of Nitrogen. A study applied to Gotland.

Björkegren, D., Från sluten till öppen kunskapsproduktion. Delrapport 3 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björkegren, D., Från tavistock till Human-relation. Delrapport 4 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björklund, L., Internationell projektförsäljning. En studie av ad hoc-samverkan mellan företag vid internationella projekt. Research report

Bojö, J., Mäler, K-G., Unemo, L., Economic Analysis of Environmental Consequences of Development Projects. Research report

Brynell, K., Davidsson, P., Det nya småföretagandet? - en empirisk jämförelse mellan små high-tech företag och konventionella småföretag. Research report

Dahlgren, G., Witt, P., Ledning av fusionsförlopp. En analys av bildandet av Ericsson Information Systems AB.

Forsell, A., Från traditionell till modern sparbank. Idé organisation och verksamhet i omvandling. Research report

Hirdman, V., The Stability and the Interest Sensitivity of Swedish Short Term Capital Flows. Research report

Hultén, S., Vad bestämmer de svenska exportmarknadsandelarnas utveckling?

Häckner, J., Biobränslenas konkurrenskraft i ett framtida perspektiv. Research report

Jennergren, P., Näslund, B., If the Supreme Court had Known Option Theory: A Reconsideration of the Gimo Case. Research report

Jennergren, P., Näslund, B., The Gimo Corporation Revisited: An Exercise in Finance Theory. Research report

Jennergren, P., Näslund, B., Valuation of Debt and Equity Through One- and Two-State Contingent Claims Models - with an Application to a Swedish Court Case. Research report

Jonung, L., Laidler, D., Are Perceptions of Inflation Rational? Some Evidence for Sweden. Research report.

Lagerstam, C., Business Intelligence. Teori samt empirisk studie av elektronik- och bioteknikbranschen i Japan. Research report

Liljegren, G., Interdependens och dynamik i långsiktiga kundrelationer. Industriell försäljning i nätverksperspektiv.

Näslund, B., Några synpunkter på börsfallet i oktober 1987. Research report

Olsson, C., The Cost-Effectiveness of Different Strategies Aimed at Reducing the Amount of Sulphur Deposition in Europe. Research report

Philips, Å., Eldsjälar. En studie av aktörsskap i arbetsorganisatoriskt utvecklingsarbete.

Sellstedt, B., Produktionsstrategier. En diskussion med utgångspunkt från litteraturen. Research report.

Skogsvik, K., Prognos av finansiell kris med redovisningsmått. En jämförelse mellan traditionell och inflationsjusterad redovisning.

Wahlund, R., Skatteomläggningen 1983 - 1985. En studie av några konsekvenser. Research report

Wahlund, R., Varför och hur olika svenska hushåll sparar. Research report

Vredin, A., Macroeconomic Policies and the Balance of Payments.

Åkerman, J., Economic Valuation of Risk Reduction: The Case of In-Door Radon. Research report.

1989

Andersson, T., Foreign Direct Investment in Competing Host Countries. A Study of Taxation and Nationalization.

Björkegren, D., Skönhetens uppfinnnare. Research report

Björkegren, D., Hur organisationer lär. Studentlitteratur.

Blomström, M., Transnational Corporations and Manufactoring Exports from Developing Countries. Research report.

Carlsson, A., Estimates of the Costs of Emission control in the Swedish Energy Sector. Research report

Blomström, M., Foreign Investment and Spillovers. London; Routledge

Bordo, M.D., Jonung, L., The Long-Run Behavior of Velocity: The Institutional Approach Revisted. Research report

Davidsson, P., Continued Entrepreneurship and Small Firm Growth.

DeJuan, A., Fiscal Attitudes and Behavior. A study of 16 - 35 years old Swedish citizens. Research report

Edlund, P-O., Preliminary Estimation of Transfer Function Weights. A Two-Step Regression Approach.

Fridman, B., Östman, L., eds. Accounting Development - some perspectives. In honour of Sven-Erik Johansson.

Gadde, L-E., Håkansson, H., Öberg, M., Stability and Change in Automobile Distribution. Research report

Glader, M., Datorer i småföretag. Teldok

Jakobsson, B., Konsten att reagera. Intressen, institutioner och näringspolitik. Carlssons Bokförlag

Jonung, L., The Economics of Private Money. Private Bank Notes in Sweden 1831 - 1902. Research report

Jonung, L., Batchelor, R.A., Confidence about Inflation Forecasts: Tests of Variance Rationality. Research report.

Kylén, B., Hur företagschefer beslutar innan de blir överraskade. Ett försök till förklaring av svarsmönster i svagsignalsitutationer.

Lagerstam, C., Jämförelse av skydd med samlad valuta/aktieoption och skydd med separata aktie- och valutaoptioner samt härledning av optionspriser.
Research report

Lagerstam, C., On the Pricing and Valuation of Forwards and Options on Futures and their Relationsship. Research report.

Larsson, B., Koncernföretaget. Ägarorganisationer eller organisation för ägare?

Lindberg, C., Holmlöv, PG., Wärneryd K-E., Telefaxen och användarna. Teldok

Löwstedt, J., Föreställningar, ny teknik och förändring. Tre organisationsprocesser ur ett kognitivt aktörsperspektiv. Doxa Förlag

Löwstedt, J., (red) Organisation och teknikförändring. Studentlitteratur.

Tollgerdt-Andersson, I., Ledarskapsteorier, företagsklimat och bedömningsmetoder.

Schuster, W., Ägandeformens betydelse för ett företag - en studie av ICA-rörelsen.

Schwartz, B., Företaget som medborgare. Samhällskontakter och reklam som legitimeringsinstrument. Research report

Spets, G., Vägen till nej. Anade och oanade konsekvenser av en OS-satsning. Research report.

Stymne, B., Information Technology and Competence Formation in the Swedish Service Sector. IMIT/EFI

1990

Björkegren, D., Litteraturproduktion - en fallstudie. Delrapport 2 i forskningsprojektet Det skapande företaget: Management of Narrativation.
Research Report.

Bojö, J., Economic Analysis of Agricultural Development Projects. A Case Study from Lesotho. Research Report.

Brunsson, N., Forsell, A., Winberg, H., Reform som tradition. Administrativa reformer i Statens Järnvägar.

Drottz-Sjöberg, B-M., Interests in Humanities, Social Science and Natural Science. A Study of High School Students.

Eklöf, J.A., Macro Ecoonomic PLanning with Quantitative Techniques - and Effects of Varying Data Quality. Research Report.

Flink, T., Hultén, S., Det svenska snabbtågsprojektet - de första 20 åren. Research Report.

Hesselman, A., Ett lokalt ekonomisystem. En fallstudie inom byggbranschen. Research Report.

Jennergren, L.P., Näslund, B., Models for the Valuation of International Convertible Bonds. Research Report.

Jonung, L., Introduction and Summary to The Stockholm School of Economics Revisited. Research Report.

Lagerstam, C., Hedging of Contracts, Anticipated Positions, and Tender Offers.

Lindkvist, H., Kapitalemigration

Normark, P., Strategiska förändringar i kooperationer fallet Lantmännen.
Research Report.

Patrickson, A., Essays in the Latin American Fertilizer Industry.

Sjöstrand, S-E., Den dubbla rationaliteten. Research Report.

Steiner, L., Ledningsfunktionen i tillväxtföretag. Ledningsteamens sammansättning och funktion i tillverkande företag inom informationsteknologiindustrin.

Warne, A., Vector Autoregressions and Common Trends in Macro and Financial Economics

Wärneryd, K., Economic Conventions. Essays in Institutional Evolution.

1991

Bergholm, M., Hemdistribution med hjälp av datorkommunikation. Research Report.

Bojö, J., The Economics of Land Degradation. Theory and Applications to Lesotho.

Brytting, T., Organizing in the Small Growing Firm - A grounded theory approach

Edlund, P-O., Soegaard, H., Business Cycle Forecasting: Tracking Time - Varying Transfer Functions. Research Report.

Ericson, M., Iggesundsaffären - Rationaliteter i en strategisk förvärvsprocess

Horn av Rantzien, M., Endogenous Fertility and Old-Age Security. Research Report.

Jennergren, L.P., Näslund, B., Options with Stochastic Lives. Research Report.

Jonung, L., Gunnarsson, E., Economics the Swedish Way 1889-1989. Research Report.

Jungenfelt, K., An Analysis of Pay as Yo Go Pension Systems as Dynastic Clubs. Research Report.

Lundgren, A., Technological Innovation and Industrial Evolution - The Emergence of Industrial Networks.

Nilsson, A. G., Anskaffning av standardsystem för att utveckla verksamheter. Utveckling och prövning av SIV-metoden.

Nilsson, J., Ekonomisk styrning i ett multinationellt företag. Research Report.

Normark, P., Swartz, E., Bolagisering av ekonomiska föreningar. Research Report.

Rutström, E., The Political Economy of Protectionism in Indonesia A Computable General Equilibrium Analysis.

Sjöstrand, S-E., Institution as Infrastructures of Human Interaction. Research Report.

Söderqvist, T., Measuring the Value of Reduced Health Risks: The Hedonic Price Technique Applied on the Case of Radon Radiation. Research Report.

Wahlund, R., Skatter och ekonomiska beteenden. En studie i ekonomisk psykologi om främst skattefusk och sparande utifrån 1982 års skatteomläggning.

Wahlund, R., Studenternas betalningsvilja för studier vid Handelshögskolan i Stockholm. Research Report.

Westelius, A., Westelius, A-S., Decentraliserade informationssystem. Två fallstudier inom ekonomistyrning. Research Report.

Westlund, K., Affärsetik. En studie av etiska bedömningar. Research Report.

Wirsäll, N-E., Julander, C-R., Den lättrörliga detaljhandeln.

1992

Charpentier, C., Ekonomisk styrning av statliga affärsverk.

Edlund, P-O., Karlsson, S., Forecasting the Swedish Unempolyment Rate: VAR vs. Transfer Function Modelling.

Eklöf, J., Varying Data Quality and Effects in Economic Analysis and Planning.

Eliasson, M., Julander, C-R., Productivity in Swedish Grocery Retailing. - changes over time and a causal model.

Ewing, P., Ekonomisk styrning av enheter med inbördes verksamhetssamband.

Fredriksson, O., Datorkommunikation i Distributionssystem.

Erfarenheter och effekter vid införandet av två multilaterala interorganisatoriska informationssystem - exemplet BASCET Infolink AB.

Fredriksson, T., Policies Towards Small and Medium Enterprises in Japan and Sweden. Research Report.

Gerdtham, U., Jönsson, B., Sjukvårdskostnader i de nordiska länderna. Research Report.

Hedvall, M., Produktkvalitet som konkurrensmedel i producentvaru-producerande industri. Research Report.

Holmberg, C., Effects of Feature Advertising and Other Promotions. Research Report.

Jansson, D., Spelet kring investeringskalkyler. Norstedts

Jennergren, L.P., Näslund, B., Valuation of Executive Stock Options. Research Report.

Ljung, A., Intressentstrategier - En longitudinell studie av utvecklingen i två svenska företag.

Kokko, A., Foreign Direct Investment, Host Country Characteristics and Spillovers.

Mårtensson, P., Mähring, M., Information Support From Staff to Executives. - An Explorative Study. Research Report

Paalzow, A., Public Debt Management. Research Report.

Sjöstrand, S-E., On Institutions and Institutional Change. Research Report.

Södergren, B., Decentralisering. Förändring i företag och arbetsliv.

Rabiee, M., A Review of Demand for Telecommunication: An Analytical Approach. Research Report.

Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1991. Research Report.

Tollgerdt-Andersson, I., Sjöberg, L., Intresse och kreativitet inom tjänsteproducerande företag. Research Report.

Wahl, A., Kvinnliga civilekonomers och civilingenjörers karriärutveckling.

Persson, P-G, Basket Analysis. A New Way of Studying Short Term Effects of Promotions in Grocery Retailing.