Five Contributions
to Econometric Theory and
the Econometrics
of Ultra-High-Frequency Data
EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization
The research activities at the Institute are organized into 22 Research Centres. Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre: Centre Director:
Management and Organisation (A) Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E) Carin Holmquist
Public Management (F) Nils Brunsson
Information Management (I) Mats Lundeborg
Centre for People and Organization (PMO) Andreas Werr (acting)
Centre for Innovation and Operations Management (T) Christer Karlsson
Centre for Risk Research (CFR) Lennart Sjöberg
Economic Psychology (P) Guje Sevón
Centre for Consumer Marketing (CCM) Magnus Söderlund
Centre for Information and Communication Research (CIC) Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D) Björn Axelsson
Centre for Strategy and Competitiveness (CSC) Örjan Sölvell
Centre for Business and Economic History (BEH) Håkan Lindgren
Accounting and Managerial Finance (B) Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC) Kenth Skogsvik
Finance (FI) Clas Bergström
Centre for Health Economics (CHE) Bengt Jönsson
International Economics and Geography (IEG) Mats Lundahl
Economics (S) Lars Bergman
Economic Statistics (ES) Anders Westlund
Law (RV) Erik Nerep
Centre for Tax Law (SR) Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address
EFI, Box 6501, SE–113 83 Stockholm, Sweden  •  Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00  •  Fax: +46(0)8-31 62 70  •  E-mail efi@hhs.se
Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data

Mika Meitz
To Everyone
Contents

Acknowledgements xi

Introduction 1

ECONOMETRICS OF ULTRA-HIGH-FREQUENCY DATA 9

Evaluating models of autoregressive conditional duration 11
1 Introduction ............................................ 14
2 Testing ACD models against general additive and multiplicative alternatives 15
3 Applications of the theory ............................ 20
4 Smooth transition ACD models .......................... 24
5 Time-varying ACD models ............................ 28
6 Simulation experiment .................................. 31
7 Application to Trades and Quotes series ................. 37
8 Conclusions ............................................ 41
 Tables .................................................. 42
 Appendix A ............................................ 47
 Appendix B ............................................ 49
 Appendix C ............................................ 50
 References ............................................. 51

On the econometrics of multivariate marked ultra-high-frequency data 55
1 Introduction ............................................ 58
2 Marked point processes ................................. 61
3 Modelling the ground process ............................ 62
4 Modelling the 'type' of trade ............................. 65
5 Modelling the returns .................................. 66
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Strict stationarity</td>
<td>69</td>
</tr>
<tr>
<td>7</td>
<td>Maximum likelihood estimation</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>Preliminary notes on the data</td>
<td>72</td>
</tr>
<tr>
<td>9</td>
<td>Conclusions and directions for future work</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td><strong>STABILITY OF TIME SERIES MODELS</strong></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models</td>
<td>83</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>86</td>
</tr>
<tr>
<td>2</td>
<td>General Markov Models</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>GARCH and ACD models</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>Conclusion</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Tables</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>A necessary and sufficient condition for the strict stationarity of a family of GARCH processes</td>
<td>115</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>118</td>
</tr>
<tr>
<td>2</td>
<td>Main result</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>Stability of nonlinear AR–GARCH models</td>
<td>123</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>126</td>
</tr>
<tr>
<td>2</td>
<td>Model and Assumptions</td>
<td>127</td>
</tr>
<tr>
<td>3</td>
<td>Geometric Ergodicity</td>
<td>132</td>
</tr>
<tr>
<td>4</td>
<td>Examples</td>
<td>135</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>149</td>
</tr>
</tbody>
</table>
Acknowledgements

Like any author, I also owe my debts, professional and personal. It is customary, and also in this case appropriate, to begin with my academic mentor. I am indebted to my supervisor, Timo Teräsvirta, first and foremost for believing in me. You have given me the opportunity to explore my academic interests freely and encouraged me even when I have embarked on projects that seem insurmountable. Without these opportunities, and without your ever-patient and careful guidance, I would never have learned so much and this thesis would now look very different. For this, I express my deepest gratitude to you.

I am in great debt also to Pentti Saikkonen, with whom I have written two of the papers in this thesis. Our journeys in the land of Markov chains have truly been a pleasure, and I could never have reached the milestones of these journeys without your continuing help and support.

One of the papers in this thesis is written together with Luc Bauwens. I thank you for your kind support and guidance, and for the opportunity to work with you on this sometimes daunting project during which I have learned a lot.

Special thanks are also due to Annastiina Silvennoinen, with whom I have shared several moments of these academic adventures. Your comments, questions, and critique have had a profound impact on my thoughts and work. For this, and for your unwithering friendship, I am forever grateful to you.

Several people have made my life much more enjoyable during these years of Serious Academic Work, and I would like to take this opportunity to thank Cristina Amado, Sune Karlsson, Tomoaki Nakatani, Rickard Sandberg, Birgit Strikholm, David Veredas, and Sophie von Pöffendahl for their most pleasant company in Serious Non-Academic Activities.

During the writing of this thesis I have had the good fortune to spend long periods visiting the Center for Operations Research and Econometrics at Université catholique de Louvain and the School of Finance and Economics at University of Technology Sydney. These visits have given me the opportunity to meet a wide range of econometricians from around the world, and have greatly broadened my academic perspective. Furthermore, without Luc Bauwens and the financial support from the Marie Curie Research Training Network “Microstructure of Financial Markets in Europe” my knowledge of Belgian beers would not be what it is today, whereas Tony Hall
and Carl Silfvén’s Foundation gave me the possibility to dive with dolphins, sharks, and turtles during my (short) absences from my office desk.

All these years of research would not have been possible without financial sponsors, and I acknowledge the support from Jan Wallander’s and Tom Hedelius’ Foundation.

Finally, the copyrights of some of the material in this thesis are held by the American Statistical Association and Cambridge University Press, and I thank them for their permission to reproduce it in this thesis.

Stockholm, March 2006

Mika Meitz
Introduction

This thesis consists of five research papers from two areas of econometrics that are rather different. The first two papers concern the econometrics of ultra-high-frequency data. This research area was born about a decade ago when a new type of data — more complete and precise than the earlier ones — became available from the financial markets. The three last papers of the thesis deal with the stability of time series models. The questions investigated in this area are some of the most fundamental properties underlying the theory of time series analysis. In what follows, we try to give brief and intuitive introductions to these two research areas and short descriptions of the topics of the five papers. For more detailed and technical accounts of the contents of the papers, as well as for further literature references, the reader is referred to the introductions of the individual papers.

Econometrics of ultra-high-frequency data

Electronic and automated trading at stock exchanges has made available precise, accurate, and complete datasets that describe the trading process at the exchanges. All transactions that take place are recorded with precise time-stamps as well as with information such as the price and volume associated with the trade. The availability of such data has created new challenges for the econometrics as a science since new methods to investigate the recorded information need to be created.

The term ultra-high-frequency data was coined by Engle (2000) to describe data in which information about financial events is recorded at the highest possible frequency. In practice, this means recording every single transaction made on a financial asset. In order to better understand the challenges in modelling such data, let us give a brief description of what the data may consist of. For concreteness, consider the trades on a single stock traded at an exchange. Over the course of one trading day, there may be up to thousands of transactions on this stock. For each of these transactions, a range of information is obtained. First of all, the exact time of the trade is recorded at the precision seconds, or even milliseconds. Knowledge of the exact timings of financial events with respect to each other is of crucial importance for the investigation of causal relationships between the events. Another critical information concerning the transaction is the price with which the trade was made. Alternatively, one may consider what was the return associated with the trade. An important aspect of
returns is their volatility, that is, how much variation is there in the returns at a given point in time. Because volatility is inherently related to financial risk, there is considerable interest in being able to model and predict it. Other characteristics of a single trade that may be recorded include the volume of the trade (how many stocks were involved in the transaction), whether the trade was initiated by a buyer or a seller, as well as various information concerning the quoted bid and ask prices at the time of the trade.

An inherent feature of ultra-high-frequency data is that the times of the events are randomly distributed in time and, hence, irregularly spaced. For this reason, many of the tools used in traditional time series econometrics are rendered inapplicable since they are typically based on the assumption that events occur at deterministic and regularly spaced points in time, say, once every day. As a consequence, a new branch of financial econometrics has emerged where data of this type are being investigated and modelled. The pioneering articles in this field are Engle and Russell (1998) and Engle (2000), and subsequently much effort has been put into this area of research by the econometric community.

Evaluating models of autoregressive conditional duration

With ultra-high-frequency data, the econometricians were faced with a type of data they had not modelled before. For this reason, the first models for ultra-high-frequency data necessarily concentrated only on some simple aspects of the data. The fact that the events are irregularly spaced in time created the first problem of how to model the timings of the events. In their landmark article, Engle and Russell (1998) considered how to model the time spans between two subsequent financial events. They proposed a class of models for these time spans, or durations, called the Autoregressive Conditional Duration (ACD) model. In the ACD model the durations are modelled in a way analogous to the one in which financial returns are modelled in the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986).

The ACD class of models contains various generalizations of the original model put forth in Engle and Russell (1998). In all of these models a critical aspect is the functional form chosen to incorporate past information into the model. The econometrician using the ACD models to ultra-high-frequency data needs to be able to evaluate how well each model fits the data at hand, and make a choice between the different functional forms. In this paper we present a framework within which the statistical fit of an ACD model can be evaluated. This evaluation is done via several misspecification tests that investigate the adequacy of the chosen model.

In econometric terms, in this paper we present a variety of Lagrange multiplier tests designed to detect misspecification of the functional form chosen for the model. These tests include ones against remaining ACD effects that could be modelled and ACD models of higher order, as well as tests of linearity and parameter constancy.

---

1This paper is joint work with Timo Teräsvirta. It has appeared in the Journal of Business and Economic Statistics 24, 104–124, 2006, and is reprinted in this thesis with permission from the American Statistical Association.
Introduction

The test statistics do not require computer-intensive calculations and their asymptotic distributions under the null hypothesis are standard, which makes their application straightforward.

In this paper we also propose two new nonlinear ACD models. Engle and Russell (1998) pointed out the need for certain types of nonlinear generalizations to their model. One of the models we propose, the Smooth Transition ACD model, addresses this need. The other model we propose, the Time Varying ACD model, is motivated by the need to investigate possible deterministic dependencies of the model parameters on time. These two new models complement the range of ACD models available in the literature.

On the econometrics of multivariate marked ultra-high-frequency data

The ACD models considered in the first paper were one of the first models used for ultra-high-frequency data. For simplicity, these models concentrated only on how to model the timings of the trades, and also considered only one stock at a time. Such a model gives a rather simplified view of the trading process at an exchange. If one wishes to provide a more complete description of this process, at least two obvious directions for extensions come into mind. First of all, it can be argued that the trading behaviour of one of the stocks may affect that of the other stocks. A more accurate description of the trading process would therefore take into account the interdependencies between the different stocks. In this sense, it is natural to consider a multivariate extension of the model. Another possible extension is to also consider some other information associated with the trade than just the timing. It would be of interest to incorporate the prices, or the related returns, into the model. This would make it possible to investigate how the timings of the trades affect the returns or their volatility.

These two extensions pose some new challenges. Let us consider the extension to the case of two stocks and their trading times. For both stocks, the trading times occur at random points in time. It may occur that over a given period of time, one of the stocks is trading actively while there are no trades for the other. As both trading and non-trading may relate to information in the market, a model should be able to take into account both the activity and the inactivity at any point in time. In other words, it is not sufficient to consider only points in time when a trade actually occurs; also the periods of non-trading need to be taken into account. This suggests that a truly continuous time approach would be an appropriate one for the trading times in the multivariate case. An approach that may be taken is to directly model the probabilities of a trade occurring at any given time point. In econometric terms, we may model the trading intensities of the two stocks. Such an approach has been used in Bauwens and Hautsch (2005) and Bowsher (2005).

Now consider the other extension, in which both the timing of the trades as well as the associated returns would be modelled. An immediate difficulty that arises is

---

\(^2\)This paper is joint work with Luc Bauwens.
the precise meaning of a return, which should be a measure of price changes. In the present context, trades occur at random and irregular times, and the price changes correspond to different durations of time. One return may correspond to a time period of five seconds, whereas another to a period of two minutes. Clearly these two returns are not directly comparable, and it would be more appropriate to consider measures of price change that would be normalized for the elapsed time durations. Such returns per time unit, and modelling them, are considered in Engle (2000).

In this paper we propose how to make both of these two extensions within one model. We consider a framework flexible enough so that several stocks may be considered simultaneously and that allows the modelling of both the timings of the trades as well as the associated returns. In technical terms, our model is one for marked point processes. For the trading times, we formulate a model using the associated multivariate intensities as in Bauwens and Hautsch (2005) and Bowsher (2005). To include information of the prices to the model, we follow the spirit of Engle (2000) in that returns per time unit are the object to be modelled, and propose a simple multivariate extension of his model.

Stability of time series models

In the second part of this thesis we consider one of the most fundamental properties of time series models, their stability. This is a topic of somewhat technical nature, but we attempt to give a brief and intuitive description of the questions considered in this area. For this purpose we consider, as an example, one of the simplest time series models, the autoregressive model of order one. To facilitate discussion, we resort to the language of mathematics and present this model as the only equation in this Introduction:

\[ X_t = \alpha X_{t-1} + \varepsilon_t. \]  

Here all the variables are thought of taking values in the set of real numbers. The time index \( t \) goes through the values 1, 2, \ldots, the process is initialized from a starting value \( X_0 \), and the error term \( \varepsilon_t \) is assumed to be independent across time, identically distributed for each \( t \), and also independent of \( X_s, 0 \leq s < t \). For concreteness, we may assume that \( \varepsilon_t \) follows the Gaussian distribution.

Let us consider the stability of this simple time series model. It is well known that there exists a unique strictly stationary process \( X_t \) satisfying equation (1) if, and only if, the parameter \( \alpha \) is strictly less than one in absolute value. This condition also ensures a form of stability called ergodicity: Regardless of the value of \( X_0 \), as \( t \) approaches infinity, the probability distribution of \( X_t \) approaches the probability distribution of the unique strictly stationary solution to (1). Therefore, the sense in which we use the word stability is the one of convergence of probability measures. Establishing conditions under which a time series model, possibly with a rather complicated structure, is stable, is the object of interest in this second part of the thesis.

Why is it then of interest to examine when a time series model is stable? We try to answer this question in reverse order. One ultimate use of a model is to apply it
Introduction

to a real-life problem, estimate it on actual data, and to draw conclusions based on
the estimated model. For our inference to be justified, we usually rely on asymptotic
distribution theory for the estimator of the model parameters. We may, for example,
wish to deduce that the parameter estimator is consistent and asymptotically normally
distributed. To show this, we often rely on limit theorems that describe the asymptotic
behaviour of functions of the underlying process. An often used phrase is that “under
usual regularity conditions” the estimator is consistent and asymptotically normally
distributed. These regularity conditions refer, among other things, to the stability
of the underlying stochastic process, which is often a prerequisite for using the limit
theorems. One motivation for our investigation is to provide easily verifiable, low-level
conditions that would ensure that the process is stable and, therefore, that some of
these vague “usual regularity conditions” would actually be satisfied. This, in turn,
will allow us to be more confident about the conclusions we draw based on our model.

Ergodicity, mixing, and existence of moments of a class of
Markov models with applications to GARCH and ACD models

Time series models such as (1) may be seen as Markov chains. Markovian processes
are such that, given the value of the process today, the next time period’s value is
independent of the history of the process before today. In this paper we consider
a class of Markov chains that consist of two parts, an observable one and a hidden
one. In investigating such a model, our ultimate purpose is to derive conditions under
which the joint process, consisting of both the observable and the hidden parts, would
be stable in the sense of being an ergodic Markov chain. We present an approach in
which this problem can be divided in two parts. First we establish conditions under
which the hidden process is stable. Then we show that under certain situations the
stability property of the hidden part is inherited by the joint process. Our approach is
related to Hidden Markov Models that have been investigated in econometric context
for example in Genon-Catalot, Jeantheau, and Larédo (2000).

We then proceed to apply these results to two classes of models used in financial
econometrics. The first of these is the class of Generalized Autoregressive Condi-
tionally Heteroskedastic (GARCH) models introduced by Engle (1982) and Bollerslev
(1986). The second is the class of Autoregressive Conditional Duration (ACD) models
put forth by Engle and Russell (1998) (and also investigated in the first part of this
thesis). We derive sufficient conditions for stability for a wide variety of models in
these two classes. Contrary to a large body of existing literature, the models are also
allowed to have highly nonlinear structures. Our results strengthen the ones in earlier
literature and, in a number of cases, no previous results have appeared.

---

3This paper is joint work with Pentti Saikkonen.
A necessary and sufficient condition for the strict stationarity of a family of GARCH processes

In the previous paper we established sufficient conditions for stability of some time series models. Even though these results may be of high value, there is at least one downside in them: As the conditions are only sufficient, and not necessary ones, we may be disregarding potentially large regions of the parameter space in which the process actually is stable. In this short note, we consider a particular family of GARCH processes introduced originally in He and Teräsvirta (1999), and establish a necessary and sufficient condition for the existence of a strictly stationary solution. Our results complement and extend the ones given in Nelson (1990), Bougerol and Picard (1992), and Duan (1997).

Stability of nonlinear AR–GARCH models

In this paper we again investigate sufficient conditions for stability for a particular class of models. The class considered is the one of autoregressive models of order $p$ with conditionally heteroskedastic errors that follow a first-order GARCH process. Both the autoregressive part and the GARCH process are allowed to take rather complicated nonlinear forms. Autoregressive and GARCH models are both widely used in econometrics and so is also their combination which we consider in this paper.

To derive the stability results we again formulate the model as a Markov chain and apply their stability theory. The approach taken is similar to the one in the third paper in this thesis, but there are also significant differences. First of all, the approach based on inheritance used in the third paper does not appear to be directly applicable in the present context. Instead, we formulate the model as a nonlinear state space model and rely on the theory available for it. Furthermore, due to the structure of the model class considered, the technical derivations to prove our results are often quite different from the ones in the third paper.

To the best of our knowledge, this paper provides the first practically applicable stability results for this class of models. The results are of importance as they make usual limit theorems available and thus open up the way to the development of rigorous asymptotic estimation theory for these models.

---

4This paper has been accepted for publication in Econometric Theory (Notes and Problems), and is printed in this thesis with permission from the Cambridge University Press.
5This paper is joint work with Pentti Saikkonen.
References


ECONOMETRICS OF ULTRA-HIGH-FREQUENCY DATA
Evaluating models of autoregressive conditional duration
Evaluating models of autoregressive conditional duration

Abstract

This paper contains two novelties. First, a unified framework for testing and evaluating the adequacy of an estimated autoregressive conditional duration (ACD) model is presented. Second, two new classes of ACD models, the smooth transition ACD model and the time-varying ACD model, are introduced and their properties discussed.

A number of new misspecification tests for the ACD class of models are introduced. They are Lagrange multiplier and Lagrange multiplier type tests against general forms of additive and multiplicative misspecification of the conditional mean function. These forms include tests against higher-order models, tests of no remaining ACD in the standardized durations, as well as tests of linearity and parameter constancy. In addition to its generality, the advantage of this testing approach is its ease of application, since all the resulting asymptotic null distributions are standard. The finite sample properties of the tests are investigated by simulation. A general observation is that the tests are well-sized and have good power. Versions of the test statistics robust to deviations from distributional assumptions other than those being explicitly tested are also given.

The smooth transition and time-varying ACD models are introduced, their main properties are examined, and they serve as alternatives in the tests of linearity and parameter constancy. Finally, the tests are applied to ACD models of the IBM stock traded at the New York Stock Exchange.

This paper is joint work with Timo Teräsvirta. It has appeared in the Journal of Business and Economic Statistics 24, 104–124, 2006, and this earlier version is reprinted in this thesis with permission from the American Statistical Association.

1 Introduction

High-frequency financial time series have become widely available during the past decade or so. Records of all transactions and quoted prices, so-called “ultra-high-frequency” data; Engle (2000), are available from many stock exchanges. An inherent feature is that such data are irregularly spaced in time. There are several approaches to tackling this feature of the data. In this paper we follow the line of work originated by Engle and Russell (1998), where the durations between events (trades, quotes, price changes etc.) are the quantities being modeled. These authors proposed a class of models called the Autoregressive Conditional Duration, or ACD, models, where conditional expected durations are modeled in a fashion similar to the way conditional variances are modeled using ARCH and GARCH models of Engle (1982) and Bollerslev (1986).

Following the GARCH literature, a number of extensions to the original linear ACD model by Engle and Russell (1998) have been suggested. These include the logarithmic ACD model of Bauwens and Giot (2000), and the threshold ACD model of Zhang, Russell, and Tsay (2001). The distribution associated with the conditional durations has also been suggested to have several different shapes. Examples include the the exponential and Weibull distributions as in Engle and Russell (1998), and the Burr and generalized gamma distributions suggested by Grammig and Maurer (2000) and Lunde (1999), respectively.

Despite the surge of different models for financial durations the issue of model evaluation using misspecification tests has not yet received as much attention as it should deserve. In time series econometrics, estimated models of the conditional mean, and more recently, ones of conditional variance, are typically subjected to a variety of evaluation tests to determine the adequacy of the specification. For models of conditional duration, some misspecification tests have also been proposed in the literature. These can be divided into two categories: tests of misspecification in the distribution of the error term, and of misspecification in the functional form of the conditional mean duration. Fernandes and Grammig (in press) as well as Bauwens, Giot, Grammig, and Veredas (2004) have suggested tests of the first type, and Hautsch (2002) proposed tests of the second type. Li and Yu (2003) derived a portmanteau test that can be used to evaluate the adequacy of an estimated ACD model. Finally, Hong and Lee (2003) considered a general diagnostic test which can be used as a misspecification test for ACD models.

In this paper, we present a framework for evaluating models of conditional duration based on Lagrange multiplier misspecification tests of the functional form of the conditional mean duration. Our goal is to derive easily applicable tests that can reveal various types of misspecification. We present our results in a general form, from which misspecification tests against specific alternatives are derived in a straightforward fashion. Our tests include ones against higher-order models and remaining ACD effects in the standardized durations, as well as tests of linearity and parameter constancy. In the process of deriving linearity and parameter constancy tests, we propose two new ACD specifications, namely the smooth transition ACD (STACD) model, and the time-varying ACD (TVACD) model.
This paper has similarities with the one by Lundbergh and Teräsvirta (2002) who derived misspecification tests for GARCH models. The two papers share the same goal: to derive easily applicable evaluation tools based on Lagrange multiplier test statistics. The types of model misspecification considered in these papers are similar. Derivations of the test statistics differ in that the error distributions of the ACD and GARCH models are not the same. Furthermore, the present paper contains a discussion of nonlinear alternatives to the standard ACD model.

The rest of the paper is organized as follows. In Section 2 we briefly review previous work on misspecification testing in ACD models and present the general results that form the basis of our misspecification tests. In Section 3 we derive tests against higher-order models and remaining ACD in the standardized durations. Section 4 presents the smooth transition ACD model and deduces a test of linearity, and Section 5 presents the time-varying ACD model and the test of parameter constancy. Section 6 contains the results of a simulation experiment. In Section 7 we estimate and evaluate ACD models using data from the New York Stock Exchange. Finally, Section 8 concludes.

2 Testing ACD models against general additive and multiplicative alternatives

2.1 Previous work on misspecification testing of ACD models

Evaluation of estimated ACD models by misspecification tests has not been commonplace in empirical work. Often the only diagnostic test applied for the purpose has been the Ljung-Box $Q$-statistic applied to the standardized or squared standardized durations. In the latter case the test is commonly called the McLeod-Li test (McLeod and Li, 1983). Nevertheless, as already mentioned, there are some papers proposing misspecification tests for ACD models. Bauwens, Giot, Grammig, and Veredas (2004) as well as Fernandes and Grammig (in press) discussed the testing for distributional misspecification. The former authors evaluated duration models using density forecast evaluation methods of Diebold, Gunther, and Tay (1998). Their method relied on the fact that the sequence of probability integral transforms of the one-step-ahead forecasts of the conditional densities of durations will be distributed as independent and identically distributed uniform $(0,1)$ random variables when the one-step-ahead forecasts of the conditional densities of the durations coincide with the true densities. This is the null hypothesis to be tested. It may be rejected either because the error distribution of the model is misspecified or because the conditional mean is misspecified. The latter alternative is due to the fact that the choice of the conditional mean function affects the one-step-ahead forecasts. Fernandes and Grammig (in press) tested the distribution of the error term by comparing parametric and nonparametric estimates of the density of the standardized durations. Their test explicitly assumes that the conditional mean is correctly specified but again a rejection may also be a consequence of a misspecified conditional mean. In order to obtain more information about the situation, complementing tests of the distributional assumption by tests of
the conditional mean specification is quite important.

The question of testing the functional form of the conditional mean of an ACD model was addressed in Hautsch (2002). He mentioned Lagrange multiplier, conditional moment, and integrated conditional moment (ICM) tests as potential tools for detecting misspecification and focuses on the latter two methods. Conditional moment tests (see e.g. Newey (1985)) are based on the fact that correct specification implies the validity of certain moment conditions. These tests are, however, known to be heavily dependent on the choice of weighting of the moment conditions. They do not require a well-specified alternative and are thus rather general misspecification tests. But then, they are based on a finite number of moment restrictions and cannot therefore be consistent against all possible alternatives. The ICM test (see e.g. Bierens (1990) or de Jong (1996)) employs an infinite amount of moment conditions and is consistent against every deviation from the null hypothesis. A consequence of this property is that the test is not very powerful against any particular alternative, which may be considered a disadvantage. Another drawback of the ICM test is that application requires approximating the asymptotic null distribution of the test statistic by simulation. This makes the use of the test computationally burdensome in the ACD case where the time series in applications can be quite long.

Recently, Li and Yu (2003) derived a portmanteau test for ACD models. Their test is based on the residual autocorrelations of an estimated ACD model in the spirit of the Ljung-Box test. More generally, Hong and Lee (2003) proposed the generalized spectrum based test of Hong (1999) as a general diagnostic test for ACD and many other models. When applied to the standardized durations resulting from the estimation of an ACD model, this test is consistent against any type of pairwise serial dependence left in the standardized durations. We provide a description of this test in Appendix C, and use it as benchmark test in our power simulations in Section 6.

The purpose of this paper is to present a unified framework for evaluating ACD models using Lagrange multiplier (LM) tests. Using LM tests makes misspecification testing easy without sacrificing power. Since the model is only estimated under the null hypothesis, the need for, say, nonlinear ACD models can be investigated without the often burdensome task of actually estimating such a model. We derive general results from which tests against specific alternatives are easily derived. Such alternatives include higher-order models, remaining ACD in the standardized durations, as well as nonlinearity and parameter nonconstancy.

2.2 General theory

Let \( t_i \) be the time at which the \( i \)th event (trade, quote, price change etc.) occurs and denote by \( x_i = t_i - t_{i-1} \) the duration between two consecutive events. Let \( \mathcal{F}_{t_{i-1}} \) be the information set consisting of all information up to and including time \( t_{i-1} \). Following Engle and Russell (1998), the class of exponential autoregressive conditional duration

\[ \text{We thank a referee for bringing this paper into our attention.} \]
Evaluating models of autoregressive conditional duration

(ACD) models is defined as follows:

\[ x_i = \psi_i \varepsilon_i \]  \hspace{1cm} (1)
\[ \psi_i = \psi_i (x_{i-1}, \ldots, x_1; \theta_1) \]  \hspace{1cm} (2)
\[ \varepsilon_i \sim \text{i.i.d. exp}(1) \]  \hspace{1cm} (3)

where \( \psi_i = \psi_i (x_{i-1}, \ldots, x_1; \theta_1) \) is the duration conditional on \( F_{i-1} \).

The types of misspecification of this structure can be divided into two broad categories: the conditional duration is either additively or multiplicatively misspecified. This means that the true process is governed either by

\[ x_i = (\psi_i + \phi_i) \varepsilon_i \]  \hspace{1cm} (4)

(additive misspecification) or

\[ x_i = \psi_i \phi_i \varepsilon_i \]  \hspace{1cm} (5)

(multiplicative misspecification). Both in (4) and (5) the additional component of the conditional duration,

\[ \phi_i = \phi_i (x_{i-1}, \ldots, x_1; \theta_1, \theta_2) \]  \hspace{1cm} (6)

is assumed to be an \( F_{i-1} \)-measurable function that depends on additional parameters \( \theta_2 \).

Let

\[ a_i (\theta_1) = \frac{1}{\psi_i (\theta_1)} \frac{\partial \psi_i (\theta_1)}{\partial \theta_1} \]
\[ b_i (\theta_1, \theta_2) = \frac{1}{\psi_i (\theta_1)} \frac{\partial \phi_i (\theta_1, \theta_2)}{\partial \theta_2} \]
\[ c_i (\theta_1) = \frac{x_i}{\psi_i (\theta_1)} - 1. \]

Furthermore, let the superscript ‘0’ denote values evaluated at the true (under the null hypothesis) parameter values, and the hat ‘\( \hat{\} \)’ values evaluated at the maximum value of the likelihood under the null hypothesis. The following theorem defines the test against a general additive alternative.

**Theorem 1** Consider the model (4), (2), (3), (6), and, in addition to assuming that standard regularity conditions (see Theorem 1 of Engle (2000)) apply, assume that under the null hypothesis \( H_0 : \theta_2 = \theta_2^0 \), the function \( \phi_i \) satisfies \( \phi_i (x_{i-1}, \ldots, x_1; \theta_1, \theta_2^0) \equiv 0. \) Then, under the null hypothesis \( H_0 : \theta_2 = \theta_2^0 \), the LM statistic

\[ LM = \left\{ \sum_{i=1}^{n} \hat{c}_i \hat{b}_i \right\} \left\{ \sum_{i=1}^{n} \hat{b}_i \hat{b}_i' - \left( \sum_{i=1}^{n} \hat{b}_i \hat{a}_i' \right) \left( \sum_{i=1}^{n} \hat{a}_i \hat{a}_i' \right)^{-1} \left( \sum_{i=1}^{n} \hat{a}_i \hat{b}_i' \right) \right\}^{-1} \left\{ \sum_{i=1}^{n} \hat{c}_i \hat{b}_i \right\} \]  \hspace{1cm} (7)

has an asymptotic \( \chi^2 \) distribution with \( \dim \theta_2 \) degrees of freedom.
Proof. See Appendix A. ■

Under the null hypothesis $H_0 : \theta_2 = \theta_2^0$, model (4) reduces to (1)–(3). In practice the LM test in Theorem 1 is most easily carried out using an auxiliary least squares regression on particular transformed variables. This can be done as follows (see for example Engle (1984)):

Procedure 1

(i) Obtain the quasi maximum likelihood estimate of $\theta_1$ under the null hypothesis, and compute

\[ \hat{a}_i' = \frac{1}{\psi_i(\hat{\theta}_1)} \frac{\partial \psi_i(\hat{\theta}_1)}{\partial \theta_1}, \hat{b}_i' = \frac{1}{\psi_i(\theta_1)} \frac{\partial \phi_i(\theta_1, \theta_2)}{\partial \theta_2}, \hat{c}_i = \frac{x_i}{\psi_i(\hat{\theta}_1)} - 1, i = 1, \ldots, n, \]

and $SSR_0 = \sum_{i=1}^{n} \hat{c}_i^2$.

(ii) Regress $\hat{c}_i$ on $\hat{a}_i'$ and $\hat{b}_i'$, $i = 1, \ldots, n$, and compute $SSR_1$.

Then, under the null hypothesis, the test statistic

\[ LM = n (SSR_0 - SSR_1)/SSR_0 \]

has an asymptotic $\chi^2$ distribution with $\text{dim} \theta_2$ degrees of freedom.

There is considerable empirical evidence in the ACD literature against exponentially distributed errors. Engle and Russell (1998) already considered the Weibull distribution, Grammig and Maurer (2000) suggested the Burr distribution, and Lunde (1999) used the generalized gamma distribution. Therefore one might consider basing the QML estimation and the derivation of the asymptotic distributions of the test statistics on these distributions. There is, however, a drawback in this approach. As Gourieroux, Monfort, and Trognon (1984) originally showed, the QML parameter estimators of a correctly specified conditional mean model are consistent if, and only if, the quasi-maximum likelihood is based on a distribution belonging to the linear exponential family, regardless of what the true density is. For a discussion of this, see for example White (1994, pp. 62–70). The exponential distribution does belong to the linear exponential family, while the Weibull, Burr, and generalized gamma distributions do not (except for special cases). Therefore, the QML approach based on the exponential distribution will produce consistent estimators regardless of the true error distribution, while QML based on these other distributions will not unless the distribution used is the true density.

Drost and Werker (2004) pointed out that the use of QML based on the ordinary gamma distribution produces consistent estimators. Furthermore, the score vectors of the gamma distribution (suitably normalized) and the exponential distribution are proportional to each other. Hence the estimators for the parameters of the conditional mean based on these two error distributions are identical.

While the QML approach based on the exponential distribution yields consistent estimators, misspecification of the conditional distribution of the durations may still affect the properties of our LM test statistics. This is because applying the test statistics involves implicit assumptions about this conditional distribution. In particular, it is tacitly assumed that the conditional variance of the durations is correctly specified under the null hypothesis. As our interest lies in the specification of the conditional mean, we do not wish other properties of the conditional distribution to affect the
properties of our test statistics. The results of Wooldridge (1991) are helpful here. Since we are using consistent estimators based on QML, his results allow us to derive ‘robust’ versions of the test statistics such that their asymptotic behaviour is unaffected by possible misspecification of the conditional distribution beyond the conditional mean. Applying these results leads to the procedure given below. They cannot be used, however, when the QML estimators are obtained assuming Weibull, Burr, or generalized gamma distributed errors, because a key requirement in Wooldridge (1991) is the consistency of the estimators.

Procedure 2
(i) Obtain the quasi maximum likelihood estimate of $\theta_1$ under the null hypothesis, and compute $\hat{a}'_i = \frac{1}{\psi_i(\hat{\theta}_1)} \frac{\partial \psi_i(\hat{\theta}_1)}{\partial \theta_1}$, $\hat{b}'_i = \frac{1}{\psi_i(\hat{\theta}_1)} \frac{\partial \psi_i(\hat{\theta}_1, \theta_0^2)}{\partial \theta_2}$ and $\hat{c}_i = \frac{1}{\psi_i(\hat{\theta}_1)} X_i$ for $i = 1, \ldots, n$.
(ii) Regress $\hat{b}'_i$ on $\hat{a}'_i$, $i = 1, \ldots, n$, and save the $(\dim \theta_2 \times 1)$ residual vectors $\hat{r}_i$.
(iii) Regress $1$ on $\hat{c}_i \hat{r}_i$, $i = 1, \ldots, n$, and compute the sum of squared residuals, $SSR$, from this regression.

Then, under the null hypothesis, the test statistic $nR^2 = n - SSR$ has an asymptotic $\chi^2$ distribution with $\dim \theta_2$ degrees of freedom.

For the general multiplicative alternative the asymptotic distribution theory has the following form:

Theorem 2 Consider the model (5), (2), (3), (6), and, in addition to assuming that standard regularity conditions (see Theorem 1 of Engle (2000)) apply, assume that under the null hypothesis $H_0: \theta_2 = \theta_0^2$, the function $\phi_i$ satisfies $\phi_i (x_{i-1}, \ldots, x_1; \theta_1, \theta_0^2) \equiv 1$. Then, under the null hypothesis $H_0: \theta_2 = \theta_0^2$, the LM statistic

$$LM = \left\{ \sum_{i=1}^n \hat{\psi}_i \hat{c}_i \hat{b}_i' \right\} \left\{ \sum_{i=1}^n \psi_i^2 \hat{b}_i \hat{a}_i' - \left( \sum_{i=1}^n \psi_i \hat{b}_i \hat{a}_i' \right) \left( \sum_{i=1}^n \hat{a}_i \hat{a}_i' \right)^{-1} \left( \sum_{i=1}^n \psi_i \hat{a}_i \hat{b}_i' \right) \right\}$$

has an asymptotic $\chi^2$ distribution with $\dim \theta_2$ degrees of freedom.

Under the null hypothesis $H_0: \theta_2 = \theta_0^2$, model (5) reduces to (1)–(3). The procedures for the practical application of the test are almost identical to the ones given in Procedures 1 and 2. The only modification needed is to replace $\hat{b}_i$ by $\hat{\psi}_i \hat{b}_i$ throughout.

It may be mentioned that there is a close connection between the LM tests presented here and the conditional moment tests discussed in Hautsch (2002) in the sense that our tests can be interpreted as particular conditional moment tests. By choosing appropriate moment conditions and weighting functions when deriving a conditional moment test one obtains a test that is asymptotically equivalent to our LM test, the only difference being a different consistent estimator of the information matrix.
3 Applications of the theory

In this section we present three misspecification tests that are applications of our general theory. The first two are tests of the original ACD model of Engle and Russell (1998) and the LOGACD model of Bauwens and Giot (2000) against higher-order alternatives. The third one is a test for remaining ACD in the standardized durations. We shall present the precise alternative under consideration, state the distributional results corresponding to Theorem 1 or 2, and give explicit formulas for the quantities $\hat{a}_i$ and $\hat{b}_i$ needed for applying the testing procedures.

3.1 Testing ACD($m,q$) against higher-order alternatives

Engle and Russell (1998) defined their original ACD($m,q$) model by parameterizing the conditional duration (2) as

\[ \psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \]  

(9)

A natural benchmark and starting point for modeling durations is a low-order ACD($m,q$) model, but then, too low an order is an obvious source of misspecification. An estimated ACD($m,q$) model is tested against higher-order alternatives in the same way as Bollerslev (1986) tested a GARCH($p,q$) model against higher-order alternatives. Consequently, either

\[ x_i = (\psi_i + \varphi_i) \varepsilon_i \]  

(10)

or

\[ x_i = (\psi_i + \varphi_i) \varepsilon_i \]  

(14)

The null hypothesis equals $H_0 : \varphi_i \equiv 0$, i.e. $\alpha_{m+1} = \cdots = \alpha_{m+r} = 0$ in the former and $\beta_{q+1} = \cdots = \beta_{q+r} = 0$ in the latter case. Under the alternative the first model is an
ACD($m+r,q$) model and the second model is an ACD($m,q+r$) model, while under the null both models collapse to an ACD($m,q$) model. The ACD($m,q$) model cannot be tested directly against an ACD($m+r,q+s$) model, $r, s > 0$, using standard techniques because of the identification problem already discussed in Bollerslev (1986).

These higher-order alternatives belong to the additive class of alternatives mentioned above. The following two corollaries of Theorem 1 define the test statistics. The tests are most easily carried out using the auxiliary regression procedures given in the previous section.

**Corollary 1** Consider the model (10)–(13) with $\theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)'$ and $\theta_2 = (\alpha_{m+1}, \ldots, \alpha_{m+r})'$. Under the null hypothesis $H_0 : \theta_2 = 0$, the statistic (7), where

$$\hat{a}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \hat{\psi}_i^{-1} \left( 1, x_{i-1}, \ldots, x_{i-m}, \hat{\psi}_{i-1}, \ldots, \hat{\psi}_{i-q} \right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_1}$$

$$\hat{b}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} \left( x_{i-m-1}, \ldots, x_{i-m-r} \right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_2}$$

has an asymptotic $\chi^2$ distribution with $r$ degrees of freedom.

**Corollary 2** Consider the model (14)–(17) with $\theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)'$ and $\theta_2 = (\beta_{q+1}, \ldots, \beta_{q+r})'$. Under the null hypothesis $H_0 : \theta_2 = 0$, the statistic (7), where

$$\hat{a}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \hat{\psi}_i^{-1} \left( 1, x_{i-1}, \ldots, x_{i-m}, \hat{\psi}_{i-1}, \ldots, \hat{\psi}_{i-q} \right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_1}$$

$$\hat{b}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} \left( \hat{\psi}_{i-q-1}, \ldots, \hat{\psi}_{i-q-r} \right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_2}$$

has an asymptotic $\chi^2$ distribution with $r$ degrees of freedom.

### 3.2 Testing LOGACD($m,q$) against higher-order alternatives

Bauwens and Giot (2000) advocate the use of a logarithmic version of the ACD model instead of the linear one. In their LOGACD($m,q$) model (2) is parameterized as

$$\ln \psi_i = \omega + \sum_{j=1}^{m} \alpha_j \ln x_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j}.$$

Also in this case the starting point for modeling would be a low-order model, which is then evaluated. Testing the logarithmic model against higher-order alternatives is
Consider the model (18)–(21) with Corollary 3 of the alternatives mentioned above. Corollaries 3 and 4 define the tests. Under the alternative the models are the LOGACD model. Under the alternative the models are the LOGACD model, respectively. The alternatives belong to the multiplicative class of alternatives mentioned above. Corollaries 3 and 4 define the tests.

**Corollary 3** Consider the model (18)–(21) with \( \theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)' \) and \( \theta_2 = (\alpha_{m+1}, \ldots, \alpha_{m+r})' \). Under the null hypothesis \( H_0 : \theta_2 = 0 \), the statistic (8), where

\[
\begin{align*}
\hat{\alpha}_i &= \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \left(1, \ln x_{i-1}, \ldots, \ln x_{i-m}, \ln \hat{\psi}_{i-1}, \ldots, \ln \hat{\psi}_{i-q}\right)' + \sum_{j=1}^{q} \beta_j \frac{\partial \ln \hat{\psi}_{i-j}}{\partial \theta_1} \\
\hat{\beta}_i &= \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} \left(\ln x_{i-m-1}, \ldots, \ln x_{i-m-r}\right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_2}
\end{align*}
\]

has an asymptotic \( \chi^2 \) distribution with \( r \) degrees of freedom.

**Corollary 4** Consider the model (22)–(25) with \( \theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)' \) and \( \theta_2 = (\beta_{q+1}, \ldots, \beta_{q+r})' \). Under the null hypothesis \( H_0 : \theta_2 = 0 \), the statistic (8), where

\[
\begin{align*}
\hat{\alpha}_i &= \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \left(1, \ln x_{i-1}, \ldots, \ln x_{i-m}, \ln \hat{\psi}_{i-1}, \ldots, \ln \hat{\psi}_{i-q}\right)' + \sum_{j=1}^{q} \beta_j \frac{\partial \ln \hat{\psi}_{i-j}}{\partial \theta_1} \\
\hat{\beta}_i &= \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} \left(\ln \hat{\psi}_{i-q-1}, \ldots, \ln \hat{\psi}_{i-q-r}\right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \ln \hat{\psi}_{i-j}}{\partial \theta_2}
\end{align*}
\]
has an asymptotic $\chi^2$ distribution with $r$ degrees of freedom.

### 3.3 Testing the hypothesis of no remaining ACD

After estimating an ACD model one may also ask whether the estimated disturbances still contain some structure. One possibility is that all the ACD effects are not captured by the estimated model but that some are still present in the disturbances. In the ACD literature it is common to evaluate the properties of standardized durations resulting from the estimation of an ACD model using Ljung-Box or McLeod-Li tests (see Ljung and Box (1978) and McLeod and Li (1983), respectively). As was shown by Li and Mak (1994) in the context of GARCH models, this is somewhat misleading. The reason is that these test statistics do not have the usual asymptotic $\chi^2$ distribution under the null hypothesis when they are applied to standardized residuals from an estimated GARCH model. Li and Mak (1994) proposed a corrected statistic and Lundbergh and Teräsvirta (2002) presented a Lagrange multiplier statistic asymptotically equivalent to it. A similar test statistic for the ACD($m,q$) model is presented next.

To this end, let

$$x_i = \psi_i \varepsilon_i$$

(26)

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}$$

(27)

$$\varphi_i = 1 + \sum_{j=1}^{m^*} \alpha_j \frac{x_{i-j}}{\psi_{i-j}}$$

(28)

$$\varepsilon_i \sim \text{i.i.d. exp}(1).$$

(29)

The null hypothesis equals $H_0: \varphi_i \equiv 1$, i.e., $\alpha_1^* = \ldots = \alpha_{m^*}^* = 0$. The test will be based on the following corollary to Theorem 2.

**Corollary 5** Consider the model (26)–(29) with $\theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)'$ and $\theta_2 = (\alpha_1^*, \ldots, \alpha_{m^*}^*)'$. Under the null hypothesis $H_0: \theta_2 = 0$, the statistic (8), where

$$\hat{\mathbf{a}}_i = \frac{1}{\psi_i} \frac{\partial \hat{\varphi}_i}{\partial \theta_1} = \hat{\psi}_i^{-1} \left(1, x_{i-1}, \ldots, x_{i-m}, \hat{\psi}_{i-1}, \ldots, \hat{\psi}_{i-q}\right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_1}$$

$$\hat{\mathbf{b}}_i = \frac{1}{\psi_i} \frac{\partial \hat{\varphi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} (x_{i-1} \hat{\psi}_{i-1}^{-1}, \ldots, x_{i-m} \hat{\psi}_{i-m}^{-1}, \hat{\psi}_{i-q})'$$

has an asymptotic $\chi^2$ distribution with $m^*$ degrees of freedom.

As $x_{i-1}/\psi_{i-1}$ are standardized durations, the test is one of the standardized durations being iid against the alternative that they follow an ACD process. When $m = q = 0$ it collapses to a test of no ACD effects in the original series. Tests of no remaining ACD effects after estimating other ACD type models are obtained by redefining $\hat{\mathbf{a}}_i$ in Corollary 5.
Recently, Li and Yu (2003) derived a portmanteau test of testing the null hypothesis that the exponentially distributed errors are independent. It turns out that their test is asymptotically equivalent to the statistic given in Corollary 5. For a proof, see Appendix B.

4 Smooth transition ACD models

Engle and Russell (1998) report that their linear ACD model generates expected durations that are on the average too long after the shortest and the longest durations. This suggests that a nonlinear specification for the conditional duration would be more appropriate than the standard linear one. Alternatives in the literature include the LOGACD model of Bauwens and Giot (2000), the Box-Cox and Exponential ACD models of Dufour and Engle (2000), and the threshold ACD model of Zhang, Russell, and Tsay (2001). A smooth transition version of the ACD model also appears to be a possibility and is considered here.

The inspiration for smooth transition ACD models comes from the GARCH literature. Smooth transition GARCH models are treated in Hagerud (1996), González-Rivera (1998) and Anderson, Nam, and Vahid (1999); see also Lundbergh and Teräsvirta (2002). In the present work, the smooth transition ACD($m,q$) model is defined as follows:

$$
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{m} \left( \omega_j^* + \alpha_j^* x_{i-j} \right) G \left( x_{i-j}; \gamma, c \right) + \sum_{j=1}^{q} \beta_j \psi_{i-j}
$$

where $G \left( x_{i-j}; \gamma, c \right)$ is a suitably chosen bounded and non-negative transition function. A natural candidate for the transition function could at first sight be the logistic function. A disadvantage of this transition function is, however, that the logistic function is defined on the whole real axis, whereas in the present case the potential transition variable $x_{i-j}$ only takes positive values. Another candidate for the transition function would be a cumulative distribution function of a random variable with a positive support. The shortcoming of this alternative is that in this case the transition function would inevitably be a non-decreasing function, which will not produce nonlinearities of the type we are interested in.

As the logarithmic transformation is a common and often convenient way of transforming positive-valued objects to ones defined on the whole real axis, we retain the logistic function but use $\ln x_{i-j}$ as the transition variable. This leads to the following smooth transition ACD($m,q$) (STACD) specification

$$
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{m} \left( \omega_j^* + \alpha_j^* x_{i-j} \right) G \left( \ln x_{i-j}; \gamma, c \right) + \sum_{j=1}^{q} \beta_j \psi_{i-j}
$$

where $G \left( \ln x_{i-j}; \gamma, c \right)$ is a suitably chosen bounded and non-negative transition function. A natural candidate for the transition function could at first sight be the logistic function. A disadvantage of this transition function is, however, that the logistic function is defined on the whole real axis, whereas in the present case the potential transition variable $x_{i-j}$ only takes positive values. Another candidate for the transition function would be a cumulative distribution function of a random variable with a positive support. The shortcoming of this alternative is that in this case the transition function would inevitably be a non-decreasing function, which will not produce nonlinearities of the type we are interested in.

As the logarithmic transformation is a common and often convenient way of transforming positive-valued objects to ones defined on the whole real axis, we retain the logistic function but use $\ln x_{i-j}$ as the transition variable. This leads to the following smooth transition ACD($m,q$) (STACD) specification

$$
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{m} \left( \omega_j^* + \alpha_j^* x_{i-j} \right) G \left( \ln x_{i-j}; \gamma, c \right) + \sum_{j=1}^{q} \beta_j \psi_{i-j}
$$

(30)
Evaluating models of autoregressive conditional duration

where

\[ G(\ln x_{i-j}; \gamma, c) = \left(1 + \exp \left\{ -\gamma \prod_{k=1}^{K} (\ln x_{i-j} - c_k) \right\} \right)^{-1}, \quad c_1 \leq \cdots \leq c_K, \quad \gamma > 0 \] (31)

and where the order \( K \in \mathbb{Z}_+ \) determines the general shape of the transition function.

We also propose the smooth transition LOGACD\((m,q)\) specification

\[
\ln \psi_i = \omega + \sum_{j=1}^{m} \alpha_j \ln x_{i-j} + \sum_{j=1}^{m} (\omega^*_j + \alpha^*_j \ln x_{i-j}) G(\ln x_{i-j}; \gamma, c) + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \\
= \omega + \sum_{j=1}^{m} \omega^*_j G(\ln x_{i-j}; \gamma, c) + \sum_{j=1}^{m} (\alpha_j + \alpha^*_j G(\ln x_{i-j}; \gamma, c)) \ln x_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j}.
\]

As \( K = 1 \) the transition function is an increasing function of the lagged duration. In order to capture the effects of very short and long durations we concentrate on the choice \( K = 2 \), which allows these extreme durations to have an impact different from the one of the more average durations. For illustration, Figure 1 contains the transition function (31) for \( K = 2 \) and a set of values for \( \gamma, c_1 \) and \( c_2 \). Assuming that the durations are transformed to take diurnal variation into account (see Section 7.1 for an explanation on how this is done) the average duration equals one, and shorter and longer than expected durations (at that time of the day) will be represented by durations less than and greater than one, respectively. For example, with the second steepest transition function in Figure 1, durations between (approximately) \( 1/2 \) and 2 would belong to the “normal” regime, whereas durations less than 1/3 and greater than 3 would belong to the “extreme” regime (here a transformed duration of \( x \) represents a duration of \( x \) times the expected duration at that time of the day).

The smooth transition ACD model is closely related to the threshold ACD model of Zhang, Russell, and Tsay (2001). We restrict our comparison to the model they refer to as the TACD(1,1) model. It is defined as follows:

\[
x_i = \psi_i \epsilon_i^{(j)} \\
\psi_i = \omega^{(j)} + \alpha^{(j)} x_{i-1} + \beta^{(j)} \psi_{i-1}
\]

whenever \( x_{i-1} \in [r_{j-1}, r_j), \) \( j = 1, 2, \ldots, J \), where \( J \) is the number of different regimes and \( 0 = r_0 < r_1 < \cdots < r_J = \infty \) are the threshold values. The parameter values \( \omega^{(j)}, \alpha^{(j)} \) and \( \beta^{(j)} \) as well as the distribution of \( \epsilon_i^{(j)} \) are allowed to vary depending on the regime. The two-regime TACD(1,1) model with the restrictions \( \beta^{(1)} = \beta^{(2)} \) and \( \epsilon_1^{(1)} \sim \epsilon_2^{(2)} \sim \exp(1) \) is achieved as the limiting case of the STACD(1,1) model with \( K = 1 \) as \( \gamma \to \infty \). Similarly, the three-regime TACD(1,1) model with \( \beta^{(1)} = \beta^{(2)} = \beta^{(3)}, \omega^{(1)} = \omega^{(3)}, \alpha^{(1)} = \alpha^{(3)} \) and \( \epsilon_i^{(j)} \sim \exp(1), \) \( j = 1, 2, 3, \) is the limiting case of the STACD(1,1) model with \( K = 2 \) as \( \gamma \to \infty \).
Figure 1: The logistic transition function $G(\ln x; \gamma, c)$ as a function of $x$ for $K = 2$, $c_1 = -1$, $c_2 = 1$ and for $\gamma = 1$ (the smoothest), 3, 5, 10 and 50 (the steepest).

4.1 Testing ACD($m,q$) against smooth transition ACD($m,q$)

We now consider testing the ACD model against its smooth transition counterpart. It is seen that model (30) is only identified under the alternative. For example, when $\gamma = 0$ (this is one form of the null hypothesis), parameters $\omega_j^*$ and $\alpha_j^*$, $J = 1, \ldots, m$, as well as $c$, are not identified. Testing when some parameters are identified only under the alternative is discussed for example in Hansen (1996). He studies the (non-standard) asymptotic distribution theory for such tests, and develops a procedure to approximate these distributions by simulation. As our goal is to derive easily applicable misspecification tests, we do not follow Hansen’s approach, but instead use the method suggested in Luukkonen, Saikkonen, and Teräsvirta (1988). In their approach the identification problem is solved by approximating the transition function with its first-order Taylor expansion around $\gamma = 0$. This will lead to an approximate alternative, which is free of nuisance parameters under the null.

To this end, define

$$x_i = \psi_i \varepsilon_i$$

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}$$

$$+ \sum_{j=1}^{m} (\omega_j^* + \alpha_j^* x_{i-j}) \bar{G}(\ln x_{i-j}; \gamma, c)$$

$$\varepsilon_i \sim \text{i.i.d. exp(1)}$$

where $\bar{G}(\ln x_{i-j}; \gamma, c) = G(\ln x_{i-j}; \gamma, c) - \frac{1}{2}$ (subtracting $\frac{1}{2}$ simplifies the derivation below but does not affect the conclusions, because we can replace $\bar{G}$ by $G$ with a
simple reparameterization). Using Taylor’s theorem one obtains as
\[
G(\ln x_{i-j}; \gamma, c) = \bar{G}(\ln x_{i-j}; 0, c) + \frac{\partial \bar{G}(\ln x_{i-j}; 0, c)}{\partial \gamma}(\gamma - 0) + G(\ln x_{i-j}; \bar{\gamma}, c)
\]
\[
= \frac{1}{4}\gamma \prod_{k=1}^{K}(\ln x_{i-j} - c_k) + \bar{G}(\ln x_{i-j}; \bar{\gamma}, c)
\]
\[
= \sum_{l=0}^{K} \gamma \tilde{c}_l (\ln x_{i-j})^l + \bar{G}(\ln x_{i-j}; \bar{\gamma}, c)
\]
(33)

where \(\bar{\gamma} \in [0, \gamma]\). Applying (33) to (32) yields
\[
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}
\]
\[
+ \sum_{j=1}^{m} \left( (\omega_j^* + \alpha_j^* x_{i-j}) \sum_{l=0}^{K} \gamma \tilde{c}_l (\ln x_{i-j})^l \right)
\]
\[
+ \sum_{j=1}^{m} (\omega_j^* + \alpha_j^* x_{i-j}) \bar{G}(\ln x_{i-j}; \bar{\gamma}, c)
\]
\[
= \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + \sum_{j=1}^{m} (\gamma \omega_j^* \tilde{c}_0 + \gamma \alpha_j^* \bar{c}_0 x_{i-j})
\]
\[
+ \sum_{j=1}^{m} \sum_{l=1}^{K} \left[ \gamma \omega_j^* \tilde{c}_l (\ln x_{i-j})^l + \gamma \alpha_j^* \bar{c}_l (\ln x_{i-j})^l \right]
\]
\[
+ \sum_{j=1}^{m} (\omega_j^* + \alpha_j^* x_{i-j}) \bar{G}(\ln x_{i-j}; \bar{\gamma}, c)
\]
(34)

This form does not lead to an operational test statistic as \(\bar{\gamma}\) is unknown. If we instead use Taylor’s theorem to approximate \(G(\ln x_{i-j}; \gamma, c)\), we can drop the remainder term \(\bar{G}(\ln x_{i-j}; \bar{\gamma}, c)\) from the last expression in (34). Doing this and renaming parameters yields the following approximation to the conditional mean of the alternative:
\[
\psi_i \approx \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + \sum_{j=1}^{m} \sum_{l=1}^{K} \left( d_{jl} (\ln x_{i-j})^l + e_{jl} x_{i-j} (\ln x_{i-j})^l \right).
\]
Using this approximation we have transformed the original testing problem into testing the ACD\((m,q)\) model against the approximate alternative

\[
x_i = (\psi_i + \varphi_i) \varepsilon_i \tag{35}
\]

\[
\psi_i + \varphi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j (\psi_{i-j} + \varphi_{i-j}) + \sum_{j=1}^{m} \sum_{l=1}^{K} \left(d_{jl} (\ln x_{i-j})^l + e_{jl} x_{i-j} (\ln x_{i-j})^l\right) \tag{36}
\]

\[
\varphi_i = \sum_{j=1}^{q} \beta_j \varphi_{i-j} + \sum_{j=1}^{m} \sum_{l=1}^{K} \left(d_{jl} (\ln x_{i-j})^l + e_{jl} x_{i-j} (\ln x_{i-j})^l\right) \tag{37}
\]

\[
\varepsilon_i \sim \text{i.i.d. exp}(1). \tag{38}
\]

Model (35)–(38) reduces to the null model when \(d_{jl} = e_{jl} = 0\) for \(j = 1, \ldots, m\) and \(l = 1, \ldots, K\), and there are no unidentified parameters under the null. This enables us to use Theorem 1 to derive the test, which is given in the following corollary.

**Corollary 6** Consider the model (35)–(38) and denote \(\theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)'\) and \(\theta_2 = ((\text{vec}D)', (\text{vec}E)')'\), where \(D = [d_{jl}]\), \(E = [e_{jl}]\), \(j = 1, \ldots, m\), \(l = 1, \ldots, K\), are \((m \times K)\) matrices, and the vec-operator stacks the columns of the matrix. Furthermore, denote \(X_{i,1} = \left[(\ln x_{i-j})^l\right]\) and \(X_{i,2} = \left[x_{i-j} (\ln x_{i-j})^l\right]\), \(j = 1, \ldots, m\), \(l = 1, \ldots, K\). Under the null hypothesis \(H_0 : \theta_2 = 0\) the statistic (7), where

\[
\hat{a}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \hat{\psi}_i^{-1} \left(1, x_{i-1}, \ldots, x_{i-m}, \hat{\psi}_{i-1}, \ldots, \hat{\psi}_{i-q}\right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_1}
\]

\[
\hat{b}_i = \frac{1}{\psi_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} \left((\text{vec}X_{i,1})', (\text{vec}X_{i,2})'\right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_2}
\]

has an asymptotic \(\chi^2\) distribution with \(2mK\) degrees of freedom.

**5 Time-varying ACD models**

In standard econometric analysis of low-frequency time series the observation period easily spans over several years or decades of data. In such a situation it may not be realistic to expect the parameters of any model to remain constant over the whole period. For this reason testing parameter constancy is important. When using ultra-high-frequency data a period of a few days or weeks often yields a sufficient amount of observations. Thus it does not seem inappropriate to assume that the parameters actually remain constant over the observation period. On the other hand, certain events affecting the economic or institutional environment could cause the structure of the trading process to change. In such a situation, fitting an ACD model with
constant parameters to the observed durations may yield unsatisfactory results. One remedy to the problem is to split the sample into several periods and estimate separate models for each of them. Identifying the number and location of the break-points becomes, however, a demanding task, but see Zhang, Russell, and Tsay (2001) for an example. An alternative to abrupt changes in the parameters would be a model where the parameters are allowed to change smoothly over time. This can be achieved for example using the logistic transition functions (31) with time as the transition variable.

Even when the model builder does not want to fit ACD models with time-varying parameters to data, they can be used as tools for detecting misspecification. For example, if an ACD model is tested and rejected against a model with time-varying parameters, this might be seen as evidence that the structure of the duration series changes during the period in question and that a more careful analysis is required. On the other hand, if an ACD model is rejected against a model with time-of-day-varying parameters, this could be an indication of the test that the approach used for removing the diurnal pattern is not satisfactory.

Two definitions of time are considered for ACD models with time-varying parameters. The first one is the total trading time (in seconds) from the beginning of the sample to the end, and is called the (total) time. In our empirical application in Section 7, each one of the samples consists of one week of data, and the time thus runs from the beginning of the first trading day of the week till the end of the last day of the week. Hence the term intraweek time is also used. The second definition, called the intraday time, is time measured in seconds from the beginning of the trading day. Both measures are for convenience and numerical stability standardized to obtain values between 0 and 1. As each financial event considered has a precise time-stamp attached to it, these two time definitions are readily available.

This leads to the following time-varying ACD (TVACD) specification

\[
\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + \left( \omega^* + \sum_{j=1}^{m} \alpha_j^* x_{i-j} + \sum_{j=1}^{q} \beta_j^* \psi_{i-j} \right) G(t_{i-1}; \gamma, c)
\]

\[
= (\omega + \omega^* G(t_{i-1}; \gamma, c)) + \sum_{j=1}^{m} (\alpha_j + \alpha_j^* G(t_{i-1}; \gamma, c)) x_{i-j}
\]

\[
+ \sum_{j=1}^{q} (\beta_j + \beta_j^* G(t_{i-1}; \gamma, c)) \psi_{i-j}
\]

where \( G(t_{i-1}; \gamma, c) = G(t_{i-1}; \gamma, c) - \frac{1}{2} \) and \( G \) is the transition function given in (31) except that the transition variable used now is \( t_{i-1} \), which can correspond to either one of the two time definitions. If we consider parameter constancy or structural breaks in the process, then the total time is the one to be used. If the issue is how well the diurnal pattern has been removed then the intraday time is the appropriate measure. Analogous definitions are available for the LOGACD model.
5.1 Testing parameter constancy

We now consider testing an estimated ACD model against these time-varying alternatives. The identification problem already discussed is present in the current situation as well. Therefore it is necessary to test for the presence of time-varying parameters before estimating any TVACD model. The identification problem is again solved using a Taylor series approximation of the transition function. Arguments similar to the ones in Section 4.1 lead to the following approximation to the TVACD(m,q) model:

\[ x_i = (\psi_i + \varphi_i) \varepsilon_i \]  
\[ \psi_i + \varphi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j (\psi_{i-j} + \varphi_{i-j}) \]
\[ + \sum_{l=1}^{K} d_l t_{i-1}^l + \sum_{j=1}^{m} \sum_{l=1}^{K} e_{jl} x_{i-j} t_{i-1}^l + \sum_{j=1}^{q} \sum_{l=1}^{K} f_{jl} \psi_{i-j} t_{i-1}^l \]  
\[ \varphi_i = \sum_{j=1}^{q} \beta_j \varphi_{i-j} + \sum_{l=1}^{K} d_l t_{i-1}^l + \sum_{j=1}^{m} \sum_{l=1}^{K} e_{jl} x_{i-j} t_{i-1}^l + \sum_{j=1}^{q} \sum_{l=1}^{K} f_{jl} \psi_{i-j} t_{i-1}^l \]
\[ \varepsilon_i \sim \text{i.i.d. exp}(1). \]

The model reduces to the null model when \( d_l = 0 \) (\( l = 1, \ldots, K \)), \( e_{jl} = 0 \) (\( j = 1, \ldots, m \) and \( l = 1, \ldots, K \)) and \( f_{jl} = 0 \) (\( j = 1, \ldots, q \) and \( l = 1, \ldots, K \)). This enables us to use Theorem 1 to derive the test, which is given in the following corollary. The total time can be replaced with the intraday time without affecting the validity of the result.

**Corollary 7** Consider the model (39)–(42) and denote \( \theta_1 = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)' \) and \( \theta_2 = (d_1, \ldots, d_K, (\text{vec}E)'(\text{vec}F)')' \), where \( E = [e_{jl}] \) (\( j = 1, \ldots, m \), \( l = 1, \ldots, K \)) is a \((m \times K)\) matrix and \( F = [f_{jl}] \) (\( j = 1, \ldots, q \), \( l = 1, \ldots, K \)) is a \((q \times K)\) matrix, and the \( \text{vec} \)-operator stacks the columns of the matrix. Furthermore, denote \( X_{i,1} = [x_{i-j} t_{i-1}^l] \) (\( j = 1, \ldots, m \), \( l = 1, \ldots, K \)) and \( X_{i,2} = [\psi_{i-j} t_{i-1}^l] \) (\( j = 1, \ldots, q \), \( l = 1, \ldots, K \)). Under the null hypothesis \( H_0 : \theta_2 = 0 \) the statistic (7), where

\[ \hat{a}_t = \frac{1}{\hat{\psi}_i} \frac{\partial \hat{\psi}_i}{\partial \theta_1} = \hat{\psi}_i^{-1} \left( 1, x_{i-1}, \ldots, x_{i-m}, \hat{\psi}_{i-1}, \ldots, \hat{\psi}_{i-q} \right)' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_1} \]
\[ \hat{b}_t = \frac{1}{\hat{\psi}_i} \frac{\partial \hat{\psi}_i}{\partial \theta_2} = \hat{\psi}_i^{-1} (t_{i-1}, \ldots, t_{i-1}^K, (\text{vec}X_{i,1})', (\text{vec}X_{i,2})')' + \hat{\psi}_i^{-1} \sum_{j=1}^{q} \beta_j \frac{\partial \hat{\psi}_{i-j}}{\partial \theta_2} \]

has an asymptotic \( \chi^2 \) distribution with \((1 + m + q)K\) degrees of freedom.
6 Simulation experiment

6.1 Size simulations

We investigate the finite sample properties of the test statistics by simulation and begin with size simulations. The data generating process we use has the form

\[ x_i = \psi_i \varepsilon_i \]

\[ \psi_i = 0.15 + 0.10x_{i-1} + 0.80\psi_{i-1} \] (43)

\[ \varepsilon_i \sim \text{i.i.d. exp}(1). \]

The parameters in (43) have been chosen such that the model is representative for the estimated ACD(1,1) models reported in the literature. We use sample sizes \( n = 1000, 5000, \) and \( 10000. \) The smallest size is extremely small in the context of ACD models, and the largest one is still considerably smaller than the sample sizes in our empirical example. To avoid initialization effects we discard 1000 observations from the beginning of each generated series. The number of replications is 10000. In each replication, an ACD(1,1) model is estimated and then evaluated using five different tests. These are the tests against ACD(2,1) and ACD(1,2) models, test of no remaining ACD (of order one) in the standardized durations, and linearity tests against smooth transition ACD of orders one and two. For each test both the ordinary version using the auxiliary regression and the robust version are computed.

Results of the experiment are presented graphically in Figure 2. For each test we calculate the actual rejection frequencies for the nominal significance levels 0.1\%, 0.2\%, \ldots, 5.0\%. The graphs show the discrepancies in size, i.e. the difference between the actual and the nominal size. In each subgraph we present the results for one of the tests using all the three sample sizes.

As can be seen, all the tests are rather well-sized for \( n = 5000 \) and 10000. The distortions are not very severe for the smallest sample size either. It can be concluded that the asymptotic null distributions of the test statistics are reasonably good approximations to the unknown finite-sample distributions for \( n \geq 5000. \) Such sample sizes are standard in the analysis of ultra-high-frequency data.

In order to complete the experiment, we also investigated the effect of having more persistence in the data generating process. The design of the experiment was the same as before except that the parameter values in (43) were changed to \( \psi_i = 0.05 + 0.09x_{i-1} + 0.90\psi_{i-1}. \) The results were similar to those reported in Figure 2 and are not presented here (they are available upon request). The only notable change was a slight increase in the empirical size of the test against the ACD(1,2) alternative.

To explore the effects of misspecification of the error distribution (and hence of the conditional distribution of the durations) on the finite-sample properties of our test statistics, we repeated the simulations with different error distributions. They were the Weibull distribution with shape parameter values 0.8 and 0.9, and the generalized gamma distribution with shape parameter values 8 and 0.3 as well as 10 and 0.5; see for example Lunde (1999) for definitions of density functions of these distributions. The distributions were scaled to have an expected value of one. With
Figure 2: Results from size simulations of the LM tests. In the figures the size discrepancy (i.e. the actual size less the nominal size) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the tests against ACD(2,1) and ACD(1,2) models, tests of no remaining ACD (of order one) in the standardized durations, and tests of no smooth transition ACD of orders one and two. Both the ordinary and robust versions of the tests are used. The three lines in each subfigure correspond to sample sizes 1000 (+), 5000 (×) and 10000 (*).
the chosen parameter values the Weibull distribution has a monotonically decreasing
hazard function, whereas for the generalized gamma distribution the hazard function
is inverted U–shaped. As the results from these simulations were similar to the ones
already reported, we do not show them here. Our general conclusion is that both the
non-robust and the robust versions of the tests remain well-sized for sample sizes over
5000. It appears that if the conditional distribution of the durations is only mildly
misspecified and if the sample size is sufficiently large, even the non-robust versions
of the tests have satisfactory size properties.

6.2 Power simulations

To evaluate the performance of our test statistics we perform a power comparison to
a benchmark test. As the test to compare with we choose the generalized spectral
density based test introduced in Hong (1999). This test has been proposed as a
general diagnostic test for a wide class of time series models, including ACD models,
by Hong and Lee (2003). In the context of ACD models, it is a test of the standardized
durations being iid against an unspecified alternative. As such it is an omnibus test
against any kind of pairwise dependence structure in the standardized durations, and
the simulation study in Hong and Lee (2003) suggests that the test has good power
against a wide variety of alternatives. Users of this test have to make some parameter
choices, and we both describe the test and discuss our choices in Appendix C.

We begin by repeating the size simulation experiment of the previous subsection
for the generalized spectral test. The design of the experiment is exactly the same as
earlier, except that we consider only sample sizes of 1000 and 5000, and perform only
1000 replications. This parsimony is due to the fact that the generalized spectral test
is computationally much more burdensome than our tests. As explained in Appendix
C, performing the test also involves the choice of a preliminary bandwidth, \( \hat{p} \). We
use the values 25, 50, 75, and 100. The size discrepancies of the test are presented
in Figure 3 (note the different scale on the y-axes compared to Figure 2). The tests
seem to be slightly undersized for all choices of the preliminary bandwidth when the
nominal significance level approaches 5%. Note, however, that the size distortion
diminishes with increasing \( \hat{p} \).

In the power simulations the restricted model to be estimated is always an ACD(1,1)
model. We consider two alternative data generating processes. The first one of these
is an ACD(2,1) specification given by

\[
x_i = \psi_i \varepsilon_i \\
\psi_i = 0.15 + 0.10x_{i-1} + 0.05x_{i-2} + 0.80\psi_{i-1} \\
\varepsilon_i \sim \text{i.i.d. exp}(1).
\]

Because of relatively large sample sizes, the coefficient of \( x_{i-2} \) is chosen to be close
to zero. This is an alternative against which our test against higher-order models is
expected to have very good power. As the second alternative model we consider a
Figure 3: Results from size simulations of the generalized spectral tests. In the figures the size discrepancy (i.e. the actual size less the nominal size) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the generalized spectral tests for preliminary bandwidths $\bar{p} = 25, 50, 75, \text{ and } 100$. The two lines in each subfigure correspond to sample sizes 1000 (+) and 5000 (×).

A three-regime threshold ACD(1,1) model given by

$$x_t = \psi_i \varepsilon_t$$

(47)

$$\psi_i = \begin{cases} 
0.05 + 0.20 x_{t-1} + 0.85 \psi_{t-1} & \text{for } 0 < x_{t-1} < 0.25 \\
0.10 + 0.05 x_{t-1} + 0.90 \psi_{t-1} & \text{for } 0.25 \leq x_{t-1} < 1.5 \\
0.20 + 0.03 x_{t-1} + 0.80 \psi_{t-1} & \text{for } 1.5 \leq x_{t-1} < \infty 
\end{cases}$$

(48)

$$\varepsilon_t \sim \text{i.i.d. exp}(1).$$

(49)

The parameter values in this model are chosen such that it resembles the threshold ACD models estimated in Zhang, Russell, and Tsay (2001). This model is not a special case of our STACD model. The first regime (for the smallest values of $x_{t-1}$) is an explosive one. The other two regimes are stable, the middle one being more persistent than the third one.

Our sample sizes are 1000 and 5000, and we perform 1000 replications. As before, we discard 1000 observations from the beginning of each series. An ACD(1,1) model is fitted to each series. The diagnostic tests performed are the tests against ACD(2,1) and ACD(1,2) models, the test of no remaining ACD (of order one) in standardized
Evaluating models of autoregressive conditional duration

**Figure 4a:** Results from power simulations of the tests using the ACD(2,1) model given in equations (44)–(46) as the alternative data generating process. In the figures the power (rejection frequency) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the test against a higher-order ACD(2,1) model, test of no remaining ACD, and the generalized spectral tests for preliminary bandwidths \( \bar{p} = 25 \) and 100. The two lines in each subfigure correspond to sample sizes 1000 (+) and 5000 (×).

---

durations, and tests of no smooth transition ACD of orders one and two. We also apply the generalized spectral test using preliminary bandwidths 25, 50, 75, and 100. We compute both the ordinary and the robustified versions of the LM tests.

Results of the power simulations are presented in Figures 4a and 4b. In order to conserve space we do not present all of the results, but they are available upon request. These figures show the rejection frequencies of the tests for both sample sizes at the nominal significance levels of 0.1%, 0.2%, ..., 5.0%. Since our size simulations indicated that all the tests in question are well-sized, the power results are not size-adjusted.

Figure 4a reports the powers of the tests when the alternative data generating process is the ACD(2,1) model given in equations (44)–(46). The upper panel presents the power levels for the non-robustified versions of the test against an ACD(2,1) model and the test of no remaining ACD in the standardized durations. Both of these tests have rather good power for \( n = 5000 \), whereas the power is still low when \( n = 1000 \). Without presenting the results we note that the test against an ACD(1,2) model has almost equally good power, whereas the tests against smooth transition ACD have
Figure 4b: Results from power simulations of the tests using the threshold ACD(1,1) model given in equations (47)–(49) as the alternative data generating process. In the figures the power (rejection frequency) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the tests of no smooth transition ACD of orders 1 and 2, and the generalized spectral tests for preliminary bandwidths \( \bar{p} = 25 \) and 100. The two lines in each subfigure correspond to sample sizes 1000 (+) and 5000 (×). 

no power at all against this alternative. Furthermore, the robustified versions of the tests have power very close to the non-robust ones.

In the lower panel we show the power of the generalized spectral test using preliminary bandwidths 25 and 100. Both of these tests have moderate power, the one with \( \bar{p} = 25 \) being somewhat more powerful. The powers of the tests with \( \bar{p} = 50 \) and 75 (not shown) are very similar to the ones shown. Our tests against higher-order ACD thus have clearly higher power than the benchmark. This is natural because these tests are designed to have power against this particular alternative. If they had been only slightly more powerful or even less powerful than the generalized spectral density based tests, that would have warned us that the small-sample properties of our test would leave much to desire.

Power results for the threshold ACD(1,1) model of equations (47)–(49) can be found in Figure 4b. The non-robustified versions of the linearity tests against smooth transition of orders one and two both have very good power for \( n = 5000 \), and low power for \( n = 1000 \) (upper panel). Without presenting the results we note that the tests against higher-order models and of no remaining ACD have no power at all, and
that the power of the robustified versions of the tests again have power very close to the non-robust ones. The generalized spectral test using preliminary bandwidths 25 and 100 has rather low power against this alternative (the powers for the tests with \( \hat{p} = 50 \) and 75 are very similar).

### 6.3 Performance of the Ljung-Box and McLeod-Li tests

As mentioned in Section 3.3, it is common in the ACD literature to evaluate the properties of estimated standardized durations using Ljung-Box or McLeod-Li tests. This practice can, however, result in misleading conclusions, because these test statistics do not have the usual asymptotic \( \chi^2 \) distribution under the null hypothesis when they are applied to standardized durations from an estimated ACD model. In order to take a closer look at this possibility, we shall investigate the properties of these test statistics when they are applied to standardized durations from an estimated ACD model.

We perform exactly the same simulation experiments as in the previous subsection, except that we also use the sample size 10000. The tests applied to the standardized durations from the estimated ACD(1,1) models are the Ljung-Box and McLeod-Li tests with lag lengths 1, 5, 10, 15, and 20. The rejection frequencies in both the size and power simulations are based on the (incorrect) asymptotic \( \chi^2 \) distribution with degrees of freedom equal to the lag length used.

The size discrepancies of the tests with lag length 15 are presented in Figure 5a (results with the other lag lengths are similar and thus omitted). The Ljung-Box test seems to be somewhat undersized. On the other hand, the McLeod-Li test is oversized: quite strongly for \( n = 1000 \) and less so for the two larger sample sizes.

In Figure 5b we present the power of the tests against the ACD(2,1) (the two upper figures) and threshold ACD(1,1) (the two lower figures) alternatives given in equations (44)–(46) and (47)–(49), respectively. Again, we only present the results with lag length 15. It can be seen that the Ljung-Box test has moderate power against the ACD(2,1) alternative. A comparison with the results in Figure 4a indicates that the test is less powerful than the tests considered there. This is not surprising because the size simulations showed that the Ljung-Box test is conservative. The McLeod-Li test has almost no power at all against this alternative, which may not be unexpected either. This is because one cannot expect to discover a misspecified lag length using a test based on squared standardized durations. Finally, neither of the tests has any notable power against the threshold ACD model.

### 7 Application to Trades and Quotes series

#### 7.1 Description of the data

In this section we apply our evaluation tests to ACD models fitted to Trades and Quotes (TAQ) data available from the New York Stock Exchange (NYSE). We use a period of six months from the beginning of July 2002 to the end of December 2002 and concentrate on intertrade durations between transactions of IBM shares.
Figure 5a: Results from size simulations of the Ljung-Box and McLeod-Li tests. In the figures the size discrepancy (i.e. the actual size less the nominal size) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the Ljung-Box (left) and McLeod-Li (right) tests with lag length 15. The three lines in each subfigure correspond to sample sizes 1000 (+), 5000 (×) and 10000 (∗).

Before using the data we remove the trades which are uncorrected and irregular. This can be done using the correction indicator attached to each trade. We also remove all trades occurring before 9:30 am. and after 4:00 pm. The NYSE was entirely closed on July 4 (Independence Day), September 2 (Labor Day), November 28 (Thanksgiving Day) and December 25 (Christmas Day), and partly closed on July 5, September 11, November 29 and December 24. For this reason all these days have been removed from the sample. Furthermore, we only consider unique trading times and hence simultaneously recorded trades are regarded as a single trade. Finally, the trades are treated consecutively from day to day, ignoring the overnight duration.

Owing to the enormous amount of trades, nearly half a million during the period in question prior to removals, we consider every week in the sample separately and perform the estimation and tests only on complete five-day weeks. This leaves us with 21 weeks with approximately 15000–20000 trades in each. The exact dates, numbers of trades and some summary statistics of durations can be found in Table 1. We note that the consecutive treating of the trades implies, in particular, that over each of the 21 weeks, estimation is not re-initialized at the beginning of every day. This practice differs from the approach of Engle and Russell (1998) but is applied by Bauwens and Giot (2000).

As is well documented in the literature, there is a strong diurnal component in the duration series. The durations tend to be shorter around the beginning and the end of a trading day, when traders open and close their positions, respectively, and longer around lunchtime. This results in an inverted U–shape pattern in the moving average of durations over the day. A common practice in the literature is to first “diurnally adjust” the series by approximating the average durations using a cubic spline and then removing this diurnal component from the durations. Such a procedure was used for example by Bauwens and Giot (2000). We do this by first averaging the durations
Figure 5b: Results from power simulations of the tests using the ACD(2,1) model given in equations (44)–(46) and the threshold ACD(1,1) model given in equations (47)–(49) as the alternative data generating process (two upper and two lower figures, respectively). In the figures the power (rejection frequency) is plotted against the nominal size. Both of them are measured in percentage points. Performed tests are the Ljung-Box (left) and McLeod-Li (right) tests with lag length 15. The three lines in each subfigure correspond to sample sizes 1000 (+), 5000 (×) and 10000 (+).

Over 30-minute periods (9:30–10:00, 10:00–10:30, etc.), setting the average durations of the midpoints of these intervals (9:45, 10:15, etc.) to the resulting averaged values, and then fitting a cubic spline over the course of the day using these midpoints as fixed. The original durations series is then diurnally adjusted by dividing it with the estimated diurnal component. This is done separately for each day of the sample, since the time-of-day component varies depending on the day of the week.

We illustrate the diurnal components in Figure 6 that contains the estimated diurnal components averaged over the 21 weeks, for each day of the week separately. It can be seen that the durations are longest around the midday, typically somewhat shorter before midday, and shortest near the closing. It should be pointed out, however, that the graphs in Figure 6 are merely for illustrating a general pattern, and that the diurnal patterns differ quite strongly from day to day and from week to week. This, however, is a casual observation not based on any statistical inference.

\[\text{The averaging is performed to produce a single graph for each day of the week and is made for illustrational purposes only.}\]
Figure 6: Graphs of the averaged estimated diurnal components. The diurnal component is estimated separately for each day of the sample. The resulting diurnal components are then averaged over the 21 weeks, for each day of the week separately. Therefore the subfigures represent the (averaged) mean duration (in seconds) at a particular time of the day versus the time of the day (in hours).

Nevertheless, the individual diurnal pattern for any particular day obtained by the cubic spline technique can be very different from the average one for the corresponding day of the week in Figure 6. This raises the question of how a diurnal component of this type should be interpreted: is it that even other events than the time of the day affect its structure? Should the diurnal pattern in fact be estimated jointly with the parameters? This problem is left for further research.

7.2 Estimation and evaluation of ACD(1,1) models

We fit an ACD(1,1) model to each of the 21 diurnally adjusted duration series assuming exponentially distributed errors. The estimation as well as all the other computations are carried out using Ox version 3.30 (see Doornik (2002)). Maximum likelihood estimation of the ACD(1,1) model is performed using the sequential quadratic programming algorithm of Lawrence and Tits (2001) with analytical first derivatives.
The parameter estimates that can be found in Table 2 are significant in all of the 21 cases.

We now subject our estimated models to a battery of evaluation tests. We perform tests against ACD(2,1) and ACD(1,2) models, remaining ACD in the standardized durations, STACD(1,1) models of order one and two, and TVACD(1,1) models of order one and two. In the TVACD case, both the intraday time and the intraweek time are used. The $p$-values of the tests are given in Tables 3 and 4. The values less than 0.01 are shown in boldface.

As to the order of the model, the ACD(1,1) model is rejected in favour of the ACD(2,1) and/or the ACD(1,2) model in seven cases out of 21. As may be expected, in these cases the hypothesis of no remaining ACD in the standardized durations is also (typically) rejected.

The ACD(1,1) model is almost always rejected against the STACD(1,1) model of orders 1 and 2. Perhaps surprisingly, given the findings of Engle and Russell (1998) mentioned in Section 4, rejections using tests based on assuming $K = 2$ (nonmonotonic change) are not systematically stronger than the ones obtained by assuming $K = 1$. More research is needed to find out what kind of nonlinear ACD model would fit the data best. Nevertheless, the results indicate that the linear ACD model does not capture the dynamics of the duration process in a satisfactory fashion and that a nonlinear model should be considered. In addition to the family of STACD models, the TACD model of Zhang, Russell, and Tsay (2001) could be a viable alternative.

Of the tests against time-varying ACD models, we first consider the ones based on intraday time. Almost all of the 21 models pass the test against the first-order TVACD model, but about one third of them are rejected against the second-order one. This rejection may be interpreted as showing that the removal of the diurnal component has not been successful, since there still is an identifiable parameter change in the process within the trading day. If this is a valid interpretation, there seems to be room for improvement in methods for diurnal adjustment of the durations. It is also possible that an erroneous linearity assumption causes these rejections.

When the time is measured as intraweek time all the 21 models pass the tests against time-varying ACD models of orders 1 and 2. It can be concluded that the structure of the duration process does not seem to change within the week in any of the cases. We have also attempted to make the diurnal adjustment using the same time-of-day-curve for all the days of the week. The results from this evaluation test (not shown) are different from the previous ones: the ACD(1,1) model is rejected or nearly rejected at the 1% level in favour of a TVACD model of order 1 and/or 2 in about two thirds of the cases. This suggests that the diurnal pattern is not the same for all days of the week. A tentative conclusion is that conditioning the diurnal adjustment on the day of the week may be of importance.

8 Conclusions

In this paper we present a general framework for evaluating ACD models using Lagrange multiplier or Lagrange multiplier type tests. We derive several misspecification...
tests of the functional form of the conditional mean of an ACD model. The alternatives considered are parametric, and hence, in case of rejection, they may suggest a direction in which to extend the model. Clearly, the test battery may also be viewed as a set of misspecification indicators that convey information about the fit instead of prompting a particular action to extend the model.

Our tests are simple to use, since the model only has to be estimated under the null hypothesis, and computation of any of the test statistics simply requires one or two additional ordinary linear regressions. Versions of the test statistics robust to deviations from distributional assumptions other than those being explicitly tested are also presented but it appears that robustifying the tests is not as important in practice as it is in some other time series applications. All the tests are found to have good size properties.

Results of the application to Trades and Quotes data clearly point out the need for nonlinear ACD models. This issue will be taken up in future work. Diurnal adjustment of durations appears to be another topic worth further consideration.
<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Number of observations</th>
<th>Original durations</th>
<th>Transformed durations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>min, mean, max</td>
<td>min, mean, max</td>
</tr>
<tr>
<td>1</td>
<td>Jul 8–12</td>
<td>18025</td>
<td>1, 6.4814, 217</td>
<td>0.0654, 0.9992, 26.8210</td>
</tr>
<tr>
<td>2</td>
<td>Jul 15–19</td>
<td>18179</td>
<td>1, 6.4278, 118</td>
<td>0.0763, 0.9982, 14.9018</td>
</tr>
<tr>
<td>3</td>
<td>Jul 22–26</td>
<td>18240</td>
<td>1, 6.4078, 92</td>
<td>0.1024, 0.9994, 11.7863</td>
</tr>
<tr>
<td>4</td>
<td>Jul 29 – Aug 2</td>
<td>17088</td>
<td>1, 6.8295, 69</td>
<td>0.0996, 1.0004, 9.8779</td>
</tr>
<tr>
<td>5</td>
<td>Aug 5–9</td>
<td>18585</td>
<td>1, 6.2917, 89</td>
<td>0.1001, 1.0015, 11.6193</td>
</tr>
<tr>
<td>6</td>
<td>Aug 12–16</td>
<td>17544</td>
<td>1, 6.6656, 126</td>
<td>0.1043, 1.0007, 17.0263</td>
</tr>
<tr>
<td>7</td>
<td>Aug 19–23</td>
<td>16900</td>
<td>1, 6.9173, 79</td>
<td>0.0972, 1.0009, 14.0726</td>
</tr>
<tr>
<td>8</td>
<td>Aug 26–30</td>
<td>15172</td>
<td>1, 7.7050, 118</td>
<td>0.0727, 1.0008, 15.3554</td>
</tr>
<tr>
<td>9</td>
<td>Sep 16–20</td>
<td>17388</td>
<td>1, 6.7126, 76</td>
<td>0.0949, 0.9985, 9.8126</td>
</tr>
<tr>
<td>10</td>
<td>Sep 23–27</td>
<td>16273</td>
<td>1, 7.1859, 117</td>
<td>0.1000, 0.9994, 14.3419</td>
</tr>
<tr>
<td>11</td>
<td>Sep 30 – Oct 4</td>
<td>17732</td>
<td>1, 6.5917, 238</td>
<td>0.0831, 1.0001, 22.4224</td>
</tr>
<tr>
<td>12</td>
<td>Oct 7–11</td>
<td>17077</td>
<td>1, 6.8220, 183</td>
<td>0.0993, 0.9987, 24.1145</td>
</tr>
<tr>
<td>13</td>
<td>Oct 14–18</td>
<td>17129</td>
<td>1, 6.8193, 90</td>
<td>0.0930, 0.9990, 12.1304</td>
</tr>
<tr>
<td>14</td>
<td>Oct 21–25</td>
<td>18500</td>
<td>1, 6.3204, 106</td>
<td>0.1090, 1.0000, 14.9705</td>
</tr>
<tr>
<td>15</td>
<td>Oct 28 – Nov 1</td>
<td>18702</td>
<td>1, 6.2453, 80</td>
<td>0.1207, 0.9997, 10.9473</td>
</tr>
<tr>
<td>16</td>
<td>Nov 4–8</td>
<td>19464</td>
<td>1, 6.0036, 83</td>
<td>0.1113, 0.9985, 13.6595</td>
</tr>
<tr>
<td>17</td>
<td>Nov 11–15</td>
<td>17383</td>
<td>1, 6.7255, 151</td>
<td>0.0888, 0.9998, 20.0673</td>
</tr>
<tr>
<td>18</td>
<td>Nov 18–22</td>
<td>20317</td>
<td>1, 5.7546, 93</td>
<td>0.1164, 1.0005, 17.5540</td>
</tr>
<tr>
<td>19</td>
<td>Dec 2–6</td>
<td>20811</td>
<td>1, 5.6180, 142</td>
<td>0.1253, 1.0003, 21.4300</td>
</tr>
<tr>
<td>20</td>
<td>Dec 9–13</td>
<td>19248</td>
<td>1, 6.0743, 81</td>
<td>0.1135, 1.0004, 13.1855</td>
</tr>
<tr>
<td>21</td>
<td>Dec 16–20</td>
<td>18818</td>
<td>1, 6.2150, 82</td>
<td>0.0951, 0.9999, 12.1802</td>
</tr>
</tbody>
</table>

Table 1: Statistics of the durations. “Transformed durations” refers to the diurnally adjusted durations.
<table>
<thead>
<tr>
<th>Week</th>
<th>Omega estimate (stddev)</th>
<th>Alpha estimate (stddev)</th>
<th>Beta estimate (stddev)</th>
<th>Value of the log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2526 (0.0304)</td>
<td>0.1681 (0.0130)</td>
<td>0.5811 (0.0396)</td>
<td>-17573.1365</td>
</tr>
<tr>
<td>2</td>
<td>0.1652 (0.0162)</td>
<td>0.1119 (0.0076)</td>
<td>0.7230 (0.0207)</td>
<td>-17888.7339</td>
</tr>
<tr>
<td>3</td>
<td>0.1333 (0.0143)</td>
<td>0.1112 (0.0073)</td>
<td>0.7557 (0.0194)</td>
<td>-17945.3494</td>
</tr>
<tr>
<td>4</td>
<td>0.1573 (0.0183)</td>
<td>0.0743 (0.0064)</td>
<td>0.7686 (0.0220)</td>
<td>-16992.9188</td>
</tr>
<tr>
<td>5</td>
<td>0.1800 (0.0210)</td>
<td>0.0767 (0.0065)</td>
<td>0.7436 (0.0248)</td>
<td>-18495.7041</td>
</tr>
<tr>
<td>6</td>
<td>0.1396 (0.0166)</td>
<td>0.0774 (0.0064)</td>
<td>0.7832 (0.0205)</td>
<td>-17415.2212</td>
</tr>
<tr>
<td>7</td>
<td>0.1415 (0.0153)</td>
<td>0.0807 (0.0064)</td>
<td>0.7780 (0.0187)</td>
<td>-16783.2353</td>
</tr>
<tr>
<td>8</td>
<td>0.1662 (0.0220)</td>
<td>0.0679 (0.0067)</td>
<td>0.7661 (0.0254)</td>
<td>-15098.1953</td>
</tr>
<tr>
<td>9</td>
<td>0.1327 (0.0229)</td>
<td>0.0741 (0.0077)</td>
<td>0.7931 (0.0288)</td>
<td>-17222.3036</td>
</tr>
<tr>
<td>10</td>
<td>0.1235 (0.0152)</td>
<td>0.0889 (0.0071)</td>
<td>0.7878 (0.0197)</td>
<td>-16059.7855</td>
</tr>
<tr>
<td>11</td>
<td>0.1060 (0.0137)</td>
<td>0.0858 (0.0072)</td>
<td>0.8085 (0.0187)</td>
<td>-17509.7025</td>
</tr>
<tr>
<td>12</td>
<td>0.1610 (0.0154)</td>
<td>0.1026 (0.0073)</td>
<td>0.7367 (0.0193)</td>
<td>-16847.6630</td>
</tr>
<tr>
<td>13</td>
<td>0.1337 (0.0147)</td>
<td>0.0823 (0.0065)</td>
<td>0.7841 (0.0182)</td>
<td>-16957.2471</td>
</tr>
<tr>
<td>14</td>
<td>0.1476 (0.0161)</td>
<td>0.0974 (0.0073)</td>
<td>0.7553 (0.0209)</td>
<td>-18287.0412</td>
</tr>
<tr>
<td>15</td>
<td>0.1634 (0.0211)</td>
<td>0.0929 (0.0076)</td>
<td>0.7439 (0.0264)</td>
<td>-18534.1774</td>
</tr>
<tr>
<td>16</td>
<td>0.1528 (0.0221)</td>
<td>0.0741 (0.0068)</td>
<td>0.7730 (0.0269)</td>
<td>-19295.7451</td>
</tr>
<tr>
<td>17</td>
<td>0.1898 (0.0246)</td>
<td>0.0977 (0.0086)</td>
<td>0.7130 (0.0299)</td>
<td>-17162.8802</td>
</tr>
<tr>
<td>18</td>
<td>0.1918 (0.0245)</td>
<td>0.0851 (0.0075)</td>
<td>0.7234 (0.0295)</td>
<td>-20199.7797</td>
</tr>
<tr>
<td>19</td>
<td>0.2095 (0.0229)</td>
<td>0.0865 (0.0073)</td>
<td>0.7042 (0.0271)</td>
<td>-20673.7455</td>
</tr>
<tr>
<td>20</td>
<td>0.1289 (0.0251)</td>
<td>0.0640 (0.0074)</td>
<td>0.8071 (0.0311)</td>
<td>-19128.6345</td>
</tr>
<tr>
<td>21</td>
<td>0.1612 (0.0278)</td>
<td>0.0781 (0.0080)</td>
<td>0.7608 (0.0339)</td>
<td>-18669.6529</td>
</tr>
</tbody>
</table>

Table 2: Results from the maximum likelihood estimation of the ACD(1,1) models with exp(1) errors.
Evaluating models of autoregressive conditional duration

<table>
<thead>
<tr>
<th>Week</th>
<th>ACD(1,1) vs ACD(2,1)</th>
<th>ACD(1,1) vs ACD(1,2)</th>
<th>No remaining ACD</th>
<th>No STACD K=1</th>
<th>No STACD K=2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(ordinary) (robust)</td>
<td>(ordinary) (robust)</td>
<td>(ordinary) (robust)</td>
<td>(ordinary) (robust)</td>
<td>(ordinary) (robust)</td>
</tr>
<tr>
<td>1</td>
<td>0.0749 0.0932</td>
<td>0.0928 0.1136</td>
<td>0.2982 0.2949</td>
<td>1 × 10⁻⁵</td>
<td>6 × 10⁻⁸</td>
</tr>
<tr>
<td>2</td>
<td>0.6507 0.6474</td>
<td>0.6047 0.5996</td>
<td>0.1500 0.1332</td>
<td>3 × 10⁻⁶</td>
<td>3 × 10⁻⁶</td>
</tr>
<tr>
<td>3</td>
<td>0.1920 0.1997</td>
<td>0.2304 0.2374</td>
<td>0.2152 0.2049</td>
<td>3 × 10⁻⁷</td>
<td>4 × 10⁻⁷</td>
</tr>
<tr>
<td>4</td>
<td>0.0037 0.0050</td>
<td>0.0013 0.0026</td>
<td>0.0057 0.0075</td>
<td>8 × 10⁻⁸</td>
<td>4 × 10⁻⁵</td>
</tr>
<tr>
<td>5</td>
<td>0.0108 0.0127</td>
<td>0.0099 0.0112</td>
<td>0.0117 0.0109</td>
<td>5 × 10⁻⁵</td>
<td>9 × 10⁻⁶</td>
</tr>
<tr>
<td>6</td>
<td>0.1177 0.1125</td>
<td>0.1647 0.1544</td>
<td>0.1675 0.1525</td>
<td>0.0020 0.0001</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0055 0.0046</td>
<td>0.0018 0.0016</td>
<td>0.0086 0.0065</td>
<td>0.0099 0.0005</td>
<td>0.0070 0.0039</td>
</tr>
<tr>
<td>8</td>
<td>0.0138 0.0139</td>
<td>0.0189 0.0167</td>
<td>0.0286 0.0303</td>
<td>0.1217 0.1202</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.3133 0.3003</td>
<td>0.3014 0.2834</td>
<td>0.3392 0.3198</td>
<td>0.0895 0.1004</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.4766 0.4804</td>
<td>0.4709 0.4758</td>
<td>0.2667 0.2699</td>
<td>0.0002 0.0012</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.0765 0.0965</td>
<td>0.0949 0.1152</td>
<td>0.0578 0.0956</td>
<td>0.0005 0.0009</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.0028 0.0028</td>
<td>0.0019 0.0018</td>
<td>0.0013 0.0007</td>
<td>0.0038 0.0012</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.0003 0.0004</td>
<td>0.0003 0.0003</td>
<td>8 × 10⁻⁸ 0.0001</td>
<td>0.0024 0.0018</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.0060 0.0066</td>
<td>0.0038 0.0046</td>
<td>0.0023 0.0032</td>
<td>0.0008 0.0011</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.1585 0.1581</td>
<td>0.0797 0.0772</td>
<td>0.1814 0.1760</td>
<td>6 × 10⁻⁷ 0.0005</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.1036 0.0962</td>
<td>0.1012 0.0926</td>
<td>0.0687 0.0577</td>
<td>6 × 10⁻⁷ 0.0001</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.0923 0.6942</td>
<td>0.7378 0.7384</td>
<td>0.8032 0.7945</td>
<td>6 × 10⁻⁷ 0.0005</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.0084 0.0130</td>
<td>0.0055 0.0084</td>
<td>0.0096 0.0146</td>
<td>3 × 10⁻⁷ 0.0001</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.0407 0.0397</td>
<td>0.0231 0.0235</td>
<td>0.0379 0.0479</td>
<td>6 × 10⁻⁷ 0.0002</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.9678 0.9679</td>
<td>0.8619 0.8607</td>
<td>0.7678 0.7661</td>
<td>4 × 10⁻⁷ 0.0002</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.8595 0.8587</td>
<td>0.6333 0.6392</td>
<td>0.8243 0.8280</td>
<td>0.0041 0.0033</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: p-values of the tests of ACD(1,1) models. The tests are the ones against ACD(2,1) and ACD(1,2) models, remaining ACD in the standardized durations, and STACD(1,1) model of orders 1 and 2. Two versions of each test are performed: the ordinary refers to the test done using the auxiliary regression, and the robust is the robustified version. The p-values less than 0.01 are shown in boldface.
<table>
<thead>
<tr>
<th>Week</th>
<th>No TVACD K=1 (robust)</th>
<th>No TVACD K=2 (robust)</th>
<th>Tests using intraday time</th>
<th>No TVACD K=1 (ordinary)</th>
<th>No TVACD K=2 (ordinary)</th>
<th>Tests using intraweek time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9800</td>
<td>0.0037</td>
<td>0.0009</td>
<td>0.1883</td>
<td>0.0386</td>
<td>0.3551</td>
</tr>
<tr>
<td>2</td>
<td>0.8940</td>
<td>0.5913</td>
<td>0.3899</td>
<td>0.5758</td>
<td>0.7698</td>
<td>0.4435</td>
</tr>
<tr>
<td>3</td>
<td>0.7564</td>
<td>0.2982</td>
<td>0.9691</td>
<td>0.1683</td>
<td>0.1168</td>
<td>0.5975</td>
</tr>
<tr>
<td>4</td>
<td>0.3564</td>
<td>0.0037</td>
<td>0.9691</td>
<td>0.3463</td>
<td>0.3774</td>
<td>0.7864</td>
</tr>
<tr>
<td>5</td>
<td>0.3215</td>
<td>0.0009</td>
<td>0.2823</td>
<td>0.3729</td>
<td>0.4872</td>
<td>0.6410</td>
</tr>
<tr>
<td>6</td>
<td>0.4771</td>
<td>0.0018</td>
<td>0.5900</td>
<td>0.5499</td>
<td>0.5800</td>
<td>0.5900</td>
</tr>
<tr>
<td>7</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>8</td>
<td>0.4618</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>9</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>10</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>11</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>12</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>13</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>14</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>15</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>16</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>17</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>18</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>19</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>20</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
<tr>
<td>21</td>
<td>0.4043</td>
<td>0.0014</td>
<td>0.3283</td>
<td>0.2789</td>
<td>0.3400</td>
<td>0.2789</td>
</tr>
</tbody>
</table>

Table 4: p-values of the tests of ACD(1,1) models (continued). The tests are the ones against TVACD(1,1) models of orders 1 and 2. Time is defined either as intraday time or as intraweek time. Two versions of each test are performed: the ordinary refers to the test done using the auxiliary regression, and the robust is the robustified version. The p-values less than 0.01 are shown in boldface.
Appendix A

In this appendix we prove Theorem 1. The proof of Theorem 2 is almost identical and is omitted.

Proof of Theorem 1. Let \( \theta = (\theta_1', \theta_2')' \) be the parameter vector and define the conditional quasi log-likelihood function for observation \( x_i \) as

\[
l_i(\theta) = -\frac{x_i}{\psi_i + \varphi_i} - \ln(\psi_i + \varphi_i).
\]  (50)

The partial derivatives of (50) with respect to \( \theta_1 \) and \( \theta_2 \) are

\[
\frac{\partial l_i(\theta)}{\partial \theta_1} = \frac{x_i}{\psi_i + \varphi_i} \left( \frac{\partial \psi_i}{\partial \theta} + \frac{\partial \varphi_i}{\partial \theta} \right) \left( \frac{1}{\psi_i + \varphi_i} - 1 \right)
\]

\[
\frac{\partial l_i(\theta)}{\partial \theta_2} = \frac{x_i}{\psi_i + \varphi_i} \left( \frac{\partial \varphi_i}{\partial \theta} \right) \left( \frac{1}{\psi_i + \varphi_i} - 1 \right).
\]

Letting \( \theta^0 = (\theta_1^0, \theta_2^0)' \) be the true (under the null hypothesis) parameter vector, the score for observation \( x_i \) evaluated at the true parameter values equals

\[
\frac{\partial l_i(\theta^0)}{\partial \theta} = \left[ \frac{x_i}{\psi_i(\theta^0_1) + \varphi_i(\theta_2^0_1, \theta_2^0_2)} \left( \frac{\partial \psi_i}{\partial \theta_1} \theta^0_1 + \frac{\partial \varphi_i}{\partial \theta_2} \theta^0_2 \right) \left( \frac{1}{\psi_i(\theta^0_1) + \varphi_i(\theta_2^0_1, \theta_2^0_2)} - 1 \right) \right]
\]

\[
\left[ \frac{1}{\psi_i(\theta^0_1) + \varphi_i(\theta_2^0_1, \theta_2^0_2)} \left( \frac{\partial \psi_i}{\partial \theta_1} \theta^0_1 + \frac{\partial \varphi_i}{\partial \theta_2} \theta^0_2 \right) \left( \frac{1}{\psi_i(\theta^0_1) + \varphi_i(\theta_2^0_1, \theta_2^0_2)} - 1 \right) \right]
\]

\[
= \left( \frac{x_i}{\psi_i(\theta^0_1)} - 1 \right) \left[ \frac{1}{\psi_i(\theta^0_1)} - \frac{\partial \psi_i(\theta^0_1)}{\partial \theta_1} \theta^0_1 \left( \frac{1}{\psi_i(\theta^0_1)} - 1 \right) \right] = c_i(\theta^0_1) \left[ \begin{bmatrix} a_i \theta^0_1 \\ b_i \theta^0_2 \end{bmatrix} \right] = c_i \left[ \begin{bmatrix} a_i^0 \\ b_i^0 \end{bmatrix} \right].
\]

Furthermore, let \( \hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')' \) be the vector of maximum likelihood estimates, estimated under the null. Then the score evaluated at the ML estimates is (note that the upper block of the score is now just a vector of zeros)

\[
\frac{\partial l(\hat{\theta})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial l_i(\hat{\theta})}{\partial \theta} = \left[ \sum_{i=1}^{n} c_i \left[ \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \right] \right] = \left[ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right].
\]

Under condition (v) of Theorem 1 in Engle (2000)

\[
n^{-1/2} \frac{\partial l(\theta)}{\partial \theta} = n^{-1/2} \sum_{i=1}^{n} \frac{\partial l_i(\theta)}{\partial \theta} \delta \left( 0, n^{-1} \sum_{i=1}^{n} E \left[ \frac{\partial l_i(\theta^0)}{\partial \theta} \frac{\partial l_i(\theta^0)}{\partial \theta^0} \right] \right).
\]

The expectation in this expression can be written as

\[
E \left[ \frac{\partial l_i(\theta^0)}{\partial \theta} \frac{\partial l_i(\theta^0)}{\partial \theta^0} \right] = E \left( c_i^0 \right)^2 \left[ \begin{bmatrix} a_i^0 \theta_i^0 \\ b_i^0 \theta_i^0 \end{bmatrix} \right] \left[ \begin{bmatrix} a_i^0 \theta_i^0 \\ b_i^0 \theta_i^0 \end{bmatrix} \right] = E \left( c_i^0 \right)^2 \cdot E \left[ \begin{bmatrix} a_i^0 b_i^0 \\ b_i^0 b_i^0 \end{bmatrix} \right]
\]

\[
= E \left( c_i^0 \right)^2 \cdot E \left[ \begin{bmatrix} a_i^0 b_i^0 \\ b_i^0 b_i^0 \end{bmatrix} \right].
\]
since \( c_1^0 = \frac{\hat{\epsilon}_1}{\psi_1(\theta_1^0)} - 1 = \epsilon_1 - 1 \) is independent of the other terms, which are measurable with respect to \( \mathcal{F}_{i-1} \). Furthermore, \( E\left(c_1^0\right)^2 = E\left(\frac{\hat{\epsilon}_1}{\psi_1(\theta_1^0)} - 1\right)^2 = \text{Var}(\epsilon_1) = 1 \), so that

\[
\frac{n^{-1/2} \partial l(\theta^0)}{\partial \theta} \xrightarrow{d} N\left(0, n^{-1} \sum_{i=1}^{n} E\left\{ \begin{bmatrix} a_i^0 & a_i^0 & b_i^0 & b_i^0 \end{bmatrix} \right\} \right).
\]

This implies that the quadratic form

\[
\left\{ n^{-1/2} \frac{\partial l(\theta^0)}{\partial \theta} \right\} \left\{ n^{-1} \sum_{i=1}^{n} E\left\{ \begin{bmatrix} a_i^0 & a_i^0 & b_i^0 & b_i^0 \end{bmatrix} \right\} \right\}^{-1} \left\{ n^{-1/2} \frac{\partial l(\theta^0)}{\partial \theta} \right\}
\]

has an asymptotic \( \chi^2 \) distribution with \( \text{dim} \theta_2 \) degrees of freedom.

Since \( \hat{\theta} \) is a consistent estimator of \( \theta^0 \) (Theorem 1, Engle (2000)) and

\[
n^{-1} \sum_{i=1}^{n} \begin{bmatrix} a_i(\hat{\theta}_1) a_i'(\hat{\theta}_1) & a_i(\hat{\theta}_1) b_i'(\hat{\theta}_1, \theta_2^0) & b_i(\hat{\theta}_1, \theta_2^0) a_i'(\hat{\theta}_1) & b_i(\hat{\theta}_1, \theta_2^0) b_i'(\hat{\theta}_1, \theta_2^0) \end{bmatrix} = n^{-1} \sum_{i=1}^{n} \begin{bmatrix} \hat{a}_i \hat{a}'_i & \hat{a}_i \hat{b}'_i \hat{b}_i \hat{b}'_i \end{bmatrix}
\]

is a consistent estimator of

\[
n^{-1} \sum_{i=1}^{n} E\left\{ \begin{bmatrix} a_i^0 & a_i^0 & b_i^0 & b_i^0 \end{bmatrix} \right\}
\]

(since the expressions are functions of \( \hat{\theta} \) and \( \theta^0 \), respectively), the LM statistic

\[
\left\{ n^{-1/2} \frac{\partial l(\hat{\theta})}{\partial \theta} \right\} \left\{ n^{-1} \sum_{i=1}^{n} \begin{bmatrix} \hat{a}_i \hat{a}'_i & \hat{a}_i \hat{b}'_i & \hat{b}_i \hat{a}'_i & \hat{b}_i \hat{b}'_i \end{bmatrix} \right\}^{-1} \left\{ n^{-1/2} \frac{\partial l(\hat{\theta})}{\partial \theta} \right\}
\]

\[
= \left\{ \begin{bmatrix} 0 \\ \sum_{i=1}^{n} \hat{c}_i \hat{b}'_i \end{bmatrix} \right\} \left\{ \begin{bmatrix} \sum_{i=1}^{n} \hat{a}_i \hat{a}'_i \\ \sum_{i=1}^{n} \hat{a}_i \hat{b}'_i \\ \sum_{i=1}^{n} \hat{b}_i \hat{a}'_i \\ \sum_{i=1}^{n} \hat{b}_i \hat{b}'_i \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} 0 \\ \sum_{i=1}^{n} \hat{c}_i \hat{b}'_i \end{bmatrix} \right\}
\]

also has an asymptotic \( \chi^2 \) distribution with \( \text{dim} \theta_2 \) degrees of freedom under the null hypothesis.
Appendix B

Here we show that the test statistic of Li and Yu (2003) is asymptotically equivalent to the one given in Corollary 5. Li and Yu (2003) only consider the case where the estimated model is an ACD\((m,0)\) model, and for ease of exposition they restrict the derivations to the case \(m = 1\). In this case \(\hat{a}_i\) in Corollary 5 reduces to \(\hat{a}_i = \hat{\psi}_i^{-1}(1, x_{i-1}')\). As usual, \(\hat{b}_i = \hat{\psi}_i^{-1}\left(x_{i-1}\hat{\psi}_{i-1}, \ldots, x_{i-m}\hat{\psi}_{i-m}\right)\) and \(\hat{c}_i = x_i\hat{\psi}_i^{-1} - 1\). The statistic of Li and Yu (2003) is

\[
Q = n\hat{r}' \left(I_{m^*} - \hat{X}G^{-1}\hat{X}'\right)^{-1} \hat{r}
\]

(51)

where \(I_{m^*}\) is the \(m^* \times m^*\) identity matrix, and \(\hat{r}\) is the following \(m^* \times 1\) vector:

\[
\hat{r} = n^{-1} \sum \left[ \left( \frac{x_i}{\hat{\psi}_i} - 1 \right) \left( \frac{x_{i-1}}{\hat{\psi}_{i-1}} - 1 \right) \right] \]

\[
= n^{-1} \sum \hat{c}_i \hat{\psi}_i \hat{b}_i - n^{-1} \sum \hat{c}_i.
\]

(52)

Furthermore, the \(m^* \times 2\) matrix \(\hat{X}\) is given by

\[
\hat{X} = n^{-1} \sum \left[ \begin{array}{c} \frac{x_i}{\hat{\psi}_i} \left( \frac{x_{i-1}}{\hat{\psi}_{i-1}} - 1 \right) \\ \vdots \\ \frac{x_i}{\hat{\psi}_i} \left( \frac{x_{i-m^*}}{\hat{\psi}_{i-m^*}} - 1 \right) \end{array} \right] = n^{-1} \sum \frac{x_i}{\hat{\psi}_i} \hat{b}_i \hat{a}_i' - n^{-1} \sum \frac{x_i}{\hat{\psi}_i} \hat{1}_{m^*} \hat{a}_i'
\]

(53)

where \(\hat{1}_{m^*}\) is an \(m^* \times 1\) vector of ones, and

\[
\hat{G} = n^{-2} \sum \left[ \begin{array}{cc} 2\frac{x_i}{\hat{\psi}_i} - \frac{1}{\hat{\psi}_i^2} & 2\frac{x_{i-1}}{\hat{\psi}_i} - \frac{x_i}{\hat{\psi}_i^2} \\ 2\frac{x_{i-1}}{\hat{\psi}_i} - \frac{x_i}{\hat{\psi}_i^2} & 2\frac{x_{i-2}}{\hat{\psi}_i} - \frac{x_{i-1}}{\hat{\psi}_i^2} \end{array} \right]
\]

\[
= n^{-2} \sum \left( 2\frac{x_i}{\hat{\psi}_i} - 1 \right) \hat{a}_i \hat{a}_i'.
\]

(54)

Now, (52), (53), and (54) converge in probability to the same quantities as \(n^{-1} \sum \hat{c}_i \hat{\psi}_i \hat{b}_i, n^{-1} \sum \hat{\psi}_i \hat{b}_i \hat{a}_i',\) and \(n^{-2} \sum \hat{a}_i \hat{a}_i',\) respectively, as \(n \to \infty\). Furthermore, \(n^{-1} \sum \hat{\psi}_i^2 \hat{b}_i \hat{b}_i'\) converges in probability to \(I_{m^*}\). It follows that the statistic (51) and the one given in Corollary 5 are asymptotically equivalent.
Appendix C

In this appendix we provide a description of the generalized spectral test introduced in Hong (1999) and suggested as a misspecification test for ACD models in Hong and Lee (2003). We present the test in the context of using it as a misspecification test for the estimated standardized durations of an ACD model. Denoting the standardized duration series by $\varepsilon_i$, consider the covariance between the empirical characteristic functions of $\varepsilon_i$ and $\varepsilon_{i-j}$,

$$\sigma_j(u, v) = \text{cov} (e^{iue_i}, e^{ive_{i-j}}) \quad (55)$$

where $i = \sqrt{-1}$ and $j = 0, \pm 1, \ldots$. Hong (1999) calls the Fourier transform of (55) (it exists under some regularity conditions)

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v)e^{-ij\omega}$$

the generalized spectral density function of $\{\varepsilon_i\}$. This is estimated using the following kernel estimator:

$$\hat{f}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{1/2}k(j/p)\hat{\sigma}_j(u, v)e^{-ij\omega} \quad (56)$$

where $k(\cdot)$ is a kernel function, $p$ is a bandwidth, $n$ is the sample size, and $\hat{\sigma}_j(\cdot, \cdot)$ is a consistent estimator for the covariance of the two empirical characteristic functions. The generalized spectral test is defined as the (suitably weighted and standardized) $L_2$ norm of the difference of the estimate of $f$ and the estimate of $f$ under the null hypothesis of serial independence of the standardized duration series. For an exact formulation of the test statistic we refer to Hong (1999) or Hong and Lee (2003).

Using the generalized spectral test involves, among other things, the choice of the bandwidth $p$ and the kernel function $k(\cdot)$. For the bandwidth $p$, Hong (1999) discusses a data-driven method for choosing an optimal bandwidth (in the sense of an integrated mean squared error criterion for the estimator $\hat{f}$ in equation (56)) given a preliminary bandwidth $\bar{p}$. The choice of $\bar{p}$ remains somewhat arbitrary, but the simulation studies of Hong (1999) and Hong and Lee (2003) suggest that the test statistic is quite robust to the choice of $\bar{p}$. We choose the values 25, 50, 75, and 100 for the preliminary bandwidth.

For the kernel function, Hong (1999) shows that the Daniell kernel is optimal in the sense that it maximizes the asymptotic power of the test over a class of kernel functions. The fact that the Daniell kernel has an unbounded support implies that when using the estimator (56), the covariance $\hat{\sigma}_j(u, v)$ has to be computed for all $j$ between 1 and $n - 1$. For ultra-high-frequency data, where sample sizes are quite large, this becomes computationally demanding. For this reason we choose to use the Parzen kernel that has a bounded support. This considerably reduces the time needed for computing the value of the statistic. According to the simulation results of Hong (1999) this should only have a minor effect on the power of the test.
References


REFERENCES


Hansen, B. E. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” Econometrica, 64, 413–430.


On the econometrics of multivariate marked ultra-high-frequency data
On the econometrics of multivariate marked ultra-high-frequency data

Abstract

In this paper we consider the modelling of multivariate marked ultra-high-frequency data. Such data consists, for example, of the trading times of several stocks together with the associated prices. We give an example how this kind of data may be modelled in the framework of marked point processes. The proposed model is a multivariate model for the trading times and the associated returns. We suggest a simple generalization of existing GARCH-type models for univariate ultra-high-frequency data to the multivariate setting. Preliminary remarks concerning the stationarity of the proposed model are given, and maximum likelihood estimation is discussed.

This paper is joint work with Luc Bauwens.
We thank Clive Bowsher, Nikolaus Hautsch, Markku Lanne, Annastiina Silvennoinen, Timo Teräsvirta, and David Veredas for useful comments and suggestions.
1 Introduction

Transaction-by-transaction data has become widely available in financial markets during the past decade or so. This data typically contains information on the precise time of a trade of an asset together with some associated characteristics, say the trading price and volume. The availability of such data has given rise to extensive research both in theoretical market microstructure as well as in econometric modelling of transaction-by-transaction, or ultra-high-frequency, data; see for example Engle (2000). The econometric approaches proposed have been models for the trading time process itself as well as for the trading times and the associated information, say prices. The latter are motivated by the finding in recent market microstructure literature that the timing of trades is of importance for the related price process; see for example Easley and O’Hara (1992) and Easley, Kiefer, and O’Hara (1997).

When the econometric modelling of univariate trading time processes is the objective, three main approaches have emerged. The first of them is concerned with modelling durations, i.e. the time intervals between subsequent trades. This approach was pioneered by Engle and Russell (1998), who introduced the autoregressive conditional duration, or ACD, model, and ever since much research has been devoted to this area. In the second approach, the number of trades in subsequent intervals of fixed length is the object of interest. These count data models have been discussed in Heinen (2003). The third alternative is a continuous time approach where the conditional intensity of observing a trade is modelled, and was first proposed in Hamilton and Jordà (2002).

Proposals for a joint model for the trading time process and the return volatility have been introduced by Engle (2000), Ghysels and Jasiak (1998), and Grammig and Wellner (2002) (see also Meddahi, Renault, and Werker (2006) and Renault and Werker (2005), which contain discussion and comparison of these approaches). In each of these papers, a univariate GARCH-type model for irregularly spaced data is devised. The motivation for a model of this kind is to investigate the relationships between the transaction rate and volatility (see Easley and O’Hara (1992) and Easley, Kiefer, and O’Hara (1997)). The conclusion in these papers is that these relationships are significant. Darolles, Gouriéroux, and Le Fol (2000) also consider the joint modelling of trading times and prices. Rydberg and Shephard (2000) and Rydberg and Shephard (2003) suggest how to model the movements in the price process. Bauwens and Giot (2003) consider the trading time process together with the direction of the price changes. In Russell and Engle (2005) the trading times and discrete price movements are modelled, whereas Manganelli (2005) models the timings, returns, and volumes simultaneously.

Multivariate extensions to models of the trading time process have been considerably more difficult to formulate. The main reason for this is that since trades with different stocks do not occur simultaneously, accounting for the arriving information becomes complicated. This has been the case, in particular, for the approach based on durations as accommodating for the arrival of intraduration information is hard. In the count data approach this problem is circumvented because the data is aggregated, and a multivariate generalization has been introduced in Heinen and Rengifo (2003). The approach based on modelling intensities, however, seems to be
the most suitable one for modelling multivariate data because it is based on a continuous time framework and therefore the asynchronous timing of trades of different stocks poses no great difficulties. Multivariate intensity models for event times have been introduced in Russell (1999) (the autoregressive conditional intensity, or ACI, model), Bauwens and Hautsch (2003, 2005) (dynamic latent factor intensity models), and Bowsher (2003). Russell (1999) applies his model to the bivariate series of transaction times and limit order submission times, Bauwens and Hautsch (2005) consider the price intensities of five stocks, and Bowsher (2003) examines the bivariate series of trading times and the times when there is a change in the midquote price. Other related papers are Hall and Hautsch (2004) and Hall and Hautsch (2006), which are applied papers using multivariate intensity models, Blaszek and Escribano (2005), who propose and apply a model similar to the one in Bauwens and Hautsch (2003) to patent citation data, Koopman, Lucas, and Monteiro (2005), who extend the model of Bauwens and Hautsch (2003), as well as Large (2005) who uses the model of Bowsher (2003) to examine the resiliency of an electronic limit order book.

In the case of usual, regularly spaced financial data, there is ample evidence that the volatilities of different stocks are interdependent. This is probably best manifested by the vast amount of research done in the area of multivariate GARCH (and stochastic volatility) models; for a review of this literature, see eg. Bauwens, Laurent, and Rombouts (2006).

Given this interest in modelling univariate and multivariate trading time processes, the relationship between the trading times and return volatilities, and multivariate volatility (in the case of regularly spaced data), a natural extension is to model the multivariate trading time process jointly with the volatilities. This extension is pointed out by Engle (2002, p. 427), who notes that the arrival rate of trades may be of importance for volatility at the ultra-high-frequency, and that solutions for modelling multivariate volatility at this frequency must be found. Furthermore, in related literature on realized volatility, Barndorff-Nielsen and Shephard (in press) emphasize the importance of the multivariate case when investigating volatility in high-frequency data (Bandi and Russell (2005) is a rare paper addressing this issue). However, in the realized volatility literature only a fraction of the available tick-by-tick data is used, a practice criticized by Ait-Sahalia, Mykland, and Zhang (2005) who argue that it would be better to use all the data available. In the present paper we follow this principle of making full use of the data at our disposal and, in particular, incorporate the precise timings of the financial events, an approach whose importance is also emphasized by Engle and Sun (2005).

An economic argument favouring this approach is to gain better measures of volatility using ultra-high-frequency data. In related work, Bandi, Russell, and Zhu (2005) use high-frequency data to obtain estimates of volatilities and co-volatilities, use these in dynamic portfolio allocation, and conclude that the gains are economically large and statistically significant. In another study, Bauwens, Ben Omrane, and Rengifo (2006) consider portfolio allocation of currencies using intraday measures of volatility, and conclude that a multivariate approach improves the results of portfolio allocation compared to using a univariate model. In that paper, returns sampled at regular intervals (30 minutes) are used, and a model which would allow
to estimate the volatilities on a more continuous time scale may bring additional flexibility in the management of the portfolio. Another motivation for our approach is the investigation of the impact of news announcements during trading sessions (for stock exchanges) or trading days (for the foreign exchange markets) on the volatility of the returns of the related assets. For example, the impact of announcements of central bank interventions on the level and the volatility of exchange rates may be measured more precisely at the ultra-high-frequency level than at a more coarse level. Bauwens, Ben Omrane, and Giot (2005) investigate this issue using returns sampled regularly (every 5 minutes), and it would be interesting to use models where the precise timing of the announcements can be incorporated. Again, news announcements typically affect several assets simultaneously so that the multivariate aspect cannot be neglected.

The purpose of this paper is therefore to make a first attempt at proposing how to model multivariate, marked, ultra-high-frequency data. In particular, we consider modelling the trading times and return volatilities jointly in a multivariate setting. We present a framework within which such modelling is possible, and give examples of simple models that can be used. Formulating more complex (and maybe more realistic) models is left for future research.

For the present purpose, multivariate ultra-high-frequency data consists of information about the precise trading times as well as some associated information, which are called ‘marks’. These kinds of observations can be regarded as a realization of a temporal-spatial point process. The temporal part consists of the locations of the points (trades) in time, whereas the spatial part (or mark) can be seen as describing the type of the point (which of the stocks traded) as well as the other characteristics associated with the trade, say the return. The theory of marked point processes is well-suited for this kind of situation, where we wish to describe both the probability distribution determining the locations of the points (times of the trades), and the probability distribution of the marks (type and return) at these points.

In this paper we show how the theory of marked point processes can be applied to the problem at hand. Marked point processes have been used to model ultra-high-frequency data already in earlier literature, but the introduction of marks in the multivariate setting appears to be a novel contribution. We give examples of how the trading time process can be modelled using the intensity function, and formulate a simple multivariate model for the return volatilities. The resulting combined model can be estimated with the usual quasi-maximum likelihood approach making the empirical implementation feasible.

This paper is organized as follows. The next section discusses some general theory of marked point processes. In Section 3 modelling the trading time process is treated, whereas Section 4 gives an example how to model which one of the stocks traded. Section 5 discusses modelling the returns and suggests a simple multivariate model for the associated conditional volatilities. Strict stationarity is the topic of Section 6, and maximum likelihood estimation is discussed in Section 7. Preliminary notes concerning some empirical issues are presented in Section 8. Section 9 concludes.

---

\( ^{9}\)We note that preliminary results of ongoing and coincidental research with a goal similar to ours are reported in Blazsek and Kamionka (2005).
Last but not least, we wish to emphasize that this manuscript presents the current status of work-in-progress, and does not constitute a finished research paper. The ideas described here will be refined and developed in further work.

2 Marked point processes

In this paper, we conform to the notation and terminology used in Daley and Vere-Jones (2003), and the reader is referred to this monograph for more details on point processes. On some occasions reference is also made to Brémaud (1981), in which a more theoretical treatment can be found. We assume some familiarity with point processes and keep the treatment brief. Useful background reading, in addition to the ones just mentioned, are Matthes, Kerstan, and Mecke (1978), Cox and Isham (1980), Karr (1986), Daley and Vere-Jones (1988), and Last and Brandt (1995) among others.

We consider a set of $S \geq 2$ stocks. Let the continuous time be denoted by $t \geq 0$, and the times of the trades (with any of the stocks) by $t_i$ ($i = 1, 2, \ldots$). Also, let $t_{si}$ denote the times of trades for stock $s$ ($s = 1, \ldots, S$, $i = 1, 2, \ldots$). Assuming that no two stocks can trade simultaneously, the series of points $\{t_i\}_{i=1}^{\infty}$ is the union of the points in the series $\{t_{si}\}_{i=1}^{\infty}$ ($s = 1, \ldots, S$). Our first objective is to describe the structure of the times of trades $0 < t_1 < \ldots < t_i < \ldots$, which form a sequence of positive and strictly increasing random variables. Following the terminology of Daley and Vere-Jones (2003), these times form a ground point process denoting the times of trades but carrying no information of which of the stocks traded nor the price associated with the trade. This ground point process can be defined through its counting function $N_g$:  

\[ N_g(A) = \sum_{i=1}^{\infty} 1(t_i \in A) \quad \text{for all } A \in \mathcal{B}(\mathbb{R}_+) \]

where $1(\cdot)$ is the indicator function and $\mathcal{B}(\mathbb{R}_+)$ denotes the Borel sets on the positive real line. In other words, $N_g(A)$ is the cardinal of the set $A \cap \{t_i\}_{i=1}^{\infty}$. In particular, if $A = (0, t]$, we also write

\[ N(t) \equiv N_g((0, t]) = \sum_{i=1}^{\infty} 1(t_i \leq t). \]

A counting function may also be defined for the times of trades for stock $s$, and in particular we write

\[ N_s(t) = \sum_{i=1}^{\infty} 1(t_{si} \leq t) \quad (s = 1, \ldots, S). \]

Furthermore, we write $N(t-)$ and $N_s(t-)$ for the number of trades strictly before time $t$, i.e. $N(t-) = N_g((0, t))$ and similarly for $N_s(t-)$. Note that $N(t-) = N(t) - 1$ if there is a trade at time $t$ and $N(t-) = N(t)$ otherwise and similarly for the $N_s(\cdot)$.

---

*We use the term ‘ground’ process and the subindex ‘g’ for non-marked point processes to differentiate them from marked point processes.*
A ground point process can also be defined through its intensity function \( \lambda_g(t) \), which has an interpretation as the instantaneous random rate of trading at time \( t \) given the past, i.e.

\[
E[N_g((t, t + dt))] \mid \mathcal{F}_{g,t} = \lambda_g(t) \, dt,
\]

where \( \mathcal{F}_{g,t} \) denotes the \( \sigma \)-algebra of all information of the ground point process up to and including time \( t \). Conditions that the function \( \lambda_g(t) \) needs to satisfy for it to define a well-defined ground point process are given in Assumption 1 in the next section.

Given these preliminaries a marked point process, where we assume the locations of the points to be on the positive real axis, may be defined as follows.

**Definition 1** A marked point process, with locations in \( \mathbb{R}_+ \) and marks in the complete separable metric space \( \mathcal{M} \), is a point process \( \{(t_i, m_i)\} \) on \( \mathbb{R}_+ \times \mathcal{M} \) with the additional property that the ground process \( N_g(\cdot) \), consisting of the locations of the process, is itself a point process (i.e. for bounded \( A \in \mathcal{B}(\mathbb{R}_+) \), \( N_g(A) = \sum_{i=1}^{\infty} 1((t_i, m_i) \in A \times \mathcal{M}) < \infty \)).

In the applications we consider, the mark space \( \mathcal{M} \) consists of the type of the trade (i.e. which of the stocks traded) and the associated return. These are denoted by the random variables \( s_i \) and \( r_i \) taking values in \( \mathcal{M}_s = \{1, \ldots, S\} \) and \( \mathcal{M}_r = \mathbb{R} \), respectively. Hence \( m_i = (s_i, r_i) \) and \( \mathcal{M} = \mathcal{M}_s \times \mathcal{M}_r \), which clearly is a complete separable metric space. Other interesting marks that could be considered include volume (taking values in \( \mathbb{Z}_+ \) or \( \mathbb{R}_+ \)) and an indicator variable carrying the information whether the trade was initiated by a buyer or a seller (taking values in \( \{0, 1\} \)). Furthermore, let \( \mathcal{F}_t, 0 \leq t < \infty \), denote the \( \sigma \)-algebra of all the information of the marked point process up to and including time \( t \), and \( \mathcal{F}_{t-} = \lim \sup_{s \downarrow t} \mathcal{F}_s \) the information immediately before time \( t \).

### 3 Modelling the ground process

To completely characterize a marked point process we need to define a probability distribution for the locations of the points as well as for the marks associated with these locations. Concerning the locations of the points, the ground process may be defined through its intensity function as mentioned above. We assume that this intensity function, \( \lambda_g(t) \), can be written as \( \lambda_g(t) = \sum_{s=1}^{S} \lambda_s(t) \), and that the \( S \times 1 \) vector \( \lambda(t) = (\lambda^1(t), \ldots, \lambda^S(t))' \) satisfies the following assumption.\(^{11}\)

**Assumption 1**

(i) The \( S \times 1 \) vector \( \lambda(t) = (\lambda^1(t), \ldots, \lambda^S(t))' \) consists of real valued processes that are nonnegative, left-continuous, adapted to \( \mathcal{F}_t \), and such that for all \( s = 1, \ldots, S \)


\(^{11}\)Ref. Def. II.D7 of Brémaud (1981) as well as I.T5, I.E11 for the first part and Def. 7.1.11 of Daley and Vere-Jones (2003) for the second part.
On the econometrics of multivariate marked ultra-high-frequency data \[ \text{and all } t \geq 0 \]
\[ \int_{(0,t)} \lambda^s(u) du < \infty. \]  
\[ (1) \]

(ii) The ground process, defined by the intensity \( \lambda_g(t) = \sum_{s=1}^{S} \lambda^s(t) \), is regular.

Assumption (i) ensures that the intensity function corresponds to a well-defined ground point process, while (ii) is needed to operationalize some of the results in Daley and Vere-Jones (2003) used below. Requiring nonnegativity of the \( \lambda^s \)-processes is natural as they, suitably rescaled, correspond to certain probabilities. The left-continuity and adaptedness to \( \mathcal{F}_t \) ensure that the processes are \( \mathcal{F}_t \)-progressive, which is required for intensity functions of point processes. Finally, the integrability condition (1) ensures that the ground process is non-explosive, i.e. that for bounded \( A \in \mathcal{B}(\mathbb{R}_+) \), \( N_g(A) < \infty \). We also note that assumption (ii) implies that the point process is ‘simple’, which rules out simultaneous occurrences of events for the ground process.\(^{12}\) The following two subsections give examples how the \( \lambda^s \)-processes may be modelled.

3.1 Multivariate nonlinear Hawkes processes

A class of point processes that has received attention in the financial literature is that of Hawkes processes, originally introduced in Hawkes (1971). Perhaps the first ones to draw attention to them were Engle and Russell (1998) and Russell (1999) who, however, questioned their usefulness in the context of transaction data due to the fact that in a Hawkes process, the marginal contribution of a trade in the past to the current intensity is independent of the number of intervening trades. However, as we are interested in how the previous trades, or maybe just some of the most recent trades, together affect the current intensity, it is not clear whether this is a significant drawback. Due to their structure (see below), the Hawkes processes also offer both a flexible class of functional forms for the intensities as well as a (relative) ease of estimation due to an inherent (log-) linear response to time. A further advantage is that due to their use in other applications of the point process literature, there exists a large set of results about their properties.

Bauwens and Hautsch (2003) used Hawkes processes in formulating their multivariate intensity model, whereas Bowsher (2003) introduced a generalization of them taking into account the non-trading periods between days. In both papers they were found to provide a reasonable approximation to the trading time process. In this paper, we consider a multivariate and nonlinear generalization of them given in Brémaud and Massoulié (1996) (see also Daley and Vere-Jones (2003, pp. 252–253)). Define the ‘\( s \)-type intensity processes’ by

\[ \lambda^s(t) = \phi^s \left( \sum_{i=1}^{S} \int_{(0,t)} h_{si}(t-u) N^i(du) \right) = \phi^s \left( \sum_{i=1}^{S} \sum_{j: t_{ij} < t} h_{si} (t - t_{ij}) \right) \]  
\[ (2) \]

\(^{12}\)See Proposition 5.4.V and Def. 3.3.II of Daley and Vere-Jones (2003).
where \( \phi^s : \mathbb{R} \rightarrow \mathbb{R}_+ \) and \( h_{si} : \mathbb{R}_+ \rightarrow \mathbb{R} \) \( (s, i = 1, \ldots, S) \) and the form of the functions \( \phi^s \) and \( h_{si} \) determines how the intensity processes are affected by the arrivals of trades. Note that the \( \lambda^s(t) \)'s are left-continuous, i.e. only points before time \( t \) are included in the summations. Concerning the functions \( h_{si} \), one common form for them is the one of exponential decay, say

\[
h_{si}(u) = \sum_{k=1}^{K} \alpha_{sik} e^{-\beta_{sik} u} \quad (u > 0)
\]

where the \( \alpha_{sik} \) and \( \beta_{sik} \) \( (s, i = 1, \ldots, S, k = 1, \ldots, K) \) are positive parameters. One major advantage of this choice is that deriving analytical results becomes much simpler than for other functional forms. These analytical results are also of considerable practical importance since they can be used to make estimation more computationally efficient, see Section 7. Concerning the functions \( \phi^s \), they may be rather general non-linear functions (satisfying some boundedness and continuity conditions). Choosing \( \phi^s(x) = \omega_s + x \), where \( \omega_s \) is a positive constant, brings us to the usual multivariate (linear) Hawkes process. In case of the linear Hawkes process with exponential decay the ‘s–type intensity processes’ thus take the form

\[
\lambda^s(t) = \omega_s + \sum_{i=1}^{S} \sum_{j : t_i^j < t} \sum_{k=1}^{K} \alpha_{sik} e^{-\beta_{sik} (t - t_i^j)}. \tag{3}
\]

The \( \alpha \)'s can be interpreted as measuring how much a trade increases the intensities for the stocks, and the \( \beta \)'s as measuring how quickly the intensities decay after the trades. The summation over \( i \) indicates that trades on all the stocks may affect the intensity of stock \( s \). If \( \alpha_{sik} = 0 \) for \( s \neq i \), the intensities are independent of the other stocks’ behaviour. In the middle summation, all trades on all stocks occurring strictly before time \( t \) are included, and are thus contributing to the intensity. The summation over \( k \) allows each trade to have an effect as a superposition of exponential decays. We note that in the case of \( K > 1 \), an identifying restriction is needed, and this restriction may be chosen to be that for each \( s \) and \( i \), \( \alpha_{s1} < \cdots < \alpha_{sK} \).

### 3.2 Autoregressive conditional intensity model

An alternative to the Hawkes processes is the Autoregressive Conditional Intensity (ACI) model introduced in Russell (1999). In this model the \( S \times 1 \) vector of intensities, \( \lambda(t) = (\lambda^1(t), \ldots, \lambda^S(t))^t \), is decomposed into three components as \( \ln \lambda(t) = \ln \Omega(t) + \ln \Psi(t) + \ln C(t) \). The first two terms are defined to be piecewise constant, changing their values only at the times of trades, and being left-continuous at these points. The last term is a deterministic component that is allowed to change its values continuously with time and may be used, for example, to capture diurnal variation within a trading day (which is discussed also in Section 8).

For \( \Omega(t) \) and \( \Psi(t) \) it is thus sufficient to define their values at times of trades \( t_{N(t)} \). To this end, first introduce the notation \( \mathbf{1}_{N(t)} \), which is an \( S \times 1 \) indicator vector with 1 in the element for which the \( N(t) \)-th trade occurred and zeros elsewhere. The term
\( \ln \Omega(t_{N(t)}) \) is the (logarithm) of the ‘baseline intensity’ (see Russell (1999)) and is defined as

\[
\ln \Omega(t_{N(t)}) = \Omega 1_{N(t)-1}
\]

where \( \Omega \) is an \( S \times S \) matrix. Thus the baseline intensity is assumed to depend on the identity of the stock for which the last trade occurred. Note that \( \ln \Omega(t) \) takes the value \( \ln \Omega(t_{N(t)}) \) for \( t \in (t_{N(t)-1}, t_{N(t)}) \) as it is left-continuous. The term \( \ln \Psi(t_{N(t)}) \) captures the dynamic structure of the intensities and is modeled as

\[
\ln \Psi(t_{N(t)}) = A \Lambda(t)_{N(t)-1} + B \cdot \ln \Psi(t_{N(t)-1})
\]

where \( A \) and \( B \) are \( S \times S \) matrices and \( \Lambda(t) = (\Lambda^1(t), \ldots, \Lambda^S(t))' \), and

\[
\Lambda_s(t) = \int_{t_{N(t)-1}}^{t_{N(t)}} \lambda_s(u) du, \quad s = 1, \ldots, S.
\]

That is, \( \Lambda_s(t_{N(t)-1}) \) is the integrated ‘\( s \)–type intensity’ over the most recent observed duration for stock \( s \). The integrated intensities for stock \( s \) form an iid sequence of \( \exp(1) \)–distributed random variables (see Russell (1999) and also Bowsher (2003, Cor. 4.2)).

4 Modelling the ‘type’ of trade

When the mark distribution is concerned, it is natural to specify the distribution of type-mark and return-mark separately. We can decompose the probability density/mass function of the marks as

\[
f^{(m)}(s_{N(t)} = s, r | F_{t-}) = f^{(t)}(s_{N(t)} = s | F_{t-}) \times f^{(r)}(r | F_{t-}, s)
\]

where the superscripts \((m)\), \((t)\), and \((r)\) refer to mark, type, and return, respectively. In writing down this decomposition we have chosen to model the marginal distribution of the type-mark and the distribution of the returns conditional on the type-mark. We first consider the distribution of the type-mark, and then define a distribution for the returns conditional on the type-mark.

First note that it only makes sense to consider the mark distribution at those time points where there actually is a trade as the marks are inherently associated with the points of the process. Hence, when discussing the mark distribution at time \( t \), the conditioning information is always assumed to be \( F_{t-} \) and the knowledge that there is a trade at time \( t \) with one of the stocks (but not necessarily the knowledge which of the stocks traded).

Given that there is a point of the ground process at time \( t = t_i \), we let the type-mark distribution be a multinomial distribution with the probabilities defined according to the respective intensities at the time of the trade. That is,

\[
P[\text{The trade at } t = t_i \text{ is with stock } s | F_{t_i-}] = \frac{\lambda^s(t_i)}{\lambda^s(t_i) + \lambda^g(t_i)}
\]  \hspace{1cm} (4)
or, in other words, the discrete probability mass function for the type of trade is for all $t > 0$ defined by

$$f^{(t)}(s_{N(t)} = s \mid \mathcal{F}_{t-}) = \frac{\lambda_s(t)}{\lambda_g(t)}, \quad s = 1, \ldots, S.$$  

Note that $\sum_{s=1}^{S} f^{(t)}(s_{N(t)} = s \mid \mathcal{F}_{t-}) = 1$, so that this discrete probability distribution is well-defined. Now the ‘conditional intensity of the $s$–type process’ is

$$\lambda(t, s_{N(t)} = s \mid \mathcal{F}_{t-}) = \lambda_g(t \mid \mathcal{F}_{t-}) f^{(t)}(s_{N(t)} = s \mid \mathcal{F}_{t-}) = \lambda_s(t)$$

which gives the $\lambda_s(t)$’s an interpretation.  

It may be mentioned that in previous literature on multivariate intensity models (Russell (1999), Bauwens and Hautsch (2003), and Bowsher (2003)) the distribution of the type-mark has been implicitly assumed to be the one defined above.  

5 Modelling the returns

Engle (2000), Ghysels and Jasiak (1998), and Grammig and Wellner (2002) make proposals of how to model the returns distribution (i.e. $f^{(r)}$) in the case of univariate irregularly spaced transactions data. A detailed discussion and comparison of these approaches can be found in Meddahi, Renault, and Werker (2006) and Renault and Werker (2005). The conclusion of the comparison is in favor of the approach in Engle (2000). Motivated by this, we opt to follow Engle (2000) for the formulation of our model and present a simple generalization of his model to the multivariate setting.

Let us first set some notation. As the returns are always associated with trades they are most conveniently indexed using the counting functions defined earlier (instead of indexing them according to continuous time). Let $y^s_{N(t)}$ be the most recent return at time $t$ on asset $s$, i.e. $y^s_{N(t)} = 100 \left( \log(p^s_{N(t)}) - \log(p^s_{N(t)-1}) \right)$, where $p^s_{N(t)}$ is the most recently recorded price of stock $s$ up until time $t$. Suitable measures of price are discussed shortly. Furthermore, let $x^s_{N(t)} = t^s_{N(t)} - t^s_{N(t)-1}$ be the most recently observed duration for stock $s$, and define $\tilde{y}^s_{N(t)} = y^s_{N(t)} (x^s_{N(t)})^{-1/2}$ as the corresponding return per square root of the most recent duration. Since return volatility is usually measured over certain time intervals, $\tilde{y}^s_{N(t)}$ is the natural object of interest as its conditional moments have the interpretation as conditional moments per unit of time. Also note that as the conditioning includes the information $s_{N(t)} = s, t^s_{N(t)} = t$ and hence $t^s_{N(t)-1} = t^s_{N(t)} - t^s_{N(t)-1} = x^s_{N(t)}$. Therefore, conditional on this information, we may consider the distribution of the returns per time unit instead of the returns themselves. Finally, for returns and durations we may also consider the left hand limit as time is strictly less than $t$ and, as noted in Section 2, the left hand limit and the value at time $t$ differ only if there is a trade at time $t$.  

\[13\] Ref. also Theorem II.T15 of Brémaud (1981) and the discussion following it.
It is argued by Engle (2000) among others that one suitable measure of price is the midquote, i.e. the average of the ask and bid prices at the time of the trade. This choice reduces the problems associated with the bid-ask bounce, but microstructure imperfections may still be argued to induce serial correlation in the returns calculated from midquotes; see Hasbrouck (1991). As Engle (2000, p. 17) demonstrates, returns per time unit calculated from the midquotes may, as an approximation, be assumed to follow an ARMA(1,1) process with conditionally heteroskedastic error terms. Hence we assume that the returns per time unit at time $t$, conditional on $\mathcal{F}_{t-}$ and $s_{N(t)} = s$, are governed by

$$
\tilde{y}_{N^s(t)} = \phi_0^s + \phi^s \tilde{y}_{N^s(t-)} + \theta^s \tilde{\epsilon}_{N^s(t-)} + \tilde{\epsilon}_{N^s(t)}
$$

Var $[\tilde{y}_{N^s(t)} | \mathcal{F}_{t-}, s_{N(t)} = s] = \text{Var} [\tilde{\epsilon}_{N^s(t)} | \mathcal{F}_{t-}, s_{N(t)} = s] = \tilde{h}_{N^s(t)}$

where $\phi_0^s$, $\phi^s$, and $\theta^s$ are parameters, $\tilde{\epsilon}_{N^s(t)}$ are error terms with a distribution specified below, and the $\tilde{h}_{N^s(t)}$ are $\mathcal{F}_{t-}$-conditionally deterministic conditional variances of the returns per time unit (possible parameterizations of the $\tilde{h}_{N^s(t)}$'s are discussed in the next subsection). We have chosen to place a tilde ($\tilde{\epsilon}$) also over $\epsilon$ and $h$ to point out that they are measured per time unit. To facilitate quasi-maximum likelihood estimation we need to make a distributional assumption on which we base our likelihood function. For convenience, although it may be rather unrealistic in practice, we assume that the returns are conditionally normally distributed with the mean and variance specified above. That is, we assume that

$$
\tilde{\epsilon}_{N^s(t)} \mid \mathcal{F}_{t-}, s_{N(t)} = s \sim \left(\tilde{h}_{N^s(t)}\right)^{1/2} \cdot z_{N(t)}
$$

where $z_{N(t)} \sim \text{nid}(0, 1)$.

### 5.1 Modelling the conditional variances of returns

We now consider how to model the $\tilde{h}_{N^s(t)}$'s, the conditional variances of the returns per time unit $\tilde{y}_{N^s(t)}$. An approach that may come to mind is an adaptation of the usual multivariate GARCH models to the facts that trades occur irregularly spaced in time and non-contemporaneously for the different stocks (i.e. that at any given point in time there is a trade with at most one stock). The non-contemporaneity of the trades, however, has some consequences on how to sensibly define multivariate GARCH type models in the present situation. In particular, concepts such as ‘contemporaneous covariance’ or ‘contemporaneous correlation’ cannot be defined in their usual sense as there are no contemporaneously observed trades.$^{14}$

The closest thing to the contemporaneous returns we do observe are the most recent returns (per time unit), which for any positive $t$ are defined by $\tilde{y}_{N(t)} = (\tilde{y}_{N^1(t)}, \ldots, \tilde{y}_{N^S(t)})^\prime$, where $\tilde{y}_{N^s(t)}$ is the return of stock $s$ over the interval $t^s_{N(t)} -$

$^{14}$Admittedly, one could assume the existence of latent return processes at each time point and define covariances or correlations through these; we leave this alternative for future research.
\[ t^s_{N^i(t^i)-1} \] divided by the square root of this duration. As returns are sensibly defined only at time points where a trade occurs, and the modelling of the returns is done conditionally also on the ‘type’ mark, the relevant conditioning information is \( F_t \), and the information that at time \( t \) there is a trade with stock \( s \). Conditional on this information, however, the \( \tilde{y}^1_{N^i(t^i)}, i \neq s, \) are known (being in the information set), and hence the conditional mean and the conditional covariance matrix of \( \tilde{y}_{N(t)} \) are

\[
\left( \tilde{y}^1_{N^1(t^1)}, \ldots, \tilde{y}^{s-1}_{N^{s-1}(t^i)}, \phi^0 + \phi^s \tilde{y}^s_{N^i(t^i)} - \tilde{h}^{s+1}_{N^{s+1}(t^i)} \right) \] \( \tilde{y}^s_{N^s(t^i)} \)

and a matrix of zeros except for \( \tilde{h}^s_{N^i(t^i)} \) in the \((s, s)\)-element, respectively. This makes the usual multivariate GARCH type of approach problematic since it does not seem sensible to model the conditional variance at the times of trades as follows. First note that it suffices to define conditional variances at \( t \), and a simpler generalization of existing univariate models to the multivariate setting as follows. First note that it suffices to define conditional variances at \( t^s, s = 1, \ldots, S, i = 1, 2, \ldots, \), as these are the occasions when they are associated with an actual trade. In a univariate setting, Engle (2000) proposed modelling the conditional variance of the error terms \( \epsilon^s_{N^i(t)} \) as a GARCH–type of equation, and named this the Ultra-High-Frequency–GARCH, or UHF–GARCH, model. Following the spirit of the UHF–GARCH model, we extend this idea to the multivariate setting: In addition to having an autoregressive formulation for the conditional variance of each of the series, we allow for the possibility that the most recent shocks for the other stocks in the equation (5) as well as the associated conditional variances may affect each stock’s conditional variance. The motivation for this is to allow for the possibility that the conditional volatility of any stock is dependent on other stocks’ lagged conditional volatilities. In formulas,

\[
\begin{align*}
\tilde{h}^1_{N^1(t^1)} &= \omega^1 + \alpha^1 \epsilon^1_{N^1(t^1)} + \cdots + \alpha^1 \epsilon^2_{N^1(t^1)} + \beta^1 \tilde{h}^1_{N^1(t^1)} + \cdots + \beta^1 \tilde{h}^S_{N^S(t^1)} \\
\tilde{h}^2_{N^2(t^2)} &= \omega^2 + \alpha^2 \epsilon^1_{N^2(t^2)} + \cdots + \alpha^2 \epsilon^2_{N^2(t^2)} + \beta^2 \tilde{h}^1_{N^1(t^2)} + \cdots + \beta^2 \tilde{h}^S_{N^S(t^2)} \\
& \vdots \\
\tilde{h}^S_{N^S(t^1)} &= \omega^S + \alpha^S \epsilon^1_{N^S(t^S)} + \cdots + \alpha^S \epsilon^2_{N^S(t^S)} + \beta^S \tilde{h}^1_{N^1(t^S)} + \cdots + \beta^S \tilde{h}^S_{N^S(t^S)}
\end{align*}
\]

Here we have written \( \tilde{e}^2_{N^i(t^i)} \) for \( \left( \tilde{\epsilon}^s_{N^i(t^i)} \right)^2, \) to make the notation slightly lighter. Also, the indexing ‘\( i \)’ denotes the \( i \)th trade on a stock and it should cause no confusion that the same index is used on all the stocks. Furthermore, for example \( N^1(t^1) = i \) and \( N^1(t^1) = i - 1 \), which could be used to make the notation lighter. However, for example \( N^2(t^2) \) has no simpler form, and to keep the notation similar throughout the equation, we retain this rather heavy notation. To simplify the expressions, note that \( \tilde{e}^i_{t^i} = \tilde{h}^i_{t^i} z^2_{t^i} \) for any \( i = 1, 2, \ldots \) (we write \( z^2_{t^i} \) for \( (z^i_{t^i})^2 \)).
Furthermore, for any $i = 1, 2, \ldots$ and $k, l = 1, \ldots, S$, denote $c_{ki}^l = \alpha_{kl} z_{ki}^l + \beta_{kl}$. Using these notations, equations (8) can be written somewhat more compactly as

$$
\tilde{h}_{1N}(t_i^1) = \omega_1 + \sum_{s=1}^{S} c_{1s}(t_i^1 - \tilde{h}_{sN}(t_i^1))
$$

$$
\tilde{h}_{2N}(t_i^2) = \omega_2 + \sum_{s=1}^{S} c_{2s}(t_i^2 - \tilde{h}_{sN}(t_i^2))
$$

$$
\vdots
$$

$$
\tilde{h}_{SN}(t_i^S) = \omega_S + \sum_{s=1}^{S} c_{Ss}(t_i^S - \tilde{h}_{sN}(t_i^S))
$$

Note that the formulation in (9) (or in (8)) is analogous to the multivariate GARCH models introduced by Jeantheau (1998) and further considered in He and Teräsvirta (2004). If $\alpha_{ij} = \beta_{ij} = 0$ for $i \neq j$ the conditional variances do not depend on the other stocks, and the equations reduce to the ones in Engle (2000). One simplification that may be of practical interest is to set $\beta_{ij} = 0$ for $i \neq j$, but let the $\alpha_{ij}$’s be positive. This allows for interdependencies between the stocks while making numerical computations less involved as the conditional variances depend only on their own past and not on that of the other stocks.

A possible extension to this formulation might be to include information from the trading time process into the volatility dynamics. For instance, one theory in the market microstructure literature (see Easley and O’Hara (1992) and also Engle (2000, Sec. 6)) interprets low trading activity as evidence that there are no news, which in turn lowers volatility. In our framework, trading activity is directly measured by the intensities $\lambda^s(\cdot)$. Because long durations mean infrequent trading, the reciprocal of the observed duration, $1/x^s$, is also a proxy for trading intensity. Including one of these variables on the right hand side of (9) might thus be interesting.

6 Strict stationarity

In this section we provide preliminary comments concerning the existence of a strictly stationary version of the models proposed above. The concept of stationarity for a marked point process is usually defined as invariance with respect to shifts in time while leaving the marks unchanged (ref. Matthes, Kerstan, and Mecke (1978, Ch. 6) and Daley and Vere-Jones (1988, Ch. 10)). In the following we only consider the properties of the ground process, and leave the investigation of the mark process for future research.

6.1 Ground process

Properties of non-marked point processes are well-established in the literature. In particular, Brémaud and Massoulié (1996) give conditions for the existence of a stationary version of the ground process in the case of a multivariate nonlinear Hawkes
process (2), and also for the asymptotic convergence to this stationary process. For ease of reference we present here a simplified version of their results.

**Proposition 1** Assume that the function \( \phi^i : \mathbb{R} \rightarrow \mathbb{R}^+ \) is \( \alpha^i \)-Lipschitz, \( i = 1, \ldots, S \), and that the functions \( h_{ij} : \mathbb{R}^+ \rightarrow \mathbb{R} \) are such that the \( S \times S \) matrix \( A \) with entries \( a_{ij} = \alpha^i \int_{(0, \infty)} |h_{ij}(t)| dt \) has a spectral radius strictly less than 1. Then there exists a unique stationary point process \( N \) with dynamics (2) and finite average intensity \( \mathbb{E}[N((0,1])] \).

As a simple example, consider the usual multivariate linear Hawkes process with exponential decay (3), in which case \( h_{ij}(u) = \sum K k=1 \alpha_{ijk} e^{-\beta_{ijk} u} \) and \( \phi(x) = \omega_i + x \) \( (i, j = 1, \ldots, S) \). The Lipschitz constant for the functions \( \phi^i \) is 1, and the conclusions of Proposition 1 thus hold whenever the matrix \( A \) with elements

\[
a_{ij} = \sum_{k=1}^K \int_{(0, \infty)} \alpha_{ijk} e^{-\beta_{ijk} t} dt = \sum_{k=1}^K \frac{\alpha_{ijk}}{\beta_{ijk}} (10)
\]

has a spectral radius strictly less than 1. Whether this condition is satisfied or not can be easily verified.

We note that in the case of the ACI model of Russell (1999) establishing stationarity appears to be problematic (Russell (1999) does present a conjecture concerning this but does not provide a proof).

### 7 Maximum likelihood estimation

A practical difficulty with models of ultra-high-frequency data is the computational burden resulting from the often vast amount of data considered. It is therefore beneficial if the estimation of the model can be performed without computer-intensive methods. This is the case for the models we propose as they can be estimated using the usual quasi-maximum likelihood approach. The following proposition gives the form of the log-likelihood for a realization of a marked point process.

**Proposition 2** Let \( N \) be a regular marked point process on \([0, T] \times M\) for some finite positive \( T \), and let \((t_1, m_1), \ldots, (t_{N_T(T)}, m_{N_T(T)})\) be a realization of \( N \) over the interval \([0, T]\). Then the log-likelihood of this realization is

\[
\sum_{i=1}^{N_T(T)} \ln \lambda^s(t_i) - \sum_{s=1}^S \int_0^T \lambda^s(u) du \right) + \left[ \sum_{i=1}^{N_T(T)} \ln f^{(r)}(r_i | t_i, s_i) \right]. \tag{11}
\]

---

\(^{16}\)Ref. Theorem 7 of Brémaud and Massoulié (1996).
Using the equations can be written as decay (3), the integration in (11) can be performed analytically, and the integral taking logarithms yields the result.

On the econometrics of multivariate marked ultra-high-frequency data

A realization is given by

By Proposition 7.3.III of Daley and Vere-Jones (2003), the likelihood of such a realization is given by

Using the equations \( f^{(m)}(s_{N(t)}) = s_i \), \( f^{(r)}(s_{N(t)}) = s \) \( |F_{t_{i}}\rangle \), and \( \lambda_g(t) = \sum_{s=1}^{S} \lambda^s(t) \) the likelihood can be written as

Using the equations \( f^{(m)}(s_{N(t)}) = s_i \), \( f^{(r)}(s_{N(t)}) = s \) \( |F_{t_{i}}\rangle \), and \( \lambda_g(t) = \sum_{s=1}^{S} \lambda^s(t) \) the likelihood can be written as

Taking logarithms yields the result.

We note that in the case of a linear multivariate Hawkes model with exponential decay (3), the integration in (11) can be performed analytically, and the integral equals\(^{17}\)

Avoiding numerical integration reduces the time needed for evaluation of the likelihood considerably, hence making estimation faster. This is of importance since the data sets considered in practice are potentially very large and the computational burden has to be kept within reasonable limits.

\(^{17}\)As all the points are in the interval \([0,T]\), the intensity (3) can be written as

where \( 1(\cdot) \) is the indicator function. Hence

\[
\int_0^T \lambda^s(t) dt = T \omega_s + \sum_{i=1}^{S} \sum_{j=1}^{N^s(T)} \sum_{k=1}^{K} \alpha_{sik} \beta_{sik} (1 - e^{-\beta_{sik}(T-t_j^i)})
\]

\[
= T \omega_s + \sum_{i=1}^{S} \sum_{j=1}^{N^s(T)} \sum_{k=1}^{K} \frac{\alpha_{sik}}{\beta_{sik}} (1 - e^{-\beta_{sik}(T-t_j^i)})
\]
We have only considered parameterizations for the intensity where the past returns do not enter the $\lambda^s(t)$-processes. Similarly, in the examples considered, the past intensities do not enter the conditional distribution of the returns (except in the possible extension discussed in the end of Section 5.1). For this reason, the first term in (11) contains only the parameters of the intensity functions, and the second term the parameters of the returns part. Hence maximization can be done separately for these two parts. If the intensities and returns are allowed to influence each other, the maximization needs to be done jointly. This is the case in particular if the trading time process is allowed to enter the volatility dynamics as discussed as a possible extension in Section 5.1.

Conditions for consistency and asymptotic normality of the quasi-maximum likelihood estimator in the case of a non-marked point process are given in Ogata (1978). Extending these results to the present case would be interesting but is far beyond the scope of this paper. We shall proceed under the assumption that the quasi-maximum likelihood estimator based on the loglikelihood (11) is consistent and asymptotically normally distributed.

8 Preliminary notes on the data

In future work, our intention is to demonstrate the use of the proposed models using Trades and Quotes (TAQ) data available from the New York Stock Exchange. For the moment this section only describes some empirical issues that need to be addressed prior to fitting a model to ultra-high-frequency data. The situation considered is one where the econometrician observes a sequence of financial events on, for simplicity, two stocks over a period of time. The events (points of the ground process) we consider are the trades on these stocks. The observable marks considered associated with these trades are (i) which of the stocks traded as well as (ii) the associated return per time unit. As the measure of price (from which the returns are computed) we use the midquote, i.e. the average of the ask and bid prices at the time of the trade, a choice that was motivated in Section 5.

8.1 Transformations to the data

One issue that needs to be dealt with is the fact that quite often there are several consecutive trades on the same stock without a change in the midquote price, which results in zero returns. These zero returns constitute a problem as our model does not accommodate for them. On the other hand, because our interest is in modelling volatility, these zero returns may not be so interesting as they do not contribute to changes in volatility. Therefore we decide to consider only those trades where there is a change in the midquote, and thus remove the trades which are associated with zero returns.

In practice it may also occur that the two stocks are traded simultaneously. The modelling approach employed does not accommodate for this possibility either, and we have chosen to adjust the data as follows (for a similar adjustment, see eg. Bowsher
(2003)). For each of these simultaneous trades, we randomly move one of the trades backward or forward by a small amount of time. If there are nearby trades with either of the stocks this is taken into consideration in order to distort the ordering and spacing of the trades as little as possible. This small adjustment of a number of the trades should only have a small effect on the final results.

8.2 Diurnal adjustment

It is well documented in the literature that there is a clear intraday seasonal (or diurnal) pattern over the trading day; see e.g. Engle (2000). Both the trading times and the associated returns are affected by this and need to be diurnally adjusted prior to fitting a model to them. We adapt the diurnal adjustment method used by Engle (2000) to the multivariate setting and do the adjustment as follows.

First consider the diurnal adjustment of the trading times, or equivalently, the durations. If the duration series of the two stocks would be adjusted separately, this would need to be done with care in order not to change the ordering of the trades between the stocks. Also, from a point process point of view it may be more natural to diurnally adjust the ground process. Because the durations of the individual processes are sums of the durations of the ground process, diurnally adjusting the ground process will in effect also diurnally adjust the individual series.

We perform the adjustment to the ground durations series by approximating the average durations using a cubic spline and then removing this diurnal component from the durations. We do this by first averaging the durations over 30-minute periods (9:30–10:00, 10:00–10:30, etc.), setting the average durations of the midpoints of these intervals (9:45, 10:15, etc.) to the resulting averaged values, and then fitting a cubic spline over the course of the day using these midpoints as fixed. The original durations series is then diurnally adjusted by dividing it with the estimated diurnal component. From this diurnally adjusted ground series the individual diurnally adjusted series can be easily recovered.

Now consider the diurnal adjustment of the returns per time unit. The returns per time unit are first calculated using the unadjusted durations. In order to retain the possibly different magnitudes of volatilities for the different stocks we apply the same diurnal adjustment to all the series. In effect, this means that we first pool all the returns into one series. Then we average the absolute returns over 30-minute periods and fit a cubic spline similarly as above. The returns are then divided by this diurnal component. Again, the individual diurnally adjusted returns series are easily recovered from this pooled diurnally adjusted series.

9 Conclusions and directions for future work

In this paper we have considered the econometric modelling of ultra-high-frequency data and, in particular, how to model the multivariate trading time process together with the volatility of the associated price process. Our formulation of the model draws heavily on the theory of marked point processes. The model for the timings
of the trades is a continuous time model formulated using the multivariate intensity function. For the returns, we proposed a simple generalization of existing GARCH-type models for ultra-high-frequency data to the multivariate setting. Preliminary remarks concerning the stationarity of the model proposed are also given. The model can be estimated using the usual quasi-maximum likelihood approach, and we note how the estimation exercise can be made easier in certain circumstances. Prior to fitting a model to ultra-high-frequency data, a number of transformations need to be done, and these are briefly discussed.

This preliminary version of the paper presents the current state of work in progress. Several issues that need to be examined and discussed in a future revision of this paper still remain to be done. From an applied point of view, a main task ahead is thorough empirical application demonstrating the usefulness of the model proposed. It is also of importance to perform a convincing simulation study examining the convergence of the parameter estimators. The empirical exercise may also reveal directions into which the model needs to be extended. For this purpose, it would be useful to devise a set of misspecification tests to evaluate the adequacy of the models proposed. From a theoretical point of view, it would be of importance to derive conditions under which a strictly stationary solution of the model exists. Examination of stationarity both validates the theoretical sensibility of the model proposed as well as provides parameter restrictions that need to be imposed in the practical application of the model. In future work, it would also be interesting to consider alternative specifications for both the trading intensity and the return volatility. Another interesting topic would be to incorporate some other information associated with the trades, such as volume or the knowledge of whether the trade was initiated by a buyer or a seller.
References


based, multivariate point process models,” Nuffield College Economics Working

Brémaud, P. (1981): Point Processes and Queues: Martingale Dynamics. Springer-
Verlag, New York.


Processes. Springer-Verlag, New York.


price dynamics,” Annales d’Economie et de Statistique, 60, 207–238.

Easley, D., N. M. Kiefer, and M. O’Hara (1997): “The information content of

Easley, D., and M. O’Hara (1992): “Time and the process of security price ad-

Engle, R. F. (2000): “The econometrics of ultra-high-frequency data,” Economet-
rica, 68, 1–22.

17, 425–446.

new model for irregularly spaced transaction data,” Econometrica, 66, 1127–1162.

Engle, R. F., and Z. Sun (2005): “Forecasting volatility using tick by tick data,”
unpublished manuscript.

Ghysels, E., and J. Jasiak (1998): “GARCH for irregularly spaced financial data:
The ACD–GARCH model,” Studies in Nonlinear Dynamics and Econometrics, 2,
133–149.


Hall, A. D., and N. Hautsch (2004): “A continuous-time measurement of the
of Economics, University of Copenhagen.
REFERENCES


REFERENCES


STABILITY OF
TIME SERIES MODELS
Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models
Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models

Abstract

This paper studies a class of Markov models which consist of two components. Typically, one of the components is observable and the other is unobservable or ‘hidden’. Conditions under which geometric ergodicity of the unobservable component is inherited by the joint process formed of the two components are given. This implies existence of initial values such that the joint process is strictly stationary and $\beta$-mixing. In addition to this, conditions for the existence of moments are also obtained and extensions to the case of nonstationary initial values are provided. All these results are applied to a general model which includes as special cases various first order generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive conditional duration (ACD) models with possibly complicated nonlinear structures.

---

This paper is joint work with Pentti Saikkonen.

We thank a co-editor of Econometric Theory and three anonymous referees for helpful comments and suggestions. First version of this paper was completed in October 2004 and subsequently material from it has been presented at the New Zealand Econometric Study Group Meeting, Christchurch, March 2005, International Workshop on Financial Econometrics and Statistics, Perth, April 2005, RUrg Workshop on Financial Econometrics, Helsinki, August 2005, Econometric Society World Congress, London, August 2005, International Conference on Finance, Copenhagen, September 2005, European Winter Meeting of the Econometric Society, Istanbul, October 2005, and in seminars at Stockholm School of Economics, Université catholique de Louvain, Université Libre de Bruxelles, University of Oxford, and Victoria University of Wellington. We thank the participants in the occasions, Michael McAleer in particular, for their comments.
1 Introduction

This paper is concerned with probabilistic properties of two common classes of models, namely generalized autoregressive conditional heteroskedasticity (GARCH) models and autoregressive conditional duration (ACD) models. GARCH models were pioneered by Engle (1982) and Bollerslev (1986), and have ever since been widely used to analyze financial time series. The more recent ACD models were introduced by Engle and Russell (1998) to model the time dimension of irregularly spaced ultra-high-frequency data.

Our study of GARCH and ACD models makes use of the theory of Markov chains. Both GARCH and ACD models can be thought of as consisting of two components of which one is observable (say, returns or durations) and the other is unobservable or ‘hidden’ (say, conditional variance or conditional expected duration). From the viewpoint of Markov chain theory, the unobservable component can be investigated as a Markov chain of its own in isolation from the observable component. However, it is also useful to consider both components jointly as a single Markov chain. For instance, in the development of statistical estimation and testing theory it is pertinent to know when the joint process formed of the two components is, for example, stationary and has finite moments of some order. To make such results readily available, we obtain conditions under which the ergodicity, or more precisely, $V$-geometric ergodicity of the hidden process (viewed as a Markov chain of its own) is inherited by the joint process (consisting of both the observable and hidden components). An immediate consequence of this is that, with an appropriate choice of initial values, the joint process is strictly stationary and $\beta$–mixing (or absolutely regular) with certain moments existing. Inspired by the recent work of Liebscher (2005) we also obtain conditions which imply $\beta$–mixing in the case of nonstationary initial values. Because other applications may emerge these results are first obtained for a general class of Markov models defined in terms of transition probability measures.

We apply the aforementioned results obtained for our general class of Markov models to a sub-class which contains many first order GARCH and ACD models as special cases. For simplicity, we concentrate on the leading case of first order GARCH and ACD models but allow for more complicated nonlinear structures than in earlier literature. Our results apply to the families of GARCH and ACD models introduced by Hentschel (1995) and Fernandes and Grammig (2006), respectively, and thereby to several commonly used GARCH and ACD models. Our results also apply to the integrated GARCH (IGARCH) model and provide a rigorous proof of its short memory nature previously demonstrated by Ding and Granger (1996) using more elementary methods. In addition to these models, the GARCH-in-mean (GARCH–M) model as well as some GARCH and ACD models with rather complicated nonlinear structures are also covered. For some of these models the obtained conditions for ergodicity, mixing, and stationarity appear new.

The approach used in this paper has predecessors. Genon-Catalot, Jeanntheau, and Larédo (2000) considered a general class of Markov models referred to as a ‘hidden Markov model’ and obtained results similar to ours for stochastic volatility models. Carrasco and Chen (2002) attempted to generalize these results by formulating a
‘generalized hidden Markov model’ which could also be applied to GARCH and ACD models. Unfortunately, however, this generalization appears too general to be useful. We show by a counterexample that the conditions required for the generalized hidden Markov model do not necessarily guarantee the validity of the ergodicity and mixing results given by Carrasco and Chen (2002). We wish to stress, however, that this only concerns their results on the generalized hidden Markov model. All their results on GARCH, stochastic volatility, and ACD models still remain valid as long as their results on the generalized hidden Markov model are corrected. Our paper presents such a correction (Proposition 1 below). An alternative correction has recently been provided by Carrasco and Chen (2005).

As far as GARCH and ACD models are concerned, it should be mentioned that related results on ergodicity, mixing, strict stationarity, and existence of moments have previously appeared in Nelson (1990), Bougerol and Picard (1992), Duan (1997), Zhang, Russell, and Tsay (2001), Carrasco and Chen (2002), Ling and McAleer (2002), Ling and McAleer (2003), Lanne and Saikkonen (2005), and Straumann and Mikosch (in press). Our contribution to this work is that we show how these models can be handled in a unified framework which also applies when very general nonlinear structures or even models, other than GARCH and ACD models, are of interest.

The rest of this paper is organized as follows. Our general class of Markov models is studied in Section 2. In Section 3 these results are specialized to a specific sub-class of models which contains various GARCH and ACD models. Concluding remarks are presented in Section 4. Proofs of all the results are given in an Appendix.

2 General Markov Models

We motivate our general Markov model by using the standard GARCH(1,1) model

\begin{align}
    u_t &= h_t^{1/2} \varepsilon_t \\
    h_t &= \omega + \beta h_{t-1} + \alpha u_{t-1}^2,
\end{align}

where \( \varepsilon_t \sim \text{i.i.d.}(0, 1) \) with \( \varepsilon_t \) independent of \((u_s, h_s), s < t, \) and the parameters satisfy \( \omega > 0, \alpha \geq 0, \beta \geq 0, \) and \( \alpha + \beta < 1. \) Here \( u_t \) is an observed process and \( h_t \) is its unobserved conditional variance. Substituting \( h_t \) from the latter equation to the former shows that \((u_t, h_t)\) is a Markov chain. On the other hand, substituting \( u_t \) from the former equation to the latter shows that \( h_t \) can be viewed as a separate Markov chain defined by the equation \( h_t = \omega + \beta h_{t-1} + \alpha u_{t-1}^2 \varepsilon_{t-1}^2. \) Thus, one can use the theory of Markov chains to study properties of either the joint process \((u_t, h_t)\) or of \( h_t \) in isolation from the process \( u_t. \) An approach like this was recently used by Carrasco and Chen (2002) who showed how to extend results on stationarity and mixing obtained for \( h_t \) to the joint process \((u_t, h_t).\) These authors also defined a ‘generalized hidden Markov model’ in order to make their approach generally applicable.

\footnote{While revising this paper we learned of the recent work of Medeiros and Veiga (2004), Franço and Zakoïan (2005), and Kristensen (2005) where related results are also obtained.}
We adopt the same idea and consider a general Markov model with a structure similar to that in the GARCH(1,1) model (1). A comprehensive reference of the needed Markov chain theory is Meyn and Tweedie (1993) whereas Chan (1990) provides a short review. As a further reference we mention Doukhan (1994) where the employed concept of $\beta$–mixing and its relation to other mixing concepts are discussed.

Consider two stochastic processes, $Y_t$ and $X_t$ ($t = 0, 1, \ldots$), taking values in measurable spaces $(Y, \mathcal{B}(Y))$ and $(X, \mathcal{B}(X))$, respectively. Suppose the joint process $Z_t = (Y_t, X_t)$ is a (time homogeneous) Markov chain on $(Z, \mathcal{B}(Z))$, where $Z = Y \times X$ and $\mathcal{B}(Z) = \mathcal{B}(Y) \otimes \mathcal{B}(X)$, and let $P^n_Z(z, n) = \Pr(Z_n \in A \mid Z_0 = z)$, $z \in Z$, $A \in \mathcal{B}(Z)$, signify its $n$–step transition probability measure $(P^n_Z(\cdot, \cdot) = P_Z(\cdot, \cdot)$ and similarly for other transition probability measures). As in the generalized hidden Markov model of Carrasco and Chen (2002) (conditions (ii) and (iii) of their Definition 3) we now assume that, for all $t \geq 1$, the conditional distribution of $Y_t$ given $(X_t, Y_{t-1}, X_{t-1}, \ldots, Y_0, X_0)$ only depends on $X_t$ and that the conditional distribution of $Y_t$ given $X_t = x$ does not depend on $t$. Let $\pi_{Y \mid X}(\cdot \mid x)$ signify this conditional probability distribution. As in the proof of Proposition 4(i) of Carrasco and Chen (2002) we can then write $P^n_Z(z, n) = \Pr(dy \mid dx, Z_0 = z) \Pr(dx \mid Z_0 = z)$ where $z = (y, x)$ and the former factor of the product can be replaced by $\pi_{Y \mid X}(dy \mid x)$. In the aforementioned proof, Carrasco and Chen (2002) use the assumption that $X_t$ is an (unobserved) stationary Markov chain (condition (i) of their Definition 3) and replace the latter factor by $P^n_X(x, dx)$, the $n$–step transition probability measure of $X_t$. However, this replacement is problematic in GARCH models, for example. Although $h_t$ in the GARCH(1,1) model (1) can be treated as a Markov chain of its own this Markov chain is not identical to the latter component of the joint process $(u_t, h_t)$. Specifically, given an initial value $(u_0, h_0)$, the joint process implies that $h_1 = \omega + \beta h_0 + \alpha u_0^2$ whereas $h_1 = \omega + \beta h_0 + \alpha h_0^2$ results when $h_2$ is treated as a separate Markov chain. Thus, if the joint process $(u_t, h_t)$ is the Markov chain considered, the conditional probability distribution of $h_1$ also depends on the initial value $u_0$, not only on $h_0$.

Motivated by the preceding discussion we denote $\tilde{P}^n_Y(z, \cdot) = \Pr(X_n \in \cdot \mid Z_0 = z)$ and conclude that

$$P^n_Z(z, d\omega) = \pi_{Y \mid X}(dy \mid x) \tilde{P}^n_Z(z, dx). \quad (2)$$

As noticed above, the dependence of the latter factor on the right hand side on the initial value $y$ is inconvenient. Fortunately, however, this matter can be handled (at least) in GARCH models. To see this, consider again the GARCH(1,1) model (1) and the related two-dimensional Markov chain. Given the initial value $(u_0, h_0)$, the joint process implies that $h_1 = \omega + \beta h_0 + \alpha u_0^2$ and, as can be easily checked, $h_2 = \omega + \beta h_0 + \alpha h_0^2$ where $h_0 = \omega + \beta h_0 + \alpha u_0^2$. On the other hand, when $h_t$ is treated as a separate Markov chain, $h_1 = \omega + \beta h_0 + \alpha h_0^2$ is obtained. Thus, the generation mechanism of $h_2$ (based on the two-dimensional Markov chain) is entirely similar to that of $h_1$ obtained when $h_t$ is treated as a separate Markov chain. Only the initial value $h_0$ that appears in $h_2$ is defined in a special way. This clearly extends to larger values of $t$ so that, apart from the definition of the initial value, the generation mechanism of $h_t$ ($t \geq 2$) based on the two-dimensional Markov chain $(u_t, h_t)$ is identical to that of $h_{t-1}$ obtained when $h_t$ is analyzed separately.
Using the above discussion on the GARCH(1,1) model (1) as a pattern we now replace the probability measure \( \tilde{P}_n^{X}(z,\cdot) \) in (2) by a counterpart which, for some \( j \geq 0 \), can be treated as an \( (n-j) \)-step transition probability measure of a separate Markov chain on \((X,\mathcal{B}(X))\). We state the following assumption.

**Assumption 1** Let \( Z_t = (Y_t, X_t) \) \((t = 0, 1, \ldots)\) be a Markov chain on \((Z, \mathcal{B}(Z))\) where \( Z = Y \times X \) and \( \mathcal{B}(Z) = \mathcal{B}(Y \times X) \). Assume the following conditions.

(a) For all \( n \geq 1 \) the \( n \)-step transition probability measure of \( Z_t \) can be expressed in the form (2) where \( \pi_{Y|X}(\cdot| x) \) is the conditional probability distribution of \( Y_t \) given \( X_t = x \).

(b) There exist a function \( \lambda : Z \rightarrow X \), an integer \( j \geq 0 \), and a transition probability measure \( P^{X}_X(\cdot,\cdot) \) of a Markov chain on \((X, \mathcal{B}(X))\) such that, for all \( n > j \),

\[
\tilde{P}_n^{X}(z,\cdot) = P^{X}_{n-j}(\tilde{x},\cdot) \quad \text{where} \quad \tilde{x} = \tilde{x}(z) = \lambda(z).
\]

It is implicit in Assumption 1(a) that conditions (ii) and (iii) defining the generalized hidden Markov model of Carrasco and Chen (2002) are satisfied. Furthermore, when \( X_t \) is viewed as a part of the joint process \((Y_t, X_t)\) its transition probability measure is assumed to agree with the transition probability measure of a separate Markov chain on \((X, \mathcal{B}(X))\) with suitably defined initial values. However, \( X_t \) is not necessarily a Markov chain because its transition probabilities may depend on the initial value of the joint process \((Y_t, X_t)\). Therefore, condition (i) of Definition 3 of Carrasco and Chen (2002) need not hold (not even without the word ‘stationary’).

Assumption 1 is more restrictive than required for the generalized hidden Markov model of Carrasco and Chen (2002). To demonstrate the need of such a restriction, we use a model which is a special case of Example 1 of Carrasco and Chen (2002). Thus, let \( \varepsilon_t \) be a sequence of n.i.d.(0,1) random variables and consider the model

\begin{align*}
Y_t &= \varepsilon_t + \varepsilon_{2t} \\
X_t &= \varepsilon_t,
\end{align*}

\((t = 1, 2, \ldots)\). The model is extended for \( t = 0 \) by assuming that \( Y_0 \) and \( X_0 \) are independent of each other and of \( \{\varepsilon_t, t \geq 1\} \) with standard normal distributions. Clearly, \( X_t \) is a stationary, geometrically ergodic, and \( \beta \)-mixing Markov chain. It is also straightforward to verify that the conditions required for the generalized hidden Markov model of Carrasco and Chen (2002) are satisfied and, by Proposition 4 of that paper, the joint process \((Y_t, X_t)\) should be a geometrically ergodic and \( \beta \)-mixing Markov chain. This, however, is not the case. Because \( \text{Cov}(Y_t, Y_{2t}) = \text{Cov}(\varepsilon_{2t}, \varepsilon_{2t}) = 1 \) for all \( t \geq 1 \) the process \( Y_t \) is not strong mixing and, hence, not \( \beta \)-mixing (cf. Proposition 1, p. 4, and Theorem 3, p. 9, in Doukhan (1994)). Also, the conditional distribution of \((Y_t, X_t)\) given its past is not a function of \((Y_{t-1}, X_{t-1})\) only and, therefore, \((Y_t, X_t)\) is not a (geometrically ergodic) Markov chain.

An important requirement in Assumption 1 is that the joint process \((Y_t, X_t)\) is
a Markov chain because then models like (3) are ruled out.\textsuperscript{19} It is straightforward to check that Assumption 1 holds for the GARCH(1,1) model (1) with \( j = 1 \), the function \( \lambda (u, h) = \omega + \beta h + \alpha u^2 \), \( \pi_{Y|X} (\cdot \mid x) \) the conditional distribution of \( u_t \) given \( h_t \), and \( P_X (\cdot, \cdot) \) the transition probability measure associated with \( h_t \) viewed as a separate Markov chain. The GARCH(1,1) model (1) is a special case of the model

\[
\begin{align*}
Y_t &= F_y (X_t, \zeta_t) \quad \text{(4)} \\
X_t &= F_x (X_{t-1}, Y_{t-1}) \quad \text{(5)}
\end{align*}
\]

where \( \zeta \) is an i.i.d. error term independent of \( (Y_s, X_s) \), \( s < t \), and the random vectors \( Y_t, X_t \), and \( \zeta \) take values on some subsets of Euclidean spaces equipped with Borel sigma fields. Substituting the right hand side of equation (5) for \( X_t \) in (4) shows that the joint process \( (Y_t, X_t) \) is a Markov chain. From (4) and (5) it also follows that the random vectors \( \zeta_t \) and \( X_t \) are independent and that the latter has the representation

\[
X_t = F_x (X_{t-1}, F_y (X_{t-1}, \zeta_{t-1})) \overset{def}{=} G_x (X_{t-1}, \zeta_{t-1}). \tag{6}
\]

Thus, \( X_t \) can be viewed as a Markov chain of its own and studied in isolation from \( Y_t \). Verifying that Assumption 1 is satisfied for the model (4)–(5) with \( \lambda (\cdot) = F_y (\cdot) \) is straightforward but somewhat technical. Details are therefore deferred to the Appendix.

We shall now show that Assumption 1 guarantees that ergodicity and mixing of the component process \( X_t \) are inherited by the joint process \( Z_t \). We use the \( V \)–geometric ergodicity of a Markov chain defined as follows (see Meyn and Tweedie (1993, p. 356)).

**Definition 1** The Markov chain \( Z_t \) is \( V \)–geometrically ergodic if there exist a real valued function \( V : \mathcal{Z} \to [1, \infty) \), a probability measure \( \pi_{Z} \) on \( \mathcal{B}(\mathcal{Z}) \), and constants \( \varrho < 1 \) and \( M_z < \infty \) (depending on \( z \)) such that

\[
\sup_{v \mid |v| \leq V} \left| \int_{\mathcal{Z}} P_{Z}^n (z, dw) v(w) - \int_{\mathcal{Z}} \pi_{Z} (dw) v(w) \right| \leq \varrho^n M_z \tag{7}
\]

for all \( z \in \mathcal{Z} \) and all \( n \geq 1 \).

The definition also assumes that the function \( V \) is integrable with respect to the probability measure \( \pi_{Z} \). When condition (7) holds we also say that the transition probability measure \( P_{Z} (\cdot, \cdot) \) is \( V \)–geometrically ergodic and similarly for other transition probability measures such as \( P_X (\cdot, \cdot) \). Note that the first integral in (7) equals the conditional expectation \( E (v (Z_n) \mid Z_0 = z) \).

\textsuperscript{19}Recently, Carrasco and Chen (2005) have provided a correction to their definition of the generalized hidden Markov model. This correction, which also rules out undesirable models such as (3), does not assume directly that the joint process \( (Y_t, X_t) \) is a Markov chain although this is implied by conditions (ii), (iii), and (iv) of the new definition. Interestingly, given conditions (ii) and (iii), condition (iv) is equivalent to the requirement that the joint process \( (Y_t, X_t) \) is a Markov chain. When the new definition is assumed, all the results in Carrasco and Chen (2002), including Proposition 4, remain valid.
Ergodicity, mixing, and existence of moments of a class of Markov models

The weakest form of $V$–geometric ergodicity is obtained when $V(\cdot) \equiv 1$ in which case the Markov chain $Z_t$ is said to be geometrically ergodic. Geometric ergodicity entails that the $n$–step transition probability measure $P_Z^n(z, \cdot)$ converges at a geometric rate to the probability measure $\pi_Z(\cdot)$ with respect to the total variation norm for all $z \in Z$. The probability measure $\pi_Z$ is often referred to as the stationary probability measure of $Z_t$. The reason is that geometric ergodicity implies stationarity of $Z_t$ if the initial value $Z_0$ is distributed according to the probability measure $\pi_Z$ (see Meyn and Tweedie (1993, p. 230–231)). A convenient feature of $V$–geometric ergodicity is that it automatically shows the existence of the expectation of $\int_Z \pi_Z(dw)v(w)$ for all $v$ such that $|v(\cdot)| \leq V(\cdot)$.

The following proposition establishes the $V$–geometric ergodicity of $Z_t$.

**Proposition 1** Suppose that the Markov chain $Z_t = (Y_t, X_t)$ satisfies Assumption 1 and that the transition probability measure $P_X(\cdot, \cdot)$ is $V_X$–geometrically ergodic. Then $Z_t$ is $V_Z$–geometrically ergodic for any function $V_Z : Z \rightarrow [1, \infty)$ such that $\int_Z \pi_{Y|X}(dy|x)V_Z(y,x) \leq cV_X(x)$ for all $x \in X$ and some $c < \infty$.

The condition imposed on the function $V_Z$ in Proposition 1 is automatically satisfied for $V_Z(y,x) = V_X(x)$, although more useful results can be obtained with other choices of $V_Z$. However, even this special case shows that the geometric ergodicity of $P_X(\cdot, \cdot)$ is inherited by $Z_t$ and, when initialized from its stationary distribution, $Z_t$ is stationary and $\beta$–mixing with geometrically decaying mixing numbers (see Meyn and Tweedie (1993, pp. 230–231) and Doukhan (1994, p. 4 and 89)). Thus, Proposition 1 provides us with results similar to those stated in Proposition 4 of Carrasco and Chen (2002).

While Proposition 1 makes it possible to apply limit theorems developed for Markov chains to functions of $Z_t$, regardless of initial values (see Meyn and Tweedie (1993, Chapter 17)), these theorems are not directly applicable when functions of $(Z_1, Z_1, ..., Z_1)$ are of interest (such functions are met, for instance, in the estimation theory of GARCH models, see e.g. Straumann and Mikosch (in press, Sections 6 and 7)). In such cases other limit theorems can be useful. Our next result gives sufficient conditions for $Z_t$ to be $\beta$–mixing. This result holds for a wide class of nonstationary initial values and provides the needed prerequisite for using limit theorems developed for near epoch dependent functions of mixing processes (see Davidson (1994, Sections 20.6 and 24.4) for such limit theorems and Francq and Zakoïan (2005) for a related discussion on the usefulness of limit theorems developed for mixing processes). We use a subscript in the expectation operator to indicate the initial distribution of the chain with respect to which the expectation is taken.

**Proposition 2** Let the assumptions of Proposition 1 be satisfied and the function $V_Z$ be as required in Proposition 1. Furthermore, let $\pi_X(\cdot)$ signify the stationary probability measure related to a Markov chain with transition probability measure $P_X(\cdot, \cdot)$. Suppose that (a) there exist constants $\varrho < 1$ and $R < \infty$ such that

$$\sup_{v:|v| \leq V_X} \left| \int_X P_X^n(x, dw)v(w) - \int_X \pi_X(dw)v(w) \right| \leq \varrho^nRV_X(x)$$
for all \( x \in \mathcal{X} \) and all \( n \geq 1 \), \( b \) \( E_\mu[V_X(\lambda(X_0, Y_0))] < \infty \) where \( \mu \) is the distribution of the initial value \( Z_0 = (Y_0, X_0) \), and \( c \) \( \int \pi_{Y|X}(dy|x) V_X(\lambda(x, y)) \leq cV_X(x) \) for all \( x \in \mathcal{X} \) and some \( c < \infty \). Then \( Z_t \) is \( \beta \)-mixing with geometrically decaying mixing numbers.

This proposition is based on recent results of Liebscher (2005) which highlight relations between \( \beta \)-mixing and geometric ergodicity. To be able to apply these results we need assumptions not needed in the case of stationary initial values. Our assumption (a) is slightly stronger than \( V_X \)-geometric ergodicity of \( P_X(\cdot, \cdot) \), but is implied by the so-called drift criterion which is a standard tool used to obtain geometric ergodicity (see Meyn and Tweedie (1993, Theorem 15.0.1)). A counterpart of our third assumption (c) was already needed in Proposition 1. This assumption is not very restrictive either in that it is automatically satisfied by the general model (4)–(5) and, therefore, by models we are mainly interested in. We state this as a lemma.

Lemma 1 Condition (c) of Proposition 2 is redundant for the model (4)–(5).

Assumptions (a) and (c) also imply that \( Z_t \) is \( V \)-geometrically ergodic with \( V(\cdot) = V_X(\lambda(\cdot)) \) and, for this case, the argument given in Meyn and Tweedie (1993, discussion following Theorem 16.1.5) could be employed to establish the strong mixing of \( Z_t \). This argument assumes condition (b) and suggests that it is also necessary in our case.

It may also be noted that the conditions of Proposition 2 ensure finiteness of certain moments. For instance, it is straightforward to establish that \( \sup_{t \geq 1} E_n[v(Z_t)] < \infty \) for any function \( v \) such that \( |v(\cdot)| \leq V_Z(\cdot) \), and that these moments converge to the ones taken with respect to the stationary distribution \( \pi_Z \) at a geometric rate.

3 GARCH and ACD models

As in Carrasco and Chen (2002), Propositions 1 and 2 can be applied to first order GARCH models. In addition to various GARCH(1,1) models Carrasco and Chen (2002) also considered higher-order GARCH models as well as examples of ACD models and autoregressive stochastic volatility models. We do not consider stochastic volatility models because for them similar results can be found in Genon-Catalot, Jeantheau, and Larédou (2000). However, as an extension of previous work we present both GARCH models and ACD models as special cases of a general model which even includes the GARCH–M model. As far as we know, these are the first results on geometric ergodicity and mixing obtained for the GARCH–M model (for the ARCH–M model similar results were obtained by Masry and Tjøstheim (1995)).

We consider a special case of the model (4)–(5) with \( Y_t \) and \( X_t \) real valued and \( X_t \) positive. Specifically, the model is defined by

\[
Y_t = f_{y1}(X_t) + f_{y2}(X_t)\varepsilon_t
\]

\[
X_t = f_{x1}(X_{t-1}) + f_{x2}(Y_{t-1} - f_{y1}(X_{t-1}), X_{t-1}),
\]
where the \( \varepsilon_t \) are i.i.d. and independent of \((Y_s, X_s), s < t\), and \( f_{y1}, f_{y2}, f_{x1}, \) and \( f_{x2} \) are Borel measurable functions to be described in detail shortly. The analog of equation (6) is obtained by substituting \( Y_{t-1} \) from (9) into (10), yielding

\[
X_t = f_{x1}(X_{t-1}) + f_{x2}(f_{y2}(X_{t-1})\varepsilon_{t-1}, X_{t-1}).
\] (11)

A model formulated in this way incorporates various GARCH and ACD models. In the GARCH context, \( f_{y1} \) is the conditional mean function whereas \( f_{y2} \) is used to model the conditional variance. In the ACD context, \( f_{y2} \) represents the conditional mean of \( Y_t \) and \( f_{y1} \) is omitted. Concrete examples will be given later.

For the development of our theory we make the following assumptions.

**Assumption 2**

(a) The i.i.d. random variables \( \varepsilon_t \) have a probability density function \( \phi_\varepsilon(\cdot) \) supported on \((\underline{\varepsilon}, \infty)\) and bounded away from zero on compact subsets of \((\underline{\varepsilon}, \infty)\). Here either \( \underline{\varepsilon} = 0 \) or \( \underline{\varepsilon} = -\infty \).

(b) The functions \( f_{x1} : \mathbb{R}_+ \to \mathbb{R}_+ \) and \( f_{x2} : (\underline{\varepsilon}, \infty) \times \mathbb{R}_+ \to \mathbb{R} \) are bounded on bounded subsets of their domains and, for some \( f > 0 \),

\[
\inf_{x \in \mathbb{R}_+, u \in (\underline{\varepsilon}, \infty)} (f_{x1}(x) + f_{x2}(u, x)) = f.
\]

(c) There exists a real number \( a \in [0, \infty) \) such that \( f_{x1}(x) \leq ax + o(x) \) as \( x \to \infty \).

(d) The function \( f_{x2} \) satisfies the following three conditions.

\( (d_1) \) There exists an unbounded interval of \( \mathbb{R}_+ \) which is, for all \( x > 0 \), contained in the image set \( f_{x2}((\underline{\varepsilon}, \infty), x) \).

\( (d_2) \) For all \( x > 0 \), the function \( f_{x2}(\cdot, x) \) is continuous from the right (or alternatively, continuous from the left).

\( (d_3) \) There exists a real number \( R > 0 \) such that, for \( u > R \) and all \( x > 0 \),

\( f_{x2}(u, x) \) is continuous and monotonically increasing, and the related inverse function \( f_{x2}^{-1}(v, x) \) has a partial derivative \( \partial f_{x2}^{-1}(v, x)/\partial v \) which is bounded away from zero on compact subsets of its domain.

(e) There exists a Borel measurable function \( b : (\underline{\varepsilon}, \infty) \to \mathbb{R}_+ \), nonconstant and continuous on some open set, and a real number \( c \in [0, \infty) \) such that

\( f_{x2}(f_{y2}(x)\varepsilon_t, x) \leq xb(\varepsilon_t) + c \) for all \( x \in \mathbb{R}_+ \). Furthermore, \( E[b(\varepsilon_t)]^k < \infty \) for some \( k \in \mathbb{R}_+ \).

(f) The function \( f_{y2} : \mathbb{R}_+ \to \mathbb{R}_+ \) is bounded on bounded subsets of its domain and bounded away from zero for \( x \geq f \).

Assumption 2(a) is satisfied in most applications. The case \( \underline{\varepsilon} = 0 \) is typical in ACD models, while in GARCH models \( \underline{\varepsilon} = -\infty \). Restricting \( \underline{\varepsilon} \) to these two values is not essential for the development of the theory, but from a practical point of view there is little need for generalizing this. In Assumptions 2(b) and (f) the functions \( f_{x1}, f_{x2}, \)
and $f_{x2}$ are assumed to be bounded on bounded subsets of their domains, while in Assumptions 2(a), (b), (d), and (f) the functions $\phi_x$, $f_{x1} + f_{x2}$, $\partial f_{x2}^2 / \partial v$, and $f_{x2}$, respectively, are assumed to be suitably bounded away from zero. These assumptions are hardly restrictive in practice. Especially the latter assumptions, though, suggest that some attention needs to be paid to the definitions of these functions.

Assumption 2(b) also ensures that the process $X_t$ is always positive. Note that there is more than one way to define the functions $f_{x1}$ and $f_{x2}$ without violating equations (10) and (11). In most cases it is natural to choose the functions $f_{x1}$ and $f_{x2}$ such that both of them are always positive. However, as a subsequent example shows, it is useful to be more flexible and only require that the sum $f_{x1} + f_{x2}$ is positive.

The conditions restricting the functions $f_{x1}$ and $f_{x2}$ in Assumptions 2(c) and (e), respectively, essentially restrict $X_t$ to depend on its past value at most in a linear fashion when arbitrarily large past values are of concern. This will be crucial in proving the geometric ergodicity of $X_t$. Similar assumptions have also been used in previous proofs for geometric ergodicity (see Lanne and Saikkonen (2005) for both Assumption 2(c) and (e) and Masry and Tjøstheim (1995), Lu (1998), and Lu and Jiang (2001), among others for Assumption 2(c)). Further conditions on the constant $a$, the function $b(\cdot)$, and moments of the random variables $\varepsilon_t$ will be imposed later. It will prove beneficial to have the values of the constant $a$ and the function $b(\cdot)$ as small values as possible.

Due to the very general nonlinear structure we wish to accommodate for, the conditions imposed on the function $f_{x2}$ in Assumption 2(d) are on the whole somewhat involved. The validity of these conditions can still be straightforwardly checked for many GARCH and ACD models, as our subsequent examples show. Often one can also use the following simple lemma (whose proof is omitted) to verify Assumptions 2(d1), (d2), and (d3).

**Lemma 2** Suppose that, for all $x > 0$, the function $f_{x2}(\cdot, x)$ is (1) surjective, (2) continuous, and (3) both monotonically increasing and continuously differentiable on $(l, \infty)$, where $l \geq 0$. Then Assumptions 2(d1), (d2), and (d3) hold.

Assumption 2(d) enables us to prove our results without knowing anything about, not even existence of, the conditional density of $X_t$ given $X_{t-1} = x$. In previous proofs of geometric ergodicity it has been quite typical to make explicit use of this conditional density and its properties (cf., e.g., Lu (1998) and Lanne and Saikkonen (2005)). While often straightforward this approach can sometimes be rather awkward to use and then our general conditions can be very convenient.

As indicated in the Introduction, our assumptions cover nonlinear specifications not covered by related previous studies. For instance, Carrasco and Chen (2002) and Straumann and Mikosch (in press) both give conditions for the ergodicity of rather general classes of GARCH models which, however, do not contain some of the nonlinear models to be discussed below. In particular, some smooth transition GARCH models (see Example 4 below) in which $f_{x1}$ and/or $f_{x2}$ in (11) are nonlinear functions of $X_{t-1}$ cannot be handled in the framework of Carrasco and Chen (2002). The
Ergodicity, mixing, and existence of moments of a class of Markov models

Theorem 3
Suppose that the assumptions of Theorems 1 and 2(a) (or 2(b)) are satisfied, and that \( Z \) is initialized from \( Z_0 = (Y_0, X_0) \) with a distribution \( \mu \) such that \( E_{\mu}[V_X(F_x(X_0, Y_0))] < \infty \), where \( F_x(\cdot, \cdot) \) denotes the function on the right hand side of (10). Then \( Z_t \) is \( \beta \)-mixing with geometrically decaying mixing numbers.

As discussed in the context of the more general model (4)–(5), \( X_t \) can be viewed as a separate Markov chain generated by (11) and with the associated transition probability measure defining the counterpart of \( P_X(\cdot, \cdot) \) in Assumption 1. The following theorem shows that, from this perspective, \( X_t \) is \( V \)-geometrically ergodic, as required for its counterpart in Proposition 1.

Theorem 1
Consider \( X_t \) as a Markov chain generated by (11) and, in addition to Assumption 2, suppose that \( E \left[ (a + b \varepsilon_t)^k \right] < 1 \). Then \( X_t \) is \( V \)-geometrically ergodic with \( V_X(x) = 1 + x^k \).

To be able to apply Proposition 1 and obtain useful results for the joint process \( (Y_t, X_t) \), concrete assumptions about the functions \( f_{y1} \) and \( f_{y2} \) are needed. In most applications of GARCH or ACD models, the function \( f_{y2} \) is assumed to be a power function, that is, \( f_{y2}(x) = x^{1/d} \) for some positive real number \( d \). This is also the assumption we will make. In the context of ACD models, the function \( f_{y1} \) is always assumed to be 0. In GARCH models, the most common specification for the ‘in-mean part’ has also been a power function, say \( f_{y1}(x) = \mu_0 + \mu_1 x^{1/e} \) for some positive real number \( e \). We will assume slightly less, only dominance by such a function. The following theorem gives an ergodicity result for the joint process \( (Y_t, X_t) \) for these cases.

Theorem 2
Suppose that the assumptions of Theorem 1 are satisfied and that \( f_{y2}(x) = x^{1/d} \) where \( d \in \mathbb{R}_+ \). (a) If \( f_{y1}(x) = 0 \) and \( E[|\varepsilon_t|^{dk}] < \infty \), where \( k \) is as in Assumption 2, then \( Z_t \) is \( V \)-geometrically ergodic with \( V_Z(y, x) = 1 + |y|^{dk} + x^k \). (b) If \( |f_{y1}(x)| \leq \mu_0 + \mu_1 x^{1/e} \), where \( e \in \mathbb{R}_+ \), \( \mu_0, \mu_1 \geq 0 \), \( \min(d, e)k \geq 1 \), \( E[|\varepsilon_t|^\min(d, e)k] < \infty \), and \( k \) as in Assumption 2, then \( Z_t \) is \( V \)-geometrically ergodic with \( V_Z(y, x) = 1 + |y|^{\min(d, e)k} + x^k \).

As already noted after Proposition 1, the \( V \)-geometric ergodicity of \( X_t \) implies the \( V \)-geometric ergodicity of \( Z_t \) with \( V(y, x) = V_X(x) \). A drawback of this choice of the function \( V \) is that nothing can be concluded about the moments of \( Y_t \). The usefulness of being able to use a more general choice of the function \( V \) becomes clear in Theorem 2 where results on the existence of moments of the stationary distribution of \( Y_t \) are also obtained. Our next result applies Proposition 2 and provides conditions which guarantee that the joint process \( (Y_t, X_t) \) is \( \beta \)-mixing.
In summary, Theorems 1–3 establish the $V$–geometric ergodicity and $\beta$–mixing for the general model defined by equations (9) and (10) or equation (11). Existence of moments also readily follows from the stated conditions (in particular, $\sup_{t \geq 1} E_\mu [Y_t^{dk}] < \infty$ (or $\sup_{t \geq 1} E_\mu [\min\{d,e\}k] < \infty$), and $\sup_{t \geq 1} E_\mu [X_t^k] < \infty$, and these moments converge to the ones taken with respect to the stationary distribution $\pi_Z$ at a geometric rate).

Concrete examples where Theorems 1–3 apply are discussed now. Depending on which formulation has been more common in the literature, the structure of each model is described by using either equation (10) or equation (11). For convenience, all the examples are summarized in Tables 1 and 2 where choices of the relevant functions and constants assumed in the preceding results are also provided. Because typical choices of the function $f_{\nu 1}$ were already discussed $f_{\nu 1}(\cdot) = 0$ is here assumed, so only GARCH and ACD models are considered. The validity of Assumption 2 can be rather straightforwardly verified for most of the considered models (see, however, some remarks in Examples 3 and 4 below). The form the condition $E \left( (a + b(\varepsilon_t))^k \right) < 1$ of Theorem 1 takes in these cases is also displayed in Table 1 with $k = 1$. The parameter restrictions implied by this condition agree in each case with the corresponding conditions reported in earlier literature.

Example 1 (The GARCH–family of Hentschel (1995) and the ACD–family of Fernandes and Grammig (2006)). Consider the family of GARCH models of Hentschel (1995), which can be written as (see eq. (A.2) and (A.3) of Hentschel (1995))

$$Y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + (\alpha \lambda f_{\nu}(\varepsilon_{t-1}) + \beta) \sigma_{t-1}^2$$

$$f(\varepsilon_{t-1}) = |\varepsilon_{t-1} - b| - c(\varepsilon_{t-1} - b),$$

where we assume that $b \in \mathbb{R}$, $|c| \leq 1$, and the remaining parameters take positive values. Defining $X_t = \sigma_t^2$ we arrive at a formulation written in the form of (9) and (11) as $Y_t = X_t^{1/\lambda} \varepsilon_t$ and $X_t = \omega + \alpha \lambda X_{t-1} f_{\nu}(\varepsilon_{t-1}) + \beta X_{t-1}$. In addition to the conventional linear GARCH model, this family also nests several other popular GARCH models (see Hentschel (1995) for a list). For brevity, the abbreviation BC–GARCH model is used in Tables 1 and 2 (here BC is due to the Box-Cox transformation used in the formulation of the model). Fernandes and Grammig (2006) consider a family of ACD models analogous to Hentschel’s family of GARCH models. This family can be defined with exactly the same equations (12) where, to use notation more conventional in the ACD–literature, $\sigma$ might be replaced with $\psi$.

Example 2 (The integrated GARCH model). Choosing $f_{\nu}(\varepsilon_{t-1}) = \varepsilon_{t-1}^2$, $\alpha \lambda = 1 - \beta$, and $\lambda = 2$ shows that the IGARCH model is a special case of the BC–GARCH model. Our results also hold for the IGARCH model but, unlike for all

\footnote{Hentschel (1995) also considers a slightly different formulation which includes the case $\lambda = 0$. We do not discuss this case.}
other cases, with the condition $E \left[ (a + b(\epsilon_t))^k \right] < 1$ only holding for $k < 1$. To see this, notice that now $E[a + b(\epsilon_t)] = E[\beta + (1 - \beta) \epsilon_t^2] = 1$ so that strict concavity and Jensen’s inequality give $E[(a + b(\epsilon_t))^k] < (E[a + b(\epsilon_t)])^k = 1$ for $0 < k < 1$. Thus, for the IGARCH model Theorem 3 should be applied with $k < 1$ and $d = 2$ implying that $Y_t$ has finite moments of orders smaller than 2. This is consistent with the well-known fact that the IGARCH process (that is $Y_t$) has a strictly stationary but not a second order stationary solution (see Nelson (1990) and note that the condition $E \left[ \ln (\alpha \epsilon_t^2 + \beta) \right] < 0$ used in that paper is automatically satisfied for the IGARCH(1,1) model). Previously, properties of the IGARCH process were also studied by Ding and Granger (1996) who demonstrated its short memory nature by showing that an ‘approximate’ autocorrelation function of $Y_t^2$ decays to zero at a geometric rate. Our results make this point more rigorous by showing that the process $Y_t$ is $\beta$–mixing with geometrically decaying mixing numbers.

Example 3 (Threshold models). A popular non-linear GARCH model is the GJR–GARCH model of Glosten, Jagannathan, and Runkle (1993), where the equation corresponding to (10) is $X_t = \omega + (\alpha + \alpha^* (Y_{t-1} > 0))Y_{t-1}^2 + \beta X_{t-1}$. Here $1(\cdot)$ is the indicator function. A restricted version of this model is nested in the family of Hentschel (1995). An ACD model resembling the GJR–GARCH model, and not nested by the family of ACD models of Fernandes and Grammig (2006), is the threshold ACD (or TACD) model of Zhang, Russell, and Tsay (2001). In a simple version of this model, the equation corresponding to (10) is given by

$$X_t = \omega_j + \alpha_j Y_{t-1} + \beta_j X_{t-1}, \quad \text{if} \quad Y_{t-1} \in [r_{j-1}, r_j), \quad j = 1, \ldots, J,$$

where $0 = r_0 < r_1 < \cdots < r_J = \infty$ are the threshold values, and the remaining parameters satisfy $\omega_j > 0, \alpha_j > 0, \text{and} \beta_j > 0$. Verifying the validity of Assumption 2 for this model is more involved than for the preceding models. Details are therefore presented in the Appendix. Here we only mention that it is convenient to express the model in a form in which the counterpart of the function $f_{\epsilon^2}$ can take negative values. Note also that in this case the moment condition $E \left[ (a + b(\epsilon_t))^k \right] < 1$ assumed in Theorem 1 takes a somewhat complicated form (see the Appendix). In Table 1 we therefore report a parameter restriction which implies the validity of this condition and also agrees with the sufficient condition for geometric ergodicity obtained by Zhang, Russell, and Tsay (2001).

Example 4 (Smooth transition models). Smooth transition GARCH and ACD models are also examples of models not nested in the families of Hentschel (1995) or Fernandes and Grammig (2006). The GARCH versions were introduced by Hagerud (1996) and González-Rivera (1998), and discussed by Lundbergh and Teräsvirta (2002) and Lanne and Saikkonen (2005), while the ACD analog was introduced by Meitz and Teräsvirta (2006). To obtain a general framework covering all these models, let $G_1$ and $G_2$ be functions with range $[0,1]$, and $\omega > 0, \alpha > 0, \beta > 0, \omega^* > 0, \omega + \omega^* > 0, \alpha + \alpha^* > 0, \text{and} \beta + \beta^* > 0$. In the GARCH variant, the equation corresponding to (10) takes the form

$$X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1} + (\omega^* + \alpha^* Y_{t-1}^2)G_1(Y_{t-1}) + (\omega^* + \beta^* X_{t-1})G_2(X_{t-1}).$$
The ACD variant is otherwise similar except that on the right hand side $Y_{t-1}$ is twice replaced by $Y_{t-1}$. For Assumption 2(d_2) to be satisfied we need to assume that the function $G_1$ is continuous from the left (or from the right). This, however, is not restrictive, because in practice $G_1$ is usually continuous or an indicator function of an interval with the latter possibility relevant for threshold type variants of the model. A sufficient condition for Assumption 2(d_3) to hold is that for large values of $y$ the function $G_1(y)$ is differentiable and $G_1'(y) = o(y^{-2})$ as $y \to \infty$. This condition is satisfied in the two typical examples where $G_1$ is the cumulative distribution function of the logistic distribution or normal distribution. For the function $G_2$ much less needs to be assumed. For convenience, we may assume that the limit $\lim_{x \to \infty} G_2(x)$ exists, in which case the constant $a$ in Table 2 has the stated form. Note that to satisfy the Lipschitz condition of Straumann and Mikosch (in press, Th. 3.1), rather complicated restrictions need to be imposed on the model parameters and on the functions $G_1$ and $G_2$. These restrictions are also quite stringent. For instance, models in which the function $G_2$ has a continuous derivative attaining a value large enough at some point are ruled out. Such models can still satisfy our Assumption 2.

As indicated earlier, the validity of Assumption 2(d) is relatively straightforward to check even for rather complicated nonlinear models. At least for some of the models discussed above alternative approaches, which require deriving the conditional density of $X_t$ given $X_{t-1} = x$ and checking that it has suitable properties, can be cumbersome. This may be the case, for instance, if one has a smooth transition GARCH model with the function $G_1$ not monotonically increasing.

4 Conclusion

In this paper we have studied a general Markov model which contains an observable and an unobservable or hidden component. We gave conditions under which the $V$–geometric ergodicity of the hidden component viewed as a Markov chain of its own is inherited by the joint process formed of the two components. Conditions for $\beta$–mixing and existence of moments for the joint process were also obtained.

Results obtained for our general Markov model were applied to a wide class of models which includes as special cases many first order GARCH, GARCH–M, and ACD models with possibly complicated nonlinear structures. As our emphasis was on allowing for nonlinearities, we only considered first order models, which are also often found adequate in practice. Due to the very general nature of the employed assumptions, the results obtained for these models should be straightforward to apply. Compared to previous counterparts they appear especially convenient for models such as smooth transition GARCH models or their ACD versions where highly nonlinear structures have been considered. Extensions of our results to general higher-order GARCH, GARCH–M, and ACD models forms an interesting topic for future research.
Ergodicity, mixing, and existence of moments of a class of Markov models

Model equations | Condition in Theorem 1 with $k = 1$
--- | ---
GARCH $X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1}$ | $\alpha + \beta < 1$
ACD $X_t = \omega + \alpha Y_{t-1} + \beta X_{t-1}$ | $\alpha + \beta < 1$
GJR–GARCH $X_t = \omega + (\alpha + \alpha^* 1(Y_{t-1} > 0)) Y_{t-1}^2 + \beta X_{t-1}$ | $\alpha + \alpha^*/2 + \beta < 1 \ (1)$
TACD $X_t = \omega + \alpha_j Y_{t-1} + \beta_j X_{t-1}$ | $\alpha_j + \beta_j < 1 \ (2)$
ST–GARCH $X_t = \omega + \beta X_{t-1} + (\omega^* + \alpha Y_{t-1}) G_1(Y_{t-1}) + (\omega^* + \beta X_{t-1}) G_2(X_{t-1})$ | $\alpha + \max\{\alpha^*, 0\} + \beta + \beta^* G_2(\infty) < 1 \ (3)$
ST–ACD $X_t = \omega + \beta X_{t-1} + (\omega^* + \alpha Y_{t-1}) G_1(Y_{t-1}) + (\omega^* + \beta X_{t-1}) G_2(X_{t-1})$ | $\alpha + \max\{\alpha^*, 0\} + \beta + \beta^* G_2(\infty) < 1 \ (3)$
BC–GARCH $X_t = \omega + \beta X_{t-1} + \alpha \lambda X_{t-1} f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda} \varepsilon_t$ | $E[\beta + \alpha \lambda f^\nu(\varepsilon_t)] < 1$
BC–ACD $X_t = \omega + \beta X_{t-1} + \alpha \lambda X_{t-1} f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda} \varepsilon_t$ | $E[\beta + \alpha \lambda f^\nu(\varepsilon_t)] < 1$

Table 1: Summary of the discussed examples: Model equations and the form the condition $E[(a + b(\varepsilon_t))^k] < 1$ of Theorem 1 takes with $k = 1$.

Notes: (1) Assuming $\varepsilon_t$ has a symmetric distribution. (2) This is a condition implying the validity of $E[(a + b(\varepsilon_t))] < 1$. (3) $G_2(\infty)$ is used as a shorthand notation for $\lim_{x \to \infty} G_2(x)$. 

Notes: (1) Assuming $\varepsilon_t$ has a symmetric distribution. (2) This is a condition implying the validity of $E[(a + b(\varepsilon_t))] < 1$. (3) $G_2(\infty)$ is used as a shorthand notation for $\lim_{x \to \infty} G_2(x)$. 

Notes: (1) Assuming $\varepsilon_t$ has a symmetric distribution. (2) This is a condition implying the validity of $E[(a + b(\varepsilon_t))] < 1$. (3) $G_2(\infty)$ is used as a shorthand notation for $\lim_{x \to \infty} G_2(x)$. 

Notes: (1) Assuming $\varepsilon_t$ has a symmetric distribution. (2) This is a condition implying the validity of $E[(a + b(\varepsilon_t))] < 1$. (3) $G_2(\infty)$ is used as a shorthand notation for $\lim_{x \to \infty} G_2(x)$.
<table>
<thead>
<tr>
<th>$f_{y2}(x)$</th>
<th>$f_{x1}(x)$</th>
<th>$f_{x2}(f_{y2}(x), x)$</th>
<th>$a$</th>
<th>$b(\varepsilon)$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>$x^{1/2}$</td>
<td>$\omega + \beta x$</td>
<td>$\alpha g^2$</td>
<td>$\beta$</td>
<td>$\alpha \varepsilon^2$</td>
</tr>
<tr>
<td>ACD</td>
<td>$x$</td>
<td>$\omega + \beta x$</td>
<td>$\alpha g$</td>
<td>$\beta$</td>
<td>$\alpha \varepsilon$</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>$x^{1/2}$</td>
<td>$\omega + \beta x$</td>
<td>$(\alpha + \alpha^* 1(y &gt; 0))g^2$</td>
<td>$\beta$</td>
<td>$(\alpha + \alpha^* 1(\varepsilon &gt; 0))\varepsilon^2$</td>
</tr>
<tr>
<td>TACD</td>
<td>$x$</td>
<td>$\beta_j x$</td>
<td>$\omega_j + \alpha_j y$</td>
<td>$\beta_j$</td>
<td>$\alpha_j \varepsilon$</td>
</tr>
<tr>
<td>ST–GARCH</td>
<td>$x^{1/2}$</td>
<td>$\omega + \beta x$</td>
<td>$\omega^* G_1(y)$</td>
<td>$\beta$</td>
<td>$(\alpha + \max {\alpha^*, 0})\varepsilon^2$</td>
</tr>
<tr>
<td>ST–ACD</td>
<td>$x$</td>
<td>$\omega + \beta x$</td>
<td>$\omega^* G_1(y)$</td>
<td>$\beta$</td>
<td>$(\alpha + \max {\alpha^*, 0})\varepsilon$</td>
</tr>
<tr>
<td>BC–GARCH</td>
<td>$x^{1/\lambda}$</td>
<td>$\omega + \beta x$</td>
<td>$\alpha \lambda x f''(\varepsilon)$</td>
<td>$\beta$</td>
<td>$\alpha \lambda f''(\varepsilon)$</td>
</tr>
<tr>
<td>BC–ACD</td>
<td>$x^{1/\lambda}$</td>
<td>$\omega + \beta x$</td>
<td>$\alpha \lambda x f''(\varepsilon)$</td>
<td>$\beta$</td>
<td>$\alpha \lambda f''(\varepsilon)$</td>
</tr>
</tbody>
</table>

Table 2: Summary of the discussed examples (continued): Choices of the relevant functions and constants. The function $f_{y1}$ is omitted as in every case $f_{y1}(x) = 0$. Notes: (1) Two different formulations, corresponding to equations (10) and (11), are used to achieve notational convenience. (2) $G_2(\infty)$ is used as a shorthand notation for $\lim_{x \to \infty} G_2(x)$. 
Appendix: Proofs

Validity of Assumption 1 for the model (4)–(5). First conclude from (4) that the conditional probability distribution of $Y_t$ given $X_t = x$ is defined by

$$\pi_{Y|X}(A | x) = \int 1 \{F_y(x, \zeta) \in A\} \, P_{\zeta}(d\zeta), \quad A \in \mathcal{B}(Y),$$

where $1(\cdot)$ is the indicator function and $P_{\zeta}(\cdot)$ signifies the probability distribution of $\zeta$. Similarly, the transition probability measure of the Markov chain $X_t$ defined by (6) is

$$P_X(x, A) = \int 1 \{G_x(x, \zeta) \in A\} \, P_{\zeta}(d\zeta), \quad A \in \mathcal{B}(X),$$

from which the corresponding $n$–step transition probability measure can be derived (cf. Meyn and Tweedie (1993, p. 78)). To derive the probability measure $\tilde{P}_X^2(z, \cdot)$ from this whereas equation (6) yields $X_2 = G_x(\tilde{z}, \zeta_1)$, $\tilde{z} \in X$. Thus, we get

$$\tilde{P}_X^2(z, A) = \int 1 \{G_x(\tilde{z}(z), \zeta) \in A\} \, P_{\zeta}(d\zeta) = \int 1 \{G_x(\tilde{z}, \zeta) \in A\} \, P_{\zeta}(d\zeta) = P_X(\tilde{z}, A), \quad A \in \mathcal{B}(X).$$

By induction it can be seen that $\tilde{P}_X^n(z, A) = P_X^{n-1}(\tilde{z}, A)$ for all $n \geq 2$ and, because the validity of equation (2) is straightforward to check, it follows that Assumption 1 applies to the model defined by (4) and (5) with $j = 1$ and the function $\lambda$ given by $\lambda = F_x$.

**Proof of Proposition 1.** The proof is based on ideas similar to those in the proof of Proposition 4 of Carrasco and Chen (2002). Set $\pi_Z(\cdot) = \pi_{Y|X}(\cdot | x) \, \pi_X(\cdot)$ where $\pi_X(\cdot)$ signifies the stationary probability measure related to a Markov chain with transition probability measure $P_X(\cdot, \cdot)$. First note that

$$\int_Z V_Z(z) \pi_Z(dz) = \int_X \pi_X(dx) \int_Y V_Z(y, x) \pi_{Y|X}(dy|x) \leq \int_X \pi_X(dx) c_V(x) < \infty,$$

where we have used the assumed condition $\int_Y V_Z(y, x) \pi_{Y|X}(dy|x) \leq c_V(x)$, for all $x \in \mathcal{X}$, and the $V_X$–geometric ergodicity of $X_t$. Then, for every $z_0 = (y_0, x_0) \in \mathcal{Y} \times \mathcal{X}$
and \( n > j \),

\[
\sup_{s:\|s\| \leq v_Z} \left| \int_{Y \times X} \left[ P^n_z(z_0, dz) - \pi_z(dz) \right] s(z) \right| = \sup_{s:\|s\| \leq v_Z} \left| \int_X \left[ \tilde{P}^n_X(z_0, dx) - \pi_X(dx) \right] \left( \int_Y \pi_{Y \mid X}(dy \mid x) s(y, x) \right) \right| \leq c \sup_{v:\|v\| \leq v_X} \left| \int_X \left[ P^{n-j}_X(\tilde{x}_0, dx) - \pi_X(dx) \right] v(x) \right|,
\]

(14)

where \( \tilde{x}_0 = \tilde{x}(z_0) = \lambda(z_0) \). Here the equality follows from (2) and the definition of \( \pi_Z(\cdot) \). In the inequality we have used Assumption 1(b) and the fact that, for any \( s \) with \( |s(\cdot)| \leq V_Z(\cdot) \),

\[
\left| \int_Y \pi_{Y \mid X}(dy \mid x) s(y, x) \right| \leq \int_Y \pi_{Y \mid X}(dy \mid x) |s(y, x)| \leq \int_Y \pi_{Y \mid X}(dy \mid x) V_Z(y, x) \leq c V_X(x).
\]

Because \( P_X(\cdot, \cdot) \) is assumed to be \( V_X \)-geometrically ergodic the last quantity in (14) can be bounded by a term of the form \( \rho^n M_{\tilde{Z}_n} \), where \( \rho < 1 \) and \( M_{\tilde{Z}_n} < \infty \). Thus, the same is true for the first quantity, implying that \( Z_t \) is \( V_Z \)-geometrically ergodic.

**Proof of Proposition 2.** By Proposition 2.4 of Liebscher (2005), \( Z_t \) is \( \beta \)-mixing with geometrically decaying mixing numbers if (i) \( E_\mu[V_X(\lambda(X, Y))] < \infty \) and (ii) \( Z_t \) is \( Q \)-geometrically ergodic in the sense of Liebscher (2005) with \( Q(\cdot) = V_X(\lambda(\cdot, y)) \).

Condition (i) holds by assumption (b). For condition (ii), we first need to show that \( E_\pi[Z_t(\lambda(X_t, Y))] < \infty \). This is obtained from (13) by replacing \( V_Z(\cdot) \) with \( V_X(\lambda(\cdot, y)) \) and using assumption (c) in conjunction with the \( V_X \)-geometric ergodicity of \( X_t \). As for the remaining part of condition (ii), notice that from (14) and (8) we find that

\[
\sup_{s:\|s\| \leq V_Z} \left| \int_{Y \times X} \left[ P^n_z(z_0, dz) - \pi_z(dz) \right] s(z) \right| \leq c \sup_{v:\|v\| \leq v_X} \left| \int_X \left[ P^{n-j}_X(\tilde{x}_0, dx) - \pi_X(dx) \right] v(x) \right| \leq \rho^n R V_X(\tilde{x}_0) = \rho^n R V_X(\lambda(x_0, y_0)).
\]

for some \( \rho \in (0, 1) \) and \( R < \infty \) (here \( \rho^{n-j} \) has been absorbed into \( R \)). Considering functions \( s(\cdot) \leq 1 \) completes the proof of condition (ii) (see the definition of \( Q \)-geometric ergodicity in Liebscher (2005, p. 671)).

**Proof of Lemma 1.** Consider the model (4)–(5) and suppose the assumptions of Proposition 2 are satisfied apart from (c). Recall that now Assumption 1 holds with the function \( \lambda \) given by \( \lambda = F_x \). This in conjunction with the definition of the
Ergodicity, mixing, and existence of moments of a class of Markov models

conditional probability distribution $\pi_{Y|X}(\cdot | x)$ and equations (4), (5), and (6) shows that

$$
\int_Y \pi_{Y|X}(dy | x) \lambda(x, y) = E[X_t(X_t, Y_t)] = E[X_t(X_{t+1}) | X_t = x] = \int_X P_X(x, dw) V_X(w).
$$

For simplicity, set $\int_X \pi_{X}(dw) V_X(w) = C$ and note that $C < \infty$ by the assumed $V_X$-geometric ergodicity of $P_X(\cdot, \cdot)$. Thus, using (8) with $n = 1$ we find that, for all $x \in X$,

$$
\left| \int_X P_X(x, dw) V_X(w) \right| \leq \left| \int_X P_X(x, dw) V_X(w) - \int_X \pi_{X}(dw) V_X(w) \right| + C \leq (\varrho R + C) V_X(x),
$$

where $\varrho R + C < \infty$. Combining the preceding inequalities yields the stated result. ■

**Proof of Theorem 1.** We use $\mu_{Leb}(\cdot)$ to signify the Lebesgue measure on $\mathbb{R}$ and $P_X(\cdot, \cdot)$ the transition probability measure obtained when $X_t$ is viewed as a separate Markov chain generated by (11). Due to the imposed assumptions, the state space of $X_t$ is $X = [l, \infty)$. The proof consists of showing that $X_t$ is irreducible and aperiodic, that an appropriate small set exists, and that the so-called drift condition is satisfied with the function $V_X$ (for definitions of these concepts, see Meyn and Tweedie (1993)). Irreducibility, the existence of a small set, and aperiodicity are first proven in Lemmas 3, 4, and 5, respectively.

**Lemma 3** If the assumptions of Theorem 1 hold there exist real numbers $l$ and $\bar{T}$ such that $(l, \bar{T}) \subset \mathcal{X}$ and the Markov chain $X_t$ is $\varphi$-irreducible with $\varphi(\cdot) = \mu_{Leb}(\cdot \cap (l, \bar{T}))$.

**Proof.** By assumption $E [(a + b(\varepsilon_t))^k] < 1$. Therefore we can choose an $\epsilon > 0$ such that

$$
E [(a + \epsilon + b(\varepsilon_t))^k] < 1. \quad (15)
$$

By Assumptions 2(c) and (e) we can now choose an $M_\epsilon \in \mathbb{R}_+$ such that

$$
f_{x2}(f_{y2}(x)\varepsilon_t, x) \leq x b(\varepsilon_t) + \frac{1}{2} x \quad (16)
$$

and

$$
f_{x1}(x) \leq ax + \frac{1}{2} cx \quad (17)
$$

for $x \in \mathcal{X}$ and $x > M_\epsilon$. Define the sets $S_{1\epsilon} = \{x \in \mathcal{X} : x > M_\epsilon\}$ and $S_{2\epsilon} = \{x \in \mathcal{X} : x \leq M_\epsilon\}$. Without loss of generality $M_\epsilon$ can be chosen large enough that $S_{2\epsilon}$ is nonempty. Clearly $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$. From (15) it follows that $a + \frac{1}{2} \epsilon < 1$, and hence
where by (15) the expectation is finite for arbitrary \( n \). The nonconstancy and continuity of \( b(\cdot) \) on some open set implies that the probability of \( \Omega_n \) is positive for every \( n \). Thus, on the event \( \Omega_n \),

\[
X_t^k \leq X_{t-1}^k \cdot E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right],
\]

(18)

where by (15) the expectation is less than 1.

Now choose an arbitrary \( x \in S_{1t} \), and denote \( X_0 = x \). Using (18) inductively we have, for arbitrary \( n \in \mathbb{Z}_+ \) and on the event \( \Omega_n \), that

\[
X_n^k \leq x^k \cdot \left\{ E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] \right\}^n
\]

(19)

as long as \( X_1, \ldots, X_{n-1} \in S_{1t} \). Since \( E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] < 1 \), the right-hand-side of (19) will eventually be less than or equal to \( M_k^k \) for such \( n \) when \( n \) is chosen large enough, and for such \( n \) we will have \( X_n \in S_{2k} \). Since the probability of the event \( \Omega_n \) is positive for every \( n \), we have thus completed the proof of I.

**Proof of II and III.** Since the functions \( f_{x1} \) and \( f_{x2} \) are bounded on bounded subsets of their domain there exist positive and finite real numbers \( M_1 \) and \( M_2 \) such that

\[
\sup_{x \in S_{2k}} f_{x1}(x) \leq M_1 \quad \text{and} \quad \sup_{x \in S_{2k}} f_{x2}(R, x) \leq M_2.
\]

(20)

Define \( \underline{l} = \max\{M_1 + M_2, M_\epsilon + 1\} \), and choose an arbitrary \( \overline{t} > \underline{l} \) (note that the fact \( \underline{l} > M_\epsilon \) is used only later in the proof of Lemma 5).
Now choose an arbitrary set \( A \) such that \( \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L})) > 0 \). Furthermore, choose an arbitrary \( x \in S_{2x} \). For the 1–step transition probability from \( x \) to \( A \) it holds that

\[
P(x, A) = \int_{-\infty}^{\infty} 1( f_{x1}(x) + f_{x2}(f_{y2}(x), x) \in A) \phi_x(\varepsilon) d\varepsilon
\]

\[
\geq \int_{R/f_{x2}(x)} 1( f_{x1}(x) + f_{x2}(f_{y2}(x), x) \in A) \phi_x(\varepsilon) d\varepsilon.
\]

According to Assumption 2(d) \( f_{x1}(x) + f_{x2}(f_{y2}(x), x) \) is monotonically increasing with respect to \( \varepsilon \) on the integration range, and thus, making a transformation of variables \( v = f_{x1}(x) + f_{x2}(f_{y2}(x), x) \), we have

\[
P(x, A) \geq \int_{\{v > f_{x2}(R, x) + f_{x1}(x)\}} 1(v \in A) \phi_x \left( \frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)} \right) \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} dv
\]

\[
\geq \int_{A \cap (\underline{L}, \overline{L})} \phi_x \left( \frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)} \right) \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} dv.
\]

The boundedness conditions for \( f_{x1}, f_{x2}, f_{y2}, \frac{\partial f_{x2}^{-1}(v, x)}{\partial v} \), and \( \phi_x \) imply that

\[
\inf_{x \in S_{2x}, \varepsilon \in A \cap (\underline{L}, \overline{L})} \phi_x \left( \frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)} \right) \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} \geq \epsilon^*_x,
\]

for some positive \( \epsilon^*_x \), and therefore \( P(x, A) \geq \epsilon^*_x \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L})) \). Because the set \( A \) can clearly be replaced by \( A \cap (\underline{L}, \overline{L}) \) both II and II are obtained from

\[
\inf_{x \in S_{2x}} P(x, A \cap (\underline{L}, \overline{L})) \geq \epsilon^*_x \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L})). \tag{21}
\]

**Proof of IV.** Choose an arbitrary set \( A \) such that \( \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L})) > 0 \), and an arbitrary \( x \in S_{1x} \). According to I, we can choose an integer \( n \) such that \( P^n(x, S_{2x}) > 0 \). Now, by the Chapman-Kolmogorov equation (Meyn and Tweedie (1993, Theorem 3.4.2, p. 67))

\[
P^{n+1}(x, A \cap (\underline{L}, \overline{L})) = \int_X P^n(x, dy) P(y, A \cap (\underline{L}, \overline{L}))
\]

\[
\geq \int_{S_{2x}} P^n(x, dy) P(y, A \cap (\underline{L}, \overline{L}))
\]

\[
\geq \int_{S_{2x}} P^n(x, dy) \epsilon^*_x \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L}))
\]

\[
= P^n(x, S_{2x}) \epsilon^*_x \mu_{\text{Leb}}(A \cap (\underline{L}, \overline{L}))
\]

\[
> 0,
\]

where the first inequality follows from the fact that \( S_{2x} \subset X \), and the second inequality follows from (21). This completes the proof of IV. 

Lemma 4 If the assumptions of Theorem 1 hold the set $S_{2\varepsilon}$ is small.

Proof. Equation (21) shows that equation (5.14) of Meyn and Tweedie (1993) holds with the measure $\epsilon_\ast \mu_{Leb}(\cdot \cap (L, T))$. Thus, the set $S_{2\varepsilon}$ is small by the definition of a small set. ■

Lemma 5 If the assumptions of Theorem 1 hold the Markov chain $X_t$ is aperiodic.

Proof. By Proposition A1.1 of Chan (1990), the aperiodicity of $X_t$ obtains if

$$\forall x \in A: (P(x, A) > 0 \text{ and } P^2(x, A) > 0)$$

for some small set $A$ such that $\varphi(A) > 0$. We shall show this holds with the set $(L, T)$.

To this end, it suffices to prove that for all open subsets $A$ of $S_{1\varepsilon}$ and for every $x \in A$, $P(x, A) > 0$ and $P^2(x, A) > 0$. Let $A$ be an arbitrary open subset of $S_{1\varepsilon}$ and $x \in A$ be arbitrary. Because $x > M_{\varepsilon}$, we have by (17) and (15), that $f_{\varepsilon}(x) \leq (a + \frac{1}{2}\varepsilon) x < x$, and furthermore that $x - f_{\varepsilon}(x) \geq x - (a + \frac{1}{2}\varepsilon) x = x(1 - a - \frac{1}{2}\varepsilon)$. As $M_{\varepsilon} > (1 - a - \frac{1}{2}\varepsilon)^{-1} \inf I$, we therefore have $x > f_{\varepsilon}(x) > \inf I$, where $I$ again denotes the interval in Assumption 2(d1). The same assumption now implies that there exists a $u$ such that $f_{\varepsilon}(x) + f_{\varepsilon}(u, x) = x$. Hence we can also find an $\xi \in (\xi, \infty)$ such that $f_{\varepsilon}(x) + f_{\varepsilon}(f_{\varepsilon}(x) \xi, x) = x$. Since the set $A$ is open, we can choose a $\delta > 0$ such that $(x - \delta, x + \delta) \subset A$, and the continuity from the right (alternatively, continuity from the left) of $f_{\varepsilon}(\cdot, x)$ ensures that for a such $\delta$, there exists an $\overline{\varepsilon} > \xi$ (alternatively, $\overline{\varepsilon} < \xi$) such that

$$\varepsilon \in (\xi, \overline{\varepsilon}) \Rightarrow f_{\varepsilon}(x) + f_{\varepsilon}(f_{\varepsilon}(x) \varepsilon, x) \in (x - \delta, x + \delta)$$

(alternatively, $\varepsilon \in (\overline{\varepsilon}, \xi)$). Thus, we can conclude that

$$P(x, A) \geq P(x, (x - \delta, x + \delta)) = \Pr(f_{\varepsilon}(x) + f_{\varepsilon}(f_{\varepsilon}(x) \varepsilon, x) \in (x - \delta, x + \delta)) \geq \Pr(\varepsilon \in (\xi, \overline{\varepsilon})) > 0,$$

where the second inequality follows from (23) and the third from the assumed positivity of $\phi_\varepsilon(\cdot)$. In addition, by the Chapman-Kolmogorov equation,

$$P^2(x, A) = \int_X P(x, dy)P(y, A) \geq \int_{(x - \delta, x + \delta)} P(x, dy)P(y, A) > 0.$$

Hence the assertion made is proven. Since this holds, in particular, for the set $(L, T)$ the condition (22) is established with $A = (L, T)$.

By Lemma 3, $\varphi((L, T)) > 0$. To establish that the set $(L, T)$ is small consider first the proof of parts II and III in Lemma 3 but with the set $S_{2\varepsilon}$ replaced by $(L, T)$. Repeating the arguments in that proof we can find an $\epsilon' > 0$ and an open interval $(L', T')$ such that the transition probabilities from $(L, T)$ to $(L', T')$ are positive and

$$\inf_{x \in (L, T)} P(x, A \cap (L', T')) \geq \epsilon_\ast \mu_{Leb}(A \cap (L', T')) > 0$$

(24)
whenever $\mu_{Leb}(A \cap (L', T)) > 0$. Equation (5.14) of Meyn and Tweedie (1993) is now satisfied with the measure $c_{i} \mu_{Leb}(\cdot \cap (L', T))$, and thus the set $(L, T)$ is small. ■

Finishing the proof of Theorem 1.

Given Lemmas 3, 4, and 5, it now suffices to show that condition (15.3) of Meyn and Tweedie (1993) holds with the function $V_X(x) = 1 + x^k$. This in turn is the case if there exist constants $c_1 > 0$ and $c_2 < \infty$ such that

$$E[V_X(X_t) \mid X_{t-1} = x] \leq (1 - c_1)V_X(x) + c_21(x \in S_{2\epsilon}) \quad \text{for all } x \in \mathcal{X}. \quad (25)$$

The expectation in (25) can be written as

$$E[1 + X_t^k \mid X_{t-1} = x] = 1 + E \left[ (f_{x_1}(x) + f_{x_2}(f_{y_2}(x) \varepsilon_{t-1}, x))^k \right].$$

Suppose first that $x \in S_{1\epsilon}$. As in the proof of part I of Lemma 3 we have

$$1 + E \left[ (f_{x_1}(x) + f_{x_2}(f_{y_2}(x) \varepsilon_{t-1}, x))^k \right] \leq 1 + x^k E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] = \left( 1 - \frac{x^k}{1 + x^k} \right) \left( 1 - \frac{1}{1 + x^k} \right) (1 + x^k).$$

Redefining $M_\epsilon$ if necessary we can without loss of generality assume that $M_\epsilon > 1$. Then $x > 1$ and $x^k/(1 + x^k) > 1/2$. Since $E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] < 1$, it follows that

$$1 + E \left[ (f_{x_1}(x) + f_{x_2}(f_{y_2}(x) \varepsilon_{t-1}, x))^k \right] \leq \left( 1 - \frac{1}{2} \right) \left( 1 - E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] \right) (1 + x^k).$$

Defining $c_1 = \frac{1}{2} \left( 1 - E \left[ (a + \epsilon + b(\varepsilon_{t-1}))^k \right] \right)$ shows that (25) holds for all $x \in S_{1\epsilon}$.

Suppose now that $x \in S_{2\epsilon}$. Then, by the first inequality in (20) and Assumption 2(e),

$$1 + E \left[ (f_{x_1}(x) + f_{x_2}(f_{y_2}(x) \varepsilon_{t-1}, x))^k \right] \leq 1 + E \left[ (M_1 + c + x b(\varepsilon_{t-1}))^k \right] \leq 1 + E \left[ (M_1 + c + M_\epsilon b(\varepsilon_{t-1}))^k \right] \leq \infty.$$

Defining $c_2 = 1 + E \left[ (M_1 + c + M_\epsilon b(\varepsilon_{t-1}))^k \right]$ and noting that $(1 - c_1)V(x)$ is always positive shows that (25) holds also for all $x \in S_{2\epsilon}$. Since $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$, this completes the proof of $V_X$-geometric ergodicity. ■
Proof of Theorem 2. The fact that the Markov chain \(Z_t = (Y_t, X_t)\) satisfies Assumption 1 follows from the discussion after this assumption. Also, \(X_t\) viewed as a separate Markov chain is \(V_X\)-geometrically ergodic by Theorem 1. Hence, by Proposition 1, it remains to be proven that \(\int_Y \pi_{Y|X}(dy|x)V_Z(y, x) \leq c V_X(x)\) for all \(x \in X\) and some \(c < \infty\).

The conditional probability distribution of \(Y_t\) given \(X_t = x\) is

\[
\pi_{Y|X}(dy|x) = \frac{1}{f_{y_2}(x)} \phi_x \left( \frac{y - f_{y_1}(x)}{f_{y_2}(x)} \right) dy.
\]

Thus, since \(V_X(x) = 1 + x^k\), part (a) follows by observing that

\[
\int_Y \pi_{Y|X}(dy|x)V_Z(y, x) = 1 + x^k + E|x^{1/d} \varepsilon_t|^d E|x^{1/d} \varepsilon_t|^d \leq (1 + x^k)(1 + E|\varepsilon_t|^d).
\]

Consider now part (b), and suppose first that \(d \leq e\). Similarly as above,

\[
\int_Y \pi_{Y|X}(dy|x)V_Z(y, x) = 1 + x^k + E|x^{1/d} \varepsilon_t + f_{y_1}(x)|^d
\]

\[
\leq 1 + x^k + \left( E|x^{1/d} \varepsilon_t|^d \right)^{1/d} \left( |f_{y_1}(x)|^d \right)^{1/d} + \left( |f_{y_1}(x)|^d \right)
\]

\[
\leq 1 + x^k + C_1 \left( E|x^{1/d} \varepsilon_t|^d + |f_{y_1}(x)|^d \right)
\]

\[
\leq 1 + x^k + C_1 \left( x^k E|\varepsilon_t|^d + (\mu_0 + \mu_1 x^{1/e})^d \right)
\]

\[
\leq 1 + x^k + C_1 \left( x^k E|\varepsilon_t|^d + C_2 (\mu_0^d + \mu_1^d x^{d/e}) \right)
\]

(26)

for some constants \(C_1\) and \(C_2\). Here the first inequality follows from the Minkowski’s inequality, and the second and fourth from the fact that for any \(p \geq 1\) there exists a constant \(C\) such that \((a + b)^p \leq C(a^p + b^p)\) for all non-negative numbers \(a\) and \(b\). In (26), \(x^{d/e} \leq \max\{1, x^k\} \leq (1 + x^k)\), and hence the expression in (26) is smaller than \(C_3(1 + x^k)\) for some constant \(C_3\). The case \(d > e\) can be proven in an analogous way.

Proof of Theorem 3. It was established in the proof of Theorem 2 that under current assumptions the Markov chain \(Z_t = (Y_t, X_t)\) satisfies the conditions of Proposition 1. The validity of condition (a) of Proposition 2 follows from Theorem 15.0.1 of Meyn and Tweedie (1993), because we have established the validity of their condition (15.3) in the proof of Theorem 1 (see equation (25)). Condition (b) is satisfied by assumption because in the present case \(\lambda = F_x\) whereas condition (c) is redundant by Lemma 1. The results follow by applying Proposition 2 with the functions \(V_Z\) from Theorem 2.

Validity of Assumption 2 for the TACD–model. Denoting \(R_j = [r_{j-1}, r_j]\) the TACD model can be written as \(X_t = \sum_{j=1}^J (\omega_j + \alpha_j Y_{t-1} + \beta_j X_{t-1}) 1(Y_{t-1} \in R_j)\).
Defining \( f_{x1}(x) = \beta_j x \) and \( f_{x2}(x, x) = \sum_{j=1}^{\ell} (\omega_j + \alpha_j x + \beta_j x) 1(x \in R_j) - \beta_j x \) we have \( X_t = f_{x1}(X_{t-1}) + f_{x2}(X_{t-1} \epsilon_{t-1}, X_{t-1}) \). The validity of conditions (b), (c), (d), and (d1) of Assumption 2 is rather clear. For condition (d1) it suffices to note that \( f_{x2}(r_{j-1}, \infty, x) \) and \( f_{x2}(r_{j-1}, \infty, x) \) for all \( x \).

For condition (e), note that \( \sum_{j=1}^{\ell} (\alpha_j x) 1(x \in R_j) \leq r_{j-1} \max \alpha_j \), \( \sum_{j=1}^{\ell} \omega_j 1(x \in R_j) \leq \max \omega_j \), and for any positive \( M \) (which is to be chosen shortly)

\[
\sum_{j=1}^{\ell} (\beta_j x) 1(x \in R_j) \leq x \max \beta_j 1(x < r_{j-1}) = x \max \beta_j 1(x < r_{j-1}) (1(x \leq M) + 1(x > M)) \leq M \max \beta_j + x \max \beta_j 1(x < r_{j-1}) 1(x > M) \leq M \max \beta_j + x \max \beta_j 1(\epsilon < r_{j-1}/M).
\]

Therefore

\[
f_{x2}(x, x) = \sum_{j=1}^{\ell} (\omega_j + \alpha_j x + \beta_j x) 1(x \in R_j) - \beta_j x \leq \max \omega_j + r_{j-1} \max \alpha_j + M \max \beta_j + x \max \beta_j 1(\epsilon < r_{j-1}/M) + \alpha_j x \leq (\max \omega_j + r_{j-1} \max \alpha_j + M \max \beta_j) + x (\alpha_j x + \max \beta_j 1(\epsilon < r_{j-1}/M))
\]

and, denoting \( c = (\max \omega_j + r_{j-1} \max \alpha_j + M \max \beta_j) \) and \( b(\epsilon) = (\alpha_j x + \max \beta_j 1(\epsilon < r_{j-1}/M)) \), we have established the validity of the inequality in condition (e).

It remains to be verified that the moment condition in (e) is satisfied. For this, we next establish that, for any \( k \geq 1 \), if \( E(\alpha_j x \epsilon_t + \beta_j)^k < 1 \), then \( E(a + b(\epsilon_t))^k < 1 \) (and thus also \( E(b(\epsilon_t)) < \infty \) in condition (e)). In addition to completing the verification of Assumption 2, this gives an easily verifiable condition which implies the validity of the moment restriction in Theorem 1. When \( k = 1 \), \( E(a + b(\epsilon_t)) = E(\alpha_j x \epsilon_t + \beta_j) + \max \beta_j E(\epsilon_t < r_{j-1}/M) \). By choosing \( M \) sufficiently large, the last term can be made arbitrarily small, and hence \( E(a + b(\epsilon_t)) < 1 \) for a suitable choice of \( M \).

Suppose now that \( k > 1 \). By Minkowski’s inequality we have

\[
E(a + b(\epsilon_t))^k = E((\alpha_j x \epsilon_t + \beta_j) + \max \beta_j 1(\epsilon < r_{j-1}/M))^k \leq \left( E(\alpha_j x \epsilon_t + \beta_j)^{1/k} + E(\max \beta_j 1(\epsilon < r_{j-1}/M))^{1/k} \right)^k.
\]

Here the second expectation can be written as \( E(\max \beta_j 1(\epsilon < r_{j-1}/M))^k = \max \beta_j^k \), \( E[1(\epsilon_t < r_{j-1}/M)] \) and hence, by choosing \( M \) sufficiently large, this term can be made small enough so that \( E(a + b(\epsilon_t))^k < 1 \). This completes establishing the validity of the conditions.
References


111
REFERENCES


A necessary and sufficient condition for the strict stationarity of a family of GARCH processes
A necessary and sufficient condition for the strict stationarity of a family of GARCH processes

Abstract

We consider a family of GARCH(1,1) processes introduced in He and Teräsvirta (1999a). This family contains various popular GARCH models as special cases. A necessary and sufficient condition for the existence of a strictly stationary solution is given.

This paper has been accepted for publication in Econometric Theory (Notes and Problems), and is printed in this thesis with permission from the Cambridge University Press.

The author thanks the editor, an anonymous referee, Pentti Saikkonen, and Timo Teräsvirta for useful comments.
1 Introduction

He and Teräsvirta (1999a) considered a general class of first-order GARCH models and examined the moment structure within this family. In their paper, the sequence of random variables \( \{ \varepsilon_t \}_{t=-\infty}^\infty \) belongs to this general class of GARCH(1,1) processes if

\[
\varepsilon_t = z_t h_t, \\
h_t^k = g(z_{t-1}) + c(z_{t-1}) h_{t-1}^k
\]  

(1)

where \( \{ z_t \} \) is a sequence of independent and identically distributed random variables with zero mean, \( k \) equals 1 or 2, and \( g_t = g(z_t) \) and \( c_t = c(z_t) \) are well-defined functions of \( z_t \). Furthermore, they assume that \( \Pr\{ h_t^k > 0 \} = 1 \).

Many GARCH(1,1) models are included in this family. For example, the choices \( k = 2, g_{t-1} = \alpha_0 \), and \( c_{t-1} = \beta + \alpha_1 z_{t-1}^2 \) yield the linear GARCH model of Bollerslev (1986). When \( k = 1, g_{t-1} = \alpha_0 \), and \( c_{t-1} = \beta + \alpha_1 |z_{t-1}| \) the absolute value GARCH model of Taylor (1986) and Schwert (1989) is obtained. For \( k = 2, g_{t-1} = \alpha_0 \), and \( c_{t-1} = \beta + (\alpha_1 + \omega I(z_{t-1})) z_{t-1}^2 \) (where \( I(z_{t-1}) = 1 \) if \( z_{t-1} < 0 \) and \( I(z_{t-1}) = 0 \) otherwise) the model reduces to the GJR–GARCH model of Glosten, Jagannathan, and Runkle (1993). For a more extensive list, see He and Teräsvirta (1999a).

He and Teräsvirta (1999a) give conditions for the existence of moments of arbitrary order for the GARCH process (1)–(2). In particular, they show that the conditions \( E[|\varepsilon_t|^{km}] < \infty \) and \( E[\varepsilon_t^m] < 1 \) are necessary and sufficient for the existence of the \( km \)th absolute moment of \( \varepsilon_t \). They also give an explicit formula for this moment under the stated conditions. He and Teräsvirta (1999b) consider a special case of the model (1)–(2), where it is assumed that \( g_{t-1} = \alpha_0 \) and \( c_{t-1} \) has a particular parametric form. The exponent \( k \) is no longer restricted to take one of the values 1 and 2, but is only assumed to be a positive real number. They show that the conditions \( E[|\varepsilon_t|^{2k}] < \infty \) and \( E[\varepsilon_t^2] < 1 \) are necessary and sufficient for the existence of \( E[|\varepsilon_t|^{2k}] \).

Ling and McAleer (2002) also considered the model (1)–(2) and complemented the results given in He and Teräsvirta (1999a, b). Ling and McAleer (2002) assume that \( k \) is a positive real number and show that if \( E[|\varepsilon_t|^{km}] < \infty \), \( E[g_{t}^{m}] < \infty \), and \( E[c_{t}^{m}] < 1 \) for some \( m \in (0, 1] \), then there exists a unique \( km \)th order stationary solution to (1)–(2), which is also strictly stationary and ergodic. Furthermore, under the conditions \( E[|\varepsilon_t|^{km}] < \infty \) and \( E[c_{t}^{m}] < \infty \) and assuming now that \( m \) is a positive integer, they show that the necessary and sufficient condition for the existence of the \( km \)th absolute moment of \( \varepsilon_t \), \( E[|\varepsilon_t|^{km}] \), is \( E[c_{t}^{m}] < 1 \) (in Theorem 2.2 of Ling and McAleer (2002) the exponent of \( g \) and \( c_t \) is \( km \) instead of \( m \), but this appears to be a typographical error; the exponent appearing in their proof is \( m \)).

A yet unresolved issue is the necessary and sufficient condition for the existence of a strictly stationary solution to (1)–(2). In the case of a linear GARCH(1,1) model of Bollerslev (1986), this condition was derived in Nelson (1990). Bougerol and Picard (1992) extended this result to the linear GARCH(p,q) process. Duan (1997) introduced the augmented GARCH(p,q) model, which overlaps with the model (1)–(2), and gave a sufficient condition for the existence of a strictly stationary solution; in
a number of cases the condition is also necessary. In all of these papers the condition is derived, in principle, using the theory of random matrices, and is formulated using the so called Lyapunov exponent.

In this short note we demonstrate that the same approach can be used to prove the necessity and sufficiency of a similar condition also in the case of the model (1)–(2). In fact, this readily follows from the results already given in Bougerol and Picard (1992) in the context of generalized autoregressive equations.

2 Main result

We consider the case in which the exponent $k$ is assumed to be a positive real number (it plays no role in the following proof). First note that $\{g_t\}$ and $\{c_t\}$ are sequences of independent and identically distributed random variables because $\{z_t\}$ is. We make the following assumption.

**Assumption 1** For all $t$, $g_t$ and $c_t$ are nonnegative and $g_t$ is strictly positive with nonzero probability.

This assumption is not very restrictive compared with the requirement $\Pr\{h_t > 0\} = 1$ made in He and Ter"asvirta (1999a). From a practical point of view, most of the models listed as special cases in He and Ter"asvirta (1999a) also satisfy this condition.

Rewriting (2) as

$$h_t^k = g_{t-1} + c_{t-1}h_{t-1}^k, \quad t \in \mathbb{Z}$$

makes it clear that the process $\{h_t^k\}$ follows an autoregressive equation in $\mathbb{R}^+$ with independent and identically distributed nonnegative coefficients $g_t$ and $c_t$. Conditions for strict stationarity in such a situation are discussed in Bougerol and Picard (1992). The following result gives necessary and sufficient conditions for the existence of a strictly stationary solution of (3). We use the notation $\ln^+(x) = \max\{\ln(x), 0\}$.

**Theorem 1** (Corollary of Theorem 3.2 of Bougerol and Picard (1992))

Suppose that Assumption 1 is satisfied and that $E[\ln^+(c_t)]$ is finite. If (3) has a strictly stationary nonnegative solution, then $E[\ln(c_t)] < 0$. Conversely, if $E[\ln^+(g_t)]$ is finite and $E[\ln(c_t)] < 0$, then for all $t \in \mathbb{Z}$, the series

$$h_t^k = g_{t-1} + \sum_{k=1}^{\infty} c_{t-1}c_{t-2} \cdots c_{t-k}g_{t-k-1}$$

converges a.s. and the process $\{h_t^k, t \in \mathbb{Z}\}$ is the unique strictly stationary solution of (3).

**Proof.** Assumption 1 ensures that condition (C) of Bougerol and Picard (1992) is satisfied (see ibid., pp. 122–123). In Theorem 3.2 of Bougerol and Picard (1992), the key condition for the existence of a strictly stationary solution is that the so called
top Lyapunov exponent associated to a sequence of certain matrices is negative. In
the present (univariate) case the top Lyapunov exponent equals

$$\inf_{t \in \mathbb{N}} E[(t + 1)^{-1} \ln(c_0 c_{-1} \cdots c_{-t})] = \inf_{t \in \mathbb{N}} \left\{ (t + 1)^{-1} \sum_{i=0}^{t} E[\ln(c_{-i})] \right\} = E[\ln(c_t)].$$

The stated result now follows from Theorem 3.2 of Bougerol and Picard (1992).

The strict stationarity of the process \( \{ \varepsilon_t \} \) follows from that of \( \{ z_t \} \) and \( \{ h_t^k \} \). In
the case of the linear GARCH(1,1) process, \( g_t \) is a constant and the condition for \( c_t \) has the form \( E[\ln(\beta + \alpha z_t^2)] < 0 \), a condition already derived in Nelson (1990). The
conditions for other members of the family of GARCH processes (1)-(2) are easily
derived from Theorem 1.

Nelson (1990, Th. 6) derived an explicit expression for the moment \( E[\ln(\beta + \alpha z_t^2)] \)
in the case of standard normal or Cauchy errors using functions standard in the
mathematical literature yet rather exotic in the econometric one. This derivation
relies on the existence of explicit integral formulas for the logarithm of a polynomial
(see the references in Nelson (1990)). Unfortunately, however, similar formulas for the
moment \( E[\ln(c_t)] \) do not seem to be available without making stringent assumptions
about the functional form of \( c(\cdot) \).
References


Stability of nonlinear AR–GARCH models
Stability of nonlinear AR–GARCH models

Abstract

This paper studies the stability of nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a nonlinear autoregression of order $p$ (AR($p$)) with the conditional variance specified as a nonlinear first order generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) model. Conditions under which the model is stable in the sense that its Markov chain representation is geometrically ergodic are provided. This implies the existence of an initial distribution such that the process is strictly stationary and $\beta$-mixing. Conditions under which the stationary distribution has finite moments are also given. The results cover several nonlinear specifications for both the conditional mean and conditional variance.

This paper is joint work with Pentti Saikkonen.

Material from this paper has been presented in seminars at Université catholique de Louvain, Université Libre de Bruxelles, and University of Oxford.
1 Introduction

This paper is concerned with the stability of nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a nonlinear autoregression of order p (AR(p)) with the conditional variance specified as a nonlinear first order generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) model. This time series model can be viewed as a Markov chain, and our study makes heavy use of the stability theory developed for Markov chains. We refer the reader to Meyn and Tweedie (1993) for a comprehensive account of the needed Markov chain theory.

The stability concept employed in the paper is that of geometric ergodicity, or more precisely, \( Q \)-geometric ergodicity as defined by Liebscher (2005). Geometric ergodicity is a useful property, for it implies the existence of an initial distribution which makes the Markov chain strictly stationary and \( \beta \)-mixing (or absolutely regular). The \( Q \)-geometric ergodicity is even more useful in that it implies that certain moments of the stationary distribution exist and, moreover, the \( \beta \)-mixing property also holds for a variety of nonstationary initial distributions. In this paper we give conditions under which the Markov chain associated with our AR–GARCH model is \( Q \)-geometrically ergodic and has moments of known order. An important consequence of these results is that usual limit theorems can be applied and, therefore, it becomes possible to develop a rigorous asymptotic estimation theory for these models.

Results similar to ours have previously been obtained for nonlinear homoskedastic autoregressions in Bhattacharya and Lee (1995), An and Huang (1996), An and Chen (1997), and Lee (1998) among many others. These results have been extended to allow for ARCH, but not GARCH, type conditional heteroskedasticity by Masry and Tjøstheim (1995), Lu (1998), Cline and Pu (1998), Cline and Pu (1999), Lu and Jiang (2001), Chen and Chen (2001), Saikkonen (2005), and Liebscher (2005). For related results for pure GARCH models, see Meitz and Saikkonen (2004) and the references therein. To the best of our knowledge, this paper provides the first practically applicable stability results for nonlinear autoregressive models with GARCH errors.

A major difficulty in establishing geometric ergodicity in the present context is to prove irreducibility and aperiodicity of the relevant Markov chain, which is typically required as a first step in the proof of geometric ergodicity. This difficulty may actually explain the aforementioned lack of related previous results. Our approach is to apply results on nonlinear state space models given in Meyn and Tweedie (1993, Chapter 7). This approach requires rather stringent smoothness assumptions about the nonlinear functions used to specify the conditional mean and conditional variance and, consequently, we are not able to handle threshold type nonlinearities characterized by discontinuous functions (see, e.g., Tong (1990) and Chen and Tsay (1993) for models for conditional mean and Glosten, Jaganathan, and Runkle (1993), and Rabemananjara and Zakoian (1993) for GARCH models). However, we are still able to cover a number of nonlinearities recently considered in both theoretical and applied literature.

A convenient feature of the assumptions needed to obtain our results is that most of them restrict the conditional mean and conditional variance of the model separately. Only one of our assumptions is common to both the conditional mean and conditional
variance and quite often this assumption can be straightforwardly checked. In such cases the verification of our assumptions reduces to separately checking the assumptions of a homoskedastic nonlinear autoregressive model and a pure GARCH model. As far as the conditional mean is concerned, our results apply to the functional-coefficient autoregressive model of Chen and Tsay (1993) which encompasses various well-known nonlinear autoregressive models such as the smooth transition autoregressive models (see Teräsvirta (1994), van Dijk, Teräsvirta, and Franses (2002), and the references therein). The conditional variance may be specified as the linear GARCH model of Bollerslev (1986) or even a GARCH model with a rather complicated nonlinear structure.

The rest of this paper is organized as follows. The model and the imposed assumptions are studied in Section 2. In Section 3 the main result of the paper is presented, and examples are given in Section 4. Section 5 concludes. Proofs of all the results are given in an Appendix.

2 Model and Assumptions

Let $y_t$, $t = 1, 2, \ldots$, be a real valued stochastic process generated by

$$y_t = f (y_{t-1}, \ldots, y_{t-p}) + h_t^{1/2} \varepsilon_t,$$

where $h_t$ is a positive function of $y_s$, $s < t$, and $\varepsilon_t$ is a sequence of (continuous) i.i.d. $(0, 1)$ random variables such that $\varepsilon_t$ is independent of $\{y_s, s < t\}$. The function $f$ is supposed to be nonlinear so that equation (1) defines a nonlinear autoregression with conditionally heteroskedastic errors. We assume that $h_t$, the conditional variance of $y_t$, is generated by a (possibly) nonlinear GARCH(1,1) process driven by regression errors. Specifically,

$$h_t = g(u_{t-1}, h_{t-1}),$$

where $g$ is a function to be described shortly and

$$u_t = y_t - f (y_{t-1}, \ldots, y_{t-p}).$$

From the definition of $u_t$ it is readily seen that $Z_t = [y_t \cdots y_{t-p} \ h_t] \overset{def}{=} [y_t' \ h_t']$ is a Markov chain on $\mathcal{Z} = \mathbb{R}^{p+1} \times \mathbb{R}_+$ (here and in what follows the notation $\mathbb{R}_+ = (0, \infty)$ is used). To make the Markov chain representation of $Z_t$ explicit, set

$$h (Z_{t-1}) = g(y_{t-1} - f (y_{t-2}, \ldots, y_{t-p}), h_{t-1})$$

and observe that then we can write

$$\begin{bmatrix}
  y_t \\
  y_{t-1} \\
  \vdots \\
  y_{t-p} \\
  h_t
\end{bmatrix}
= \begin{bmatrix}
  f (y_{t-1}, \ldots, y_{t-p}) \\
  y_{t-1} \\
  \vdots \\
  y_{t-p} \\
  h_t
\end{bmatrix} + \begin{bmatrix}
  0 \\
  0 \\
  \vdots \\
  0
\end{bmatrix},$$

$$h (Z_{t-1})^{1/2} \varepsilon_t$$

(5)
or, more briefly,

\[ Z_t = F (Z_{t-1}, \varepsilon_t) , \tag{6} \]

where the function \( F : \mathbb{Z} \times \mathbb{R} \to \mathbb{Z} \) is defined in an obvious way.

We set \( F_1 = F \) and, for \( k \geq 1 \), \( F_{k+1} (z, e_1, \ldots, e_{k+1}) = F (F_k (z, e_1, \ldots, e_k), e_{k+1}) \), where \( z \in \mathbb{R}^{p+1} \) and \( e_i \in \mathbb{R} \). Then, for any initial condition \( Z_0 = z_0 \) and any \( k \geq 1 \), \( Z_k = F_k (z_0, e_1, \ldots, e_k) \). Following Meyn and Tweedie (1993) we call \( \{ e_i \} \) a control sequence and \( z_k = F_k (z_0, e_1, \ldots, e_k) \) \((k = 1, 2, \ldots)\) the associated deterministic control model for the nonlinear state space model (6). Our analysis of the Markov chain \( Z_t \) makes use of this deterministic control model.

We make the following assumptions about the error term \( \varepsilon_t \) and the function \( f \).

We call a function smooth if its (partial) derivatives exist up to any order and are continuous.

**Assumption 1** The i.i.d. \((0, 1)\) random variables \( \varepsilon_t \) have a (Lebesgue) density which is positive and lower semicontinuous on \( \mathbb{R} \). Furthermore, for some real \( r \geq 1 \), \( \mathbb{E} [\varepsilon_t^{2r}] < \infty \).

**Assumption 2** The function \( f \) is of the form

\[ f (x) = a (x)' x + b (x), \quad x \in \mathbb{R}^p, \]

where the functions \( a : \mathbb{R}^p \to \mathbb{R}^p \) and \( b : \mathbb{R}^p \to \mathbb{R} \) are bounded and smooth.

Assumption 1 is mild and met in most applications where no bounds for the values of the considered process are assumed. Assumption 2 imposes a certain structure on the nonlinear function \( f \) which specifies the conditional expectation of the process. As mentioned in the introduction, similar structures have previously appeared in the functional-coefficient autoregressive models of Chen and Tsay (1993) and its special cases such as smooth transition autoregressive models. For these models the required smoothness assumption is also satisfied. This, as well as Assumption 1, is needed to make use of the results for nonlinear state space models in Meyn and Tweedie (1993, Chapter 7). For this reason we also need similar smoothness assumptions for the function \( g \) used to model the conditional variance. These assumptions are satisfied by several well-known models but, as discussed in the introduction, they rule out threshold type nonlinearities.

For later purposes it will be convenient to introduce some notation. For any integer \( p \geq 2 \) and \( x \in \mathbb{R}^p \) we define the \( p \times p \) matrix

\[
\bar{A}_p (x) = \begin{bmatrix}
x_1 & x_2 & \cdots & x_{p-1} & x_p \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}.
\]
Then, using the function \( a(x) \) in Assumption 2, set \( a(x) = [a_1(x) \cdots a_p(x)]' \) and define the \((p+1) \times (p+1)\) matrix

\[
A(x) = \bar{A}_{p+1} \left( [a(x)' 0]' \right) = \begin{bmatrix}
a_1(x) & a_2(x) & \cdots & a_p(x) & 0 \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}.
\]

With this notation the model takes the form

\[
Y_t = A(S'Y_{t-1})Y_{t-1} + \epsilon h(Z_{t-1})^{1/2} \varepsilon_t
\]

\[
h_t = h(Z_t),
\]

where \( \varepsilon = [1 \ 0 \ \cdots \ 0]' \ ( (p+1) \times 1 \) and \( S = [I_p : 0]' \ ((p+1) \times p) \). To be able to establish geometric ergodicity we need to restrict the matrix \( A(x) \). A general way to do this is provided by the following assumption where \( \mathcal{A}_* = \{ A(x) : x \in \mathbb{R}^p \} \).

**Assumption 3** There exists a matrix norm \( \| \cdot \|^{*} \) induced by a vector norm, also denoted by \( \| \cdot \|^{*} \), such that \( \| A \|^{*} \leq \rho \) for all \( A \in \mathcal{A}_* \) and some \( 0 < \rho < 1 \).

To make Assumption 3 operational in practice, two concrete cases are considered. For the first one we need the concept of the joint spectral radius of a (bounded) set of (square) matrices. To introduce this concept, let \( \mathcal{A} \) be a set of bounded square matrices and \( \mathcal{A}^k = \{ A_1 \cdots A_k : A_i \in \mathcal{A}, i = 1, \ldots, k \} \). Then the joint spectral radius of the set \( \mathcal{A} \) is defined by

\[
\rho (\mathcal{A}) = \limsup_{k \to \infty} \left( \sup_{A \in \mathcal{A}^k} \| A \|^{1/k} \right),
\]

where \( \| \cdot \| \) can be any matrix norm (the value of \( \rho (A) \) does not depend on the choice of this norm). If the set \( \mathcal{A} \) only contains a single matrix \( A \) then the joint spectral radius of \( A \) coincides with \( \rho (A) \), the spectral radius of \( A \). Several useful results about the joint spectral radius are given in the recent paper by Liebscher (2005) where further references can also be found.

Sufficient conditions for Assumption 3 can now be given.

**Lemma 1** Either of the following conditions is sufficient for Assumption 3 to hold.

(i) \( \rho (\mathcal{A}_*) < 1 \) or, equivalently, \( \rho (\mathcal{A}_1) < 1 \), where \( \mathcal{A}_1 = \{ A_1(x) : x \in \mathbb{R}^p \} \) with the \( p \times p \) matrix \( A_1(x) \) defined by deleting the last row and last column of \( A(x) \).

(ii) \( \sum_{j=1}^{p} \alpha_j < 1 \) or, equivalently, the roots of the characteristic polynomial \( \lambda^p - \alpha_1 \lambda^{p-1} - \cdots - \alpha_p = 0 \) are inside the unit circle, where \( \alpha_j = \sup_{x \in \mathbb{R}^p} |a_j(x)| \ (j = 1, \ldots, p) \).

As already indicated, Assumption 3 is needed to prove the geometric ergodicity of the Markov chain \( Z_t \). A similar assumption based on the concept of joint spectral radius or Lemma 1(i) was recently used by Liebscher (2005) who established
geometric ergodicity for various nonlinear autoregressive models. In these models conditional heteroskedasticity was also allowed but only of a limited nature. In particular, GARCH type or even ARCH type conditional heteroskedasticity was ruled out. A practical difficulty with the application of Lemma 1(i) is that the computation of the joint spectral radius is very computer-intensive unless the dimension of the matrix $A(x)$ is reasonably small (for a discussion, see Liebscher (2005)). In practice one should therefore consider $\rho(A_1)$ rather than $\rho(A_x)$. This computational difficulty has also been a motivation for the second part of Lemma 1 which gives the condition used by Chen and Tsay (1993) to provide a sufficient condition for geometric ergodicity in their functional-coefficient autoregressive models. The main advantage of this latter condition is its simplicity, for Liebscher (2005, Section 7) shows by an example that the condition based on the joint spectral radius can provide a larger region in the parameter space ensuring geometric ergodicity than the condition given in Lemma 1(ii).

The following assumption contains conditions which restrict the dynamics of the conditional variance process.

Assumption 4

(a) The function $g : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is smooth and, for some $g > 0$, 
$$\inf_{(u,x)} \in \mathbb{R} \times \mathbb{R}_+ g(u,x) = g.$$ 
(b) For all $x \in \mathbb{R}_+$, $g(u,x) \rightarrow \infty$ as $u \rightarrow \infty$.
(c) There exists a real number $h^* \in \mathbb{R}_+$ such that the sequence $h_k$ ($k = 1, 2, \ldots$) defined by $h_k = g(0,h_{k-1})$, $k = 1, 2, \ldots$, converges to $h^*$ as $k \rightarrow \infty$ for all $h_0 \in \mathbb{R}_+$. If $g(u,x) \geq h^*$ for all $u \in \mathbb{R}$ and all $x \geq h^*$ it suffices that this convergence holds for all $h_0 \geq h^*$.
(d) There exist nonnegative real numbers $a$ and $c$, and a Borel measurable function $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ such that
$$g(x^{1/2} \epsilon_t, x) \leq (a + \varphi(\epsilon_t)) x + c$$ 
for all $x \in \mathbb{R}_+$. Furthermore, $a + \varphi(0) < 1$ and $E[(a + \varphi(\epsilon_t))^r] < 1$ where the real number $r \geq 1$ is as in Assumption 1.

As mentioned above, the smoothness condition in Assumption 4(a) is needed to make use of the results for nonlinear state space models in Meyn and Tweedie (1993, Chapter 7). The same is true for Assumption 4(c) which is a high level assumption. Sufficient conditions for this assumption are discussed below. The latter condition in Assumption 4(a) implies that the conditional variance $h_t$ is bounded away from zero, a property shared by most GARCH models. Assumption (b) is technical and needed in the proofs. It is also satisfied by most commonly used first order GARCH models. Assumption 4(d) supplements Assumption 3 in that it is needed to prove the geometric ergodicity of the Markov chain $Z_t$. Assumptions closely related to Assumption 4(d) have also been used by Lanne and Saikkonen (2005) and Meitz and Saikkonen (2004).
In Assumption 4(c) the existence of a fixed point $h^*$ of the function $g(0, x)$ is assumed. A well-known sufficient condition which implies that a unique fixed point exists and can be found by the stated recursion is the Lipschitz condition

$$|g(0, x_1) - g(0, x_2)| \leq \kappa |x_1 - x_2| \quad \text{for some } 0 \leq \kappa < 1 \text{ and all } x_1, x_2 \in \mathbb{R}^+ \quad (7)$$

(this follows from the contraction map principle, see for example Simmons (1963, Appendix 1)). This condition applies to the standard (linear) GARCH(1,1) model and, more generally, when Assumption 4(d) holds with $g(0, x) = (a + \varphi(0)) x + c$. However, when the function $g(0, x)$ is nonlinear the Lipschitz condition (7) may not hold or it can be difficult to verify. Then the second condition of Assumption 4(c) may be useful. For instance, in some cases $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}^+$ and it suffices that the function $g(0, x)$ is nondecreasing for $x \geq h^*$.

The second condition of Assumption 4(c) combined with the other conditions of this assumption implies the convergence of the stated recursion. This can be seen as follows. Note first that from Assumption 4(d) it follows that $g(0, x) \leq (a + \varphi(0)) x + c \leq (a + \varphi(0) + \epsilon) x$ for all $x$ large enough and some $\epsilon > 0$ such that $a + \varphi(0) + \epsilon < 1$. From this and Assumption 4(a) it is straightforward to check that the function $g(0, x)$ has a maximal fixed point $h^*$ such that $g(0, h^*) = h^*$ and $g(0, x) < x$ for all $x > h^*$. This, together with the latter condition of Assumption 4(c), implies that for any initial value $h_0 > h^*$ the sequence $h_k, k \geq 0$, is nonincreasing and bounded from below by $h^*$. Therefore it converges to, say, $h_*$ ($\geq h^*$) and, because $g(0, h_k) = h_{k+1}$, we also have $g(0, h_k) \to h_*$. On the other hand, by the continuity of $g(0, \cdot)$, $g(0, h_k) \to g(0, h_*)$. Thus we must have $g(0, h_*) = h_*$ and, since $h^*$ is the maximal fixed point, $h_* = h^*$.

Our final assumption concerns the deterministic control model $z_k = F_k (z_0, e_1, \ldots, e_k)$ ($k = 1, 2, \ldots$) associated with the nonlinear state space model (6) and the concept of forward accessibility (for a definition, see Meyn and Tweedie (1993, p. 151)).

**Assumption 5** For each initial value $z_0 \in \mathbb{Z}$, there exists a control sequence $e_1^{(0)}, \ldots, e_{p+2}^{(0)}$ such that the $(p + 2) \times (p + 2)$ matrix

$$\nabla F_{p+2}^{(0)} = \left[ \frac{\partial}{\partial e_1} F_{p+2}^{(0)} (z_0, e_1^{(0)}, \ldots, e_{p+2}^{(0)}) \right] : \cdots : \left[ \frac{\partial}{\partial e_{p+2}^{(0)}} F_{p+2}^{(0)} (z_0, e_1^{(0)}, \ldots, e_{p+2}^{(0)}) \right]$$

is nonsingular.

By Proposition 7.1.4 of Meyn and Tweedie (1993), this assumption implies that the deterministic control model $z_k = F_k (z_0, e_1, \ldots, e_k)$ is forward accessible. This property is needed to apply the results obtained in Chapter 7 of Meyn and Tweedie (1993). Note that although Assumption 5 is sufficient for forward accessibility it is not necessary, as the aforementioned proposition of Meyn and Tweedie (1993) shows.

Although Assumption 5 may look difficult to verify in practice that is fortunately not the case for several commonly used models. To get an idea of the structure of the derivative matrix of $F_{p+2}^{(0)}$, denote the components of the vector $F_{p+2}^{(0)} (z_0, e_1, \ldots, e_{p+2})$
briefly by \( y_{p+2}, \ldots, y_2 \) and \( h_{p+2} \) (cf. equations (5) and the subsequent discussion). Then it is straightforward to check that

\[
\nabla F_{p+2} = \begin{bmatrix}
\frac{\partial y_{p+2}}{\partial e_1} & \frac{\partial y_{p+2}}{\partial e_2} & \frac{\partial y_{p+2}}{\partial e_3} & \cdots & \frac{\partial y_{p+2}}{\partial e_{p+1}} & h_{p+2}^{1/2} \\
\frac{\partial y_{p+1}}{\partial e_1} & \frac{\partial y_{p+1}}{\partial e_2} & \frac{\partial y_{p+1}}{\partial e_3} & \cdots & h_{p+1}^{1/2} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{\partial y_{2}}{\partial e_1} & h_{2}^{1/2} & 0 & \cdots & 0 & 0 \\
\frac{\partial h_{p+2}}{\partial e_1} & \frac{\partial h_{p+2}}{\partial e_2} & \frac{\partial h_{p+2}}{\partial e_3} & \cdots & \frac{\partial h_{p+2}}{\partial e_{p+1}} & 0 \\
\end{bmatrix},
\]

where \( h_{i}^{1/2} = h (z_{i-1})^{1/2} > 0 \) \( (i = 1, \ldots, p+1) \) and the superscript has been suppressed from \( \nabla F_{p+2}^{(0)} \) to indicate that the derivatives are evaluated at an arbitrary control sequence. Thus, for Assumption 5 to hold it suffices to find \( e_{1}^{(0)}, \ldots, e_{p+2}^{(0)} \) such that, for all initial values, \( \partial h_{p+2}/\partial e_1 \) is nonzero and \( \partial h_{p+2}/\partial e_2, \ldots, \partial h_{p+2}/\partial e_{p+1} \) are zero when evaluated at \( [e_{1}^{(0)}, \ldots, e_{p+2}^{(0)}] \). As will be seen in Section 4, this holds for the standard linear GARCH model and even for some nonlinear GARCH models without any further assumptions. However, for some models, including pure ARCH models, the situation turns out to be different.

We close this section by noting that a convenient feature of the assumptions imposed on the conditional mean and conditional variance is that, except for Assumption 5, they are separate. Specifically, as Lemma 1 shows, Assumptions 2 and 3 restrict only the conditional mean in (1) and this is done in the same way as in previous models without conditional heteroskedasticity. On the other hand, Assumption 4 only concerns the GARCH model specified for the error term in (1) and restricts it by conditions which are very similar to previous counterparts used in pure GARCH(1,1) models. As for Assumption 5, it concerns both the conditional mean and conditional variance but, as the examples of Section 4 show, this assumption is often easy to check by only considering the model specified for conditional heteroskedasticity.

### 3 Geometric Ergodicity

Under the assumptions stated in the previous section we are able to show that the Markov chain \( Z_t \) is geometrically ergodic. We use the \( Q \)-geometric ergodicity of a Markov chain introduced by Liebscher (2005). For convenience, we repeat the definition here in a slightly different, though equivalent, form. We use \( P^n(z, A) = \Pr(Z_n \in A \mid Z_0 = z), z \in \mathcal{Z}, A \in \mathcal{B}(\mathcal{Z}) \), to signify the \( n \)-step transition probability measure of the Markov chain \( Z_t \) defined on \( \mathcal{B}(\mathcal{Z}) \), the Borel sets of \( \mathcal{Z} \). (When \( n = 1 \) the notation \( P(z, A) \) will be used.)

**Definition 1** (Liebscher (2005)) The Markov chain \( Z_t \) on \( \mathcal{Z} \) is \( Q \)-geometrically ergodic if there exists a function \( Q : \mathcal{Z} \to [0, \infty] \), a probability measure \( \pi \) on \( \mathcal{B}(\mathcal{Z}) \), and
constants $a > 0$, $b > 0$, and $0 < \varrho < 1$ such that

$$\sup_{\|v\| \leq 1} \left| \int_{\mathcal{Z}} P^n(z, dw) v(w) - \int_{\mathcal{Z}} \pi(dw) v(w) \right| \leq (a + bQ(z)) \varrho^n \quad (8)$$

for all $z \in \mathcal{Z}$ and all $n \geq 1$.

Observing that the left hand side of (8) equals the total variation norm of the signed measure $P^n(z, \cdot) - \pi(\cdot)$ shows that our definition of $Q$–geometric ergodicity is equivalent to that in Liebscher (2005). Thus, geometric ergodicity entails that the $n$–step transition probability measure $P^n(z, \cdot)$ converges at a geometric rate to the probability measure $\pi(\cdot)$ with respect to the total variation norm for all $z \in \mathcal{Z}$. The probability measure $\pi$ is often referred to as the stationary probability measure of $Z_t$. The reason is that geometric ergodicity implies stationarity of $Z_t$ if the initial value $Z_0$ is distributed according to the probability measure $\pi$ (see Meyn and Tweedie (1993, p. 230–231)). Another useful consequence of $Q$–geometric ergodicity is that it implies that the Markov chain $Z_t$ is $\beta$–mixing for any initial value $Z_0$ with a distribution such that the expectation of $Q(Z_0)$ is finite (see Liebscher (2005)). Also, once $Q$–geometric ergodicity has been established the finiteness of the expectation $\int_{\mathcal{Z}} \pi(dw)Q(w)$ is automatically obtained. This fact can be used to show that the stationary distribution of $Z_t$ has finite moments of some order.

Note that one should be careful with the term $Q$–geometric ergodicity because, except for the prefix $Q$, another similar concept is in use. This concept is defined by assuming $Q \geq 1$ and replacing the inequality $|v| \leq 1$ and the bound $a + bQ(z)$ in (8) by $|v| \leq Q$ and $M_z$, respectively (see Meyn and Tweedie (1993, p. 356)). This clearly results in a stronger convergence than assumed in (8). This stronger convergence has also been established in various nonlinear autoregressive and GARCH models (see Meyn and Tweedie (1993), Saikkonen (2005), Lanne and Saikkonen (2005), and Meitz and Saikkonen (2004)). However, we have found it difficult to establish it in the present context. Therefore, we use the weaker $Q$–geometric ergodicity which, as discussed above, also provides us with useful results.

The standard method to establish $Q$–geometric ergodicity, as well as its aforementioned stronger counterpart, is based on the so called drift criterion (see Meyn and Tweedie (1993, Theorem 15.0.1) or Liebscher (2005)). Before the application of this criterion one needs to show that the considered Markov chain is irreducible and aperiodic. In many nonlinear autoregressions of the type (6) this can be done in a fairly straightforward way. That also applies to our model if the function $g$ in (2) is independent of $h_{t-1}$. Then the analysis can be reduced to that of the process $Y_t$ which is a Markov chain and one can employ the ideas in Cline and Pu (1998) and Lu (1998) to show irreducibility and aperiodicity. However, when the function $g$ also depends on $h_{t-1}$ we have to consider the larger Markov chain $Z_t$ in which the deterministic dependence of $h_t$ on past values of the process $y_t$ through the nonlinear function $f$ makes the analysis complicated and the approach described in Cline and Pu (1998) and Lu (1998) gets difficult. Similar difficulties occur when one tries to establish the $T$–continuity of $Z_t$ which, in conjunction with irreducibility and aperiodicity, implies
that compact subsets of $\mathcal{Z}$ are small, a fact also pertinent for the application of the drift criterion (see Theorems 6.2.5(ii) and 5.5.7 of Meyn and Tweedie (1993)).

Due to the aforementioned difficulties we establish the irreducibility, aperiodicity, and $T$–continuity of $Z_t$ by using the approach described in Chapter 7 of Meyn and Tweedie (1993). This approach is based on the deterministic control model associated with the nonlinear state space model (6) and, as already discussed, its application assumes the smoothness conditions imposed in Assumptions 2 and 4(a). We have the following lemma.

**Lemma 2** If Assumptions 1–4 hold then the Markov chain $Z_t$ on $\mathcal{Z}$ is an irreducible and aperiodic $T$–chain and, hence, all compact subsets of $\mathcal{Z}$ are small. Moreover, the set $A_N = \{ z \in \mathcal{Z} : \|y\|^{2r} \leq N, \|h^r(z)\| \leq N \}$ is small for any vector norm and for all positive $r$ and $N$ such that $g^r < N$.

Thus, Lemma 1 provides the necessary prerequisites for the application of the drift criterion. Note that Lemma 1 also shows that certain noncompact subsets of $\mathcal{Z}$ are small. Unlike in many previous cases this result greatly facilitates the application of the drift criterion. This part of the lemma is based on ideas used by Cline and Pu (1998, Theorem 2.5) who also discuss its usefulness.

The following theorem presents the main result of the paper. In the proof of this theorem we apply an $m$–step ahead drift criterion for a sufficiently large value of $m$ (cf. Theorem 19.1.3 of Meyn and Tweedie (1993)). In most previous cases 1–step ahead versions of this criterion have sufficed, but in the present model the combination of the assumed nonlinear autoregressive structure both in the conditional mean and conditional variance seems to make the application of the this more conventional approach difficult. Although the possibility to make use of the $m$–step ahead drift criterion in nonlinear autoregressions was already pointed out by Tjøstheim (1990) it seems that its previous applications have been rather rare and confined to cases where a 1–step ahead drift criterion would have worked without any difficulty. A new $m$–step ahead drift criterion for $\mathcal{Q}$–geometric ergodicity (Lemma 6), potentially of independent interest, is proven in the Appendix.

**Theorem 1** Suppose that Assumptions 1–4 hold, and let $\|\cdot\|$ be any vector norm. Then the Markov chain $Z_t$ on $\mathcal{Z}$ is $\mathcal{Q}^*$–geometrically ergodic in the sense of Liebscher (2005) with a function $\mathcal{Q}^*(z) \geq 1 + \|y\|^{2r} + h^r(z)$.

As discussed after Definition 1, Theorem 1 implies that, with appropriate initial distributions, the process $(y_t, h_t)$ is $\beta$–mixing and that there exists a stationary initial distribution such that $y_t$ and $h_t$ have moments of orders $2r \beta$ and $r$, respectively (the latter moment result follows because $h_t = h(Z_t-1)$). An important consequence of Theorem 1 is that usual limit theorems apply. As far as we know, there is no equivalent to this result in the previous literature on nonlinear autoregressions with GARCH errors.
4 Examples

We shall now consider concrete examples to which Theorem 1 applies. According to what was said after Assumption 1, it suffices to discuss Assumptions 2–5 of which Assumptions 2 and 3 concern the conditional mean of the model, that is, the function $f$, whereas Assumption 4 restricts the form of permitted conditional heteroskedasticity. As already indicated, Assumption 5 can often be checked without paying attention to the conditional mean. In such cases it is only necessary to check conditions imposed on the conditional mean and conditional variance separately.

First consider the conditional mean. A very general specification only assumes that the function $f$ has the general structure imposed in Assumption 2. In this case, general sufficient conditions for Assumption 3 are obtained from Lemma 1. This approach is relevant for the general functional-coefficient autoregressive model of Chen and Tsay (1993). For more concrete examples, we have to be more specific about the function $a$ in Assumption 2. For instance, suppose that $f(y_{t-1}, \ldots, y_{t-p}) = \phi_0 + \psi_0 G(y_{t-1}, \ldots, y_{t-p}) + \sum_{j=1}^{p} (\phi_j + \psi_j G(y_{t-1}, \ldots, y_{t-p})) y_{t-j}$, (9)

where $\phi_j, \psi_j \in \mathbb{R}$, $j = 0, \ldots, p$, and $G$ is a smooth function with range $[0, 1]$. In this case Lemma 1(ii) shows that a sufficient condition for Assumption 3 is

$$\sum_{j=1}^{p} \max \{|\phi_j|, |\phi_j + \psi_j|\} < 1,$$

a condition previously obtained by Chen and Tsay (1993, Example 2) for a special choice of the function $G$.

To apply Lemma 1(i) to the specification (9), define $A_1 = \bar{A}_p((\phi_1, \ldots, \phi_p)'$ and $A_2 = \bar{A}_p((\phi_1 + \psi_1, \ldots, \phi_p + \psi_p)')$. From Theorem 1 and Proposition 5 of Liebscher (2005) we can then conclude that a sufficient condition for Lemma 1(i), and hence, Assumption 3 is that the joint spectral radius of the set of two matrices $\{A_1, A_2\}$ is smaller than one, or that

$$\rho(\{A_1, A_2\}) < 1.$$ (11)

Liebscher (2005, Section 7) provides a numerical example of this condition with $p = 2$ and $G(y_{t-1}, \ldots, y_{t-p}) = \exp(-\gamma y_{t-1}^2)$ ($\gamma > 0$). This choice of the function $G$ corresponds to the exponential autoregressive (EXPAR) model introduced by Haggan and Ozaki (1981) and also studied by Tong (1990, p. 108). In this numerical example condition (11) holds but the simpler condition (10) is violated. Clearly, the same conclusion is obtained even if the general function $G$ is assumed.

The EXPAR model discussed above is closely related to the exponential smooth transition autoregressive (ESTAR) model which, along with other smooth transition autoregressive models, have been considered by Teräsvirta (1994) and van Dijk, Teräsvirta, and Franses (2002) amongst others. In the ESTAR case, (9) applies
with \( G(y_{t-1}, \ldots, y_{t-p}) = 1 - \exp\left(-\gamma(y_{t-d} - c)^2\right) \) whereas the logistic smooth transition autoregressive (LSTAR) specification is given by \( G(y_{t-1}, \ldots, y_{t-p}) = (1 + \exp\left(-\gamma(y_{t-d} - c)^2\right))^{-1} \) \((\gamma > 0, c \in \mathbb{R}, 1 \leq d \leq p)\). A generalization of the latter is obtained by

\[
G(y_{t-1}, \ldots, y_{t-p}) = \left(1 + \exp\left(-\gamma \prod_{j=1}^{k} (y_{t-d} - c_j)\right)\right)^{-1},
\]

where \(\gamma\) and \(d\) are as above and \(c_1 < \cdots < c_k\). Of course, conditions (10) and (11) which apply to the general specification (9) also apply to all these special cases. It may also be noted that, although our smoothness assumption rules out the possibility that \(G\) is an indicator function, an approximating smooth counterpart such as a logistic function is allowed.

Now consider the conditional variance. Although the conditions imposed on the function \(g\) in Assumption 4 rule out threshold GARCH models they apply to smooth transition GARCH models introduced in Hagerud (1996) and González-Rivera (1998), and further discussed in Lundbergh and Teräsvirta (2002), Lanne and Saikkonen (2005), and Meitz and Saikkonen (2004). In one variant of this model the dynamics of the conditional variance process are governed by

\[
h_t = g(u_{t-1}, h_{t-1}) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + \alpha^* G(u_{t-1}) u_{t-1}^2,
\]

where \(G\) is a smooth function with range \([0, 1]\) and the parameters satisfy \(\omega > 0, \alpha > 0, \beta > 0, \) and \(\alpha + \alpha^* > 0\). This model reduces to the linear GARCH model of Bollerslev (1986) when \(\alpha^* = 0\). Again, the possibility that \(G\) is an indicator function is ruled out but an approximating smooth counterpart is allowed giving a smooth version of the GJR specification of Glosten, Jaganathan, and Runkle (1993). Checking the validity of Assumption 4 for this model is straightforward. Assumptions 4(a) and 4(b) clearly hold with the lower bound in the former given by \(g(0)\) when \(\alpha > 0\) and \(\alpha^* = 0\), and since the moment condition \(E[\beta + \max\{\alpha, \alpha + \alpha^*\}] < 1\) when \(r = 1\) we calculate that \(g(0, x) = \omega + \beta x\) so that, because \(\beta < 1\), the Lipschitz condition (7), and hence Assumption 4(c), is satisfied.

We shall now demonstrate that Assumption 5 holds for the GARCH model (12). Following the discussion in Section 2 we consider the last row of the derivative matrix \(\nabla F_{p+2}^{(0)}\). The needed derivatives can be straightforwardly obtained from equation (12) and, unless otherwise stated, all derivatives below are evaluated at \(e_2^{(0)} = \cdots = e_{p+2}^{(0)} = 0\). First note that \(\partial h_i/\partial e_j = \beta \partial h_{i-1}/\partial e_j, i = 3, \ldots, p+2, j = 1, \ldots, i-2\), and \(\partial h_i/\partial e_{i-1} = 0, i = 3, \ldots, p+2\), and thus the last row of the matrix \(\nabla F_{p+2}^{(0)}\) becomes 

\[
\begin{bmatrix}
\partial h_{p+2}/\partial e_1 & 0 & \cdots & 0 \\
\end{bmatrix}.
\]

To obtain \(\partial h_{p+2}/\partial e_1\) we calculate \(\partial h_{2}/\partial e_1\) (evaluated at an arbitrary \(e_1\)) and find that

\[
\partial h_{p+2}/\partial e_1 = 2\beta^p h(z_0) e_1 \left(\alpha + \alpha^* \left(h(z_0)^{1/2} e_1\right)\right) + \beta^p \alpha^* h(z_0)^{3/2} G' \left(h(z_0)^{1/2} e_1\right) e_1^2.
\]
For the standard linear GARCH model $\alpha^* = 0$ so that, since $\beta > 0$ is assumed, 
$\partial h_{p+2}/\partial e_1 = 2B \alpha h(z_0) e_1$ is nonzero for any $e_1 \neq 0$ whereas 
$\partial h_{p+2}/\partial e_2, \ldots,$ 
$\partial h_{p+2}/\partial e_{p+1}$ are zero. Thus, as discussed in Section 2, Assumption 5 holds. The 
same conclusion is obtained even if $\alpha^* \neq 0$. In this case $\partial h_{p+2}/\partial e_1 \neq 0$ may not hold 
for all $e_1 \neq 0$ without further assumptions on the derivative $G'$ but it clearly holds 
for some $e_1 \neq 0$, which suffices for Assumption 5.

In an alternative smooth transition GARCH model, suggested by Lanne and 
Saikkonen (2005), the conditional heteroskedasticity is specified as 
$h_t = g(u_{t-1}, h_{t-1}) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + (\omega^* + \beta^* h_{t-1})G(h_{t-1}), \quad (13)$

where $G$ is again a smooth function with range $[0, 1]$ and the parameters satisfy $\omega > 0,$ 
$\alpha > 0,$ $\beta > 0,$ $\omega^* \geq 0,$ and $\beta^* \geq 0$. Again, the validity of Assumptions 4(a) and (b) 
is clear with $g = \omega$ in the former. For Assumption 4(d), we may choose $a = \beta + \beta^*$, 
$\varphi(\epsilon_i) = \alpha \epsilon_i^2,$ and $c = \omega + \omega^*$. This gives the condition $E[(\beta + \beta^* + \alpha \epsilon_i^2)] < 1$ or 
$\beta + \beta^* + \alpha < 1$ when $r = 1$. To verify Assumption 4(c), the Lipschitz condition 
(7) appears inconvenient, and the latter part of this assumption becomes useful. For 
this model, $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$. Thus, if we assume that 
the function $G(x)$ is nondecreasing, the same is true for $g(0, x)$ $(x > 0)$ and it follows that 
Assumption 4(c) holds (see the discussion after Assumption 4). That Assumption 5 
holds will be discussed below in the context of a related model.

In the preceding model one can also consider the case $\omega > 0,$ $\omega^* > 0,$ $\alpha > 0,$ and 
$\beta = \beta^* = 0$. This special case was applied by Lanne and Saikkonen (2005) with the 
function $G$ (strictly) increasing. As above, one can check that Assumption 4 holds 
for this specification with the moment condition in part (d) given by $E[\alpha \epsilon_i^2] < 1$ or 
$\alpha < 1$ when $r = 1$. More generally, the assumption that the (increasing) function $G$ 
is bounded can be relaxed by requiring that $G(x) = o(x)$ as $x \to \infty$. It suffices to 
discuss Assumptions 4(c) and (d). Regarding the latter, one can write 
$g(x^{1/2} \epsilon_t, x) = \omega + \omega^* G(x) + \alpha \epsilon_t^2$

where $1(\cdot)$ signifies the indicator function. Choosing $M$ large enough the last expression 
can be bounded from above by $(\epsilon + \alpha \epsilon_t^2) x + c$ where $0 < c < 1$ is so small 
that $E[\epsilon + \alpha \epsilon_t^2] < 1$ holds whenever $E[\alpha \epsilon_i^2] < 1$. Thus, Assumption 4(d) holds. 
Since $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$ also Assumption 4(c) holds.

Now consider Assumption 5 in the preceding model. In the same way as in the 
standard GARCH model discussed in the context of model (12), it is straightforward 
to check that when evaluated at $\epsilon_{t-1} = \cdots = \epsilon_{p+2} = 0$, $\partial h_i/\partial e_j = \omega^* G(h_{t-1}) \partial h_{i-1}/\partial e_j,$ 
i = 3, \ldots, p + 2, j = 1, \ldots, i - 2, and $\partial h_i/\partial e_{i-1} = 2a h_{i-1} e_{i-1}, i = 3, \ldots, p + 2$. Thus, 
since the function $G$ is increasing we can choose $e_1 = e^{(0)}_i \neq 0$ such that 
$\partial h_{p+2}/\partial e_1$ 
evaluated at $\epsilon^{(0)}_i$ and $\epsilon^{(0)}_2 = \cdots = \epsilon^{(0)}_{p+2} = 0$ becomes nonzero while 
$\partial h_{p+2}/\partial e_2 = \cdots = \partial h_{p+2}/\partial e_{p+1} = 0$. Hence Assumption 5 holds. The same reasoning applies to model 
(13), for it suffices to consider the (increasing) function $\beta x + (\omega^* + \beta^* x)G(x)$ in place 
of $\omega^* G(x)$.
In the preceding cases the verification of Assumption 5 required that the considered GARCH models do not reduce to pure ARCH models. We shall now demonstrate that Assumption 5 can also be verified when the conditional heteroskedasticity is modeled by a pure ARCH model. Because in many fields of application pure ARCH models are seldom adequate and, as discussed in the introduction, stability results for them are already available we shall only consider the standard ARCH model

\[ h_t = \omega + \alpha u_{t-1}^2 \quad (\alpha > 0) \]

and, for simplicity, assume that \( p = 1 \). It suffices to show that the matrix

\[
D_1 = \begin{bmatrix}
\frac{\partial y_2}{\partial e_1} & \frac{\partial y_2}{\partial e_2} \\
\frac{\partial h_3}{\partial e_1} & \frac{\partial h_3}{\partial e_2}
\end{bmatrix}
\]

is nonsingular when the partial derivatives are evaluated at suitable values of \( e_1 \) and \( e_2 \). By straightforward derivation,

\[
\frac{\partial y_2}{\partial e_1} = h(z_0)^{1/2} f' (y_0) + h(z_0)^{1/2} e_1 + h(z_0) h_2^{-1/2} \alpha e_1 e_2
\]

\[
\frac{\partial y_2}{\partial e_2} = h_2^{1/2}
\]

\[
\frac{\partial h_3}{\partial e_1} = 2\alpha^2 h(z_0)e_1 e_2
\]

\[
\frac{\partial h_3}{\partial e_2} = 2\alpha h_2 e_2.
\]

Assume that \( e_2 \) is nonzero. Then, multiplying the first row of \( D_1 \) by \( h_2^{1/2} \) and dividing the second row and second column by \( 2\alpha e_2 \) and \( h_2 \), respectively, we can transform \( D_1 \) to

\[
D_2 = \begin{bmatrix}
\frac{h_2^{1/2}}{\alpha h(z_0)} e_1 e_2 & 1 \\
1 & 1
\end{bmatrix}.
\]

This matrix can further be transformed to

\[
D_3 = \begin{bmatrix}
\frac{h_2^{1/2}}{\alpha h(z_0)} e_1 e_2 & 0 \\
\frac{h(z_0)^{1/2} f' (y_0) + h(z_0)^{1/2} e_1}{\alpha h(z_0)} & 1
\end{bmatrix}
\]

by subtracting the second row of \( D_2 \) from its first row. The rank of \( D_3 \) clearly equals the the rank of \( D_1 \). Thus, assuming that the function \( f \) is not constant we can find an \( e_1 \) such that the element of \( D_3 \) in the upper left hand corner is nonzero. Then \( D_3 \), or equivalently, \( D_1 \) is nonsingular, and because this can be done for each initial value \( z_0 \), Assumption 5 holds.

5 Conclusion

In this paper we have studied a nonlinear autoregressive model of order \( p \) with conditionally heteroskedastic errors specified as a nonlinear GARCH(1,1) model. We gave conditions under which the Markov chain representation of the model is \( Q \)-geometrically ergodic in the sense of Liebscher (2005) and, hence, \( \beta \)-mixing. Conditions for existence of moments of the stationary distribution were also obtained. The assumptions needed to obtain these results are convenient because in most cases they restrict the conditional mean and conditional variance separately. To the best of our knowledge, these are the first practically applicable stability results for nonlinear
autoregressions with GARCH errors. They are of importance as they open up the way for the development of rigorous asymptotic estimation theory for these models. Due to the approach taken to obtain the results of the paper, rather stringent smoothness assumptions on the permitted nonlinearity were needed, and hence threshold type nonlinear models could not be covered. It would be of interest to consider alternative approaches in which such smoothness assumptions would not be required. For simplicity, we also focused on the leading case of GARCH(1,1) errors and left the extension to the general GARCH(r,s) case for future work.
Appendix: Proofs

Proof of Lemma 1. First note that $A_s$ and $A_1$ are both bounded sets of matrices. That $\rho(A_s) < 1$ implies Assumption 3 follows from Theorem 1 of Liebscher (2005).

To see that $\rho(A_1) < 1$ is equivalent to this condition, notice that

$$A(x_1)A(x_2)\cdots A(x_k) = \begin{bmatrix} A_1(x_1)A_1(x_2)\cdots A_1(x_k) & 0 \\ A_{\nu}(x_1)A_1(x_2)\cdots A_1(x_k) & 0 \end{bmatrix},$$

where $\nu = [0\cdots 0 1]^\top (p \times 1)$. Thus, the stated equivalence can be established by choosing the norm in the definition of the joint spectral radius as the maximum of absolute row sums (the matrix norm induced by the $l_\infty$-norm).

To justify the second part of the lemma, use the notation introduced in Section 2 and denote $A = A_{p+1}((\alpha_1,\ldots,\alpha_p,0)^\top)$. By direct calculation, the characteristic polynomial of $A$ is (up to a factor $\pm 1$) $\lambda^p - \alpha_1 \lambda^{p-1} - \cdots - \alpha_p$. To see the equivalence of the two conditions, denote $f_1(\lambda) = \lambda^p - \alpha_1 \lambda^{p-1} - \cdots - \alpha_p$, $f_2(\lambda) = \lambda^p - f_1(\lambda)$, and first suppose that $\sum_{j=1}^{p} \alpha_j < 1$. Now, for $|\lambda| \geq 1$, $|f_1(\lambda)| \geq |\lambda|^p - |f_2(\lambda)| \geq |\lambda|^p (1 - \sum_{j=1}^{p} \alpha_j) > 0$, and hence the roots are inside the unit circle. On the other hand, if $\sum_{j=1}^{p} \alpha_j \geq 1$, then $f_1(1) \leq 0$ while $f_1(\lambda) \to +\infty$ as $\lambda \to +\infty$ and hence there is a root on or outside the unit circle. Thus under either condition $\rho(A) < 1$ (cf. Chen and Tsay (1993, Proof of Theorem 1.1)).

Now, using the same argument as in Ling and McAleer (2003, Proof of Lemma A.2) we can find a $(p+1) \times 1$ vector $\kappa$ with positive components such that the components of the row vector $\nu' = \kappa' (I_{p+1} - A)$ are positive and, furthermore, $0 < \nu' / \kappa < 1$ where $\nu$ and $\kappa$ are the smallest and largest components of $\nu$ and $\kappa$, respectively. Next define the vector norm $\|\cdot\|^*$ in $\mathbb{R}^{p+1}$ by $\|y\|^* = \sum_{j=1}^{p+1} \kappa_j |y_j| = \kappa' |y|$ where $|y| = [|y_1| \cdots |y_{p+1}|]^\top$. For arbitrary $A = A(x) \in A_s$ and $y \in \mathbb{R}^{p+1}$, $y \neq 0$, we have

$$\|A(x)y\|^* = \kappa' |A(x)y|$$

$$\leq \kappa_1 \sum_{j=1}^{p} \alpha_j |y_j| + \sum_{j=1}^{p} \kappa_{j+1} |y_j|$$

$$= \kappa' A |y|$$

$$= \kappa' |y| - \kappa' (I_{p+1} - A) |y|$$

$$= \kappa' |y| - \nu' |y|$$

$$= \kappa' |y| (1 - \nu' / \kappa)$$

$$\leq \|y\|^* (1 - \frac{\nu'}{\kappa}),$$

where $0 < 1 - \nu' / \kappa < 1$. This shows that the matrix norm induced by $\|\cdot\|^*$ satisfies Assumption 3.

Proof of Lemma 2. We consider $Z_t$ as a nonlinear state space model and use the results in Chapter 7 of Meyn and Tweedie (1993) (note that under our assumptions,
the conditions (NSS1)–(NSS3) in Meyn and Tweedie (1993, pp. 32 and 156) are satisfied. For this we need to show that the deterministic control model associated with $Z_t$ is forward accessible and attains a globally attracting state (for definitions of these concepts, see pp. 155 and 160 of Meyn and Tweedie (1993), respectively). As discussed in Section 2, forward accessibility follows from Assumption 5 and Proposition 7.1.4 of (Meyn and Tweedie (1993)). The existence of a globally attracting state is shown below in Lemma 3. Thus, from Propositions 7.1.5 and 7.2.5(i), and Theorem 7.2.6 of Meyn and Tweedie (1993) we can conclude that the Markov chain shown below in Lemma 3. Thus, from Propositions 7.1.5 and 7.2.5(i), and Theorem 7.2.6 of Meyn and Tweedie (1993) we can conclude that the Markov chain $Z_t$ is an irreducible $T$-chain. Aperiodicity is obtained from Theorems 7.3.3 and 7.3.5(ii) of the same reference (see also the proof of Proposition 7.4.1) because any cycle of the associated control model must contain the globally attracting state (in Lemma 3 we also show that there exists a control sequence such that the deterministic control model converges to the globally attracting state, and thus the period in Theorem 7.3.3 of ibid. necessarily equals one). That every compact set is small now follows from Theorems 6.2.5(ii) and 5.5.7 of Meyn and Tweedie (1993)). Finally, in Lemma 4 below it is shown that the set $A_N$ is also small.

Thus, the proof of Lemma 2 is completed by the following two lemmas.

**Lemma 3** Under Assumptions 1–4 the deterministic control model associated with the Markov chain $Z_t$ attains a globally attracting state.

**Proof.** For a $z^* \in Z$ to be a globally attracting state for the associated deterministic control model it suffices to establish that, for any initial value $z_0 \in Z$, there exists a control sequence $e_t$ such that $z_t$ converges to $z^*$ as $t \to \infty$ (see Meyn and Tweedie (1993, p. 160)). First suppose that the convergence in Assumption 4(c) holds for all $h_0 \in \mathbb{R}_+$ so that for every $z_0 \in Z$, $h_t \to h^*$ as $t \to \infty$.

By Assumption 3 there exist an induced matrix norm $\|\cdot\|^*$ and a real number $\rho \in (0, 1)$ such that $\|A\|^* \leq \rho$ for all $A \in \mathbb{A}$. As in Assumption 3 we also use $\|\cdot\|^*$ for the vector norm corresponding to the matrix norm $\|\cdot\|^*$. Because the function $b$ is bounded by assumption we can find a positive real number $c$ such that $\|b(x)\|^* \leq c/2$ for all $x \in \mathbb{R}^p$. Define the compact set $K = \{y \in \mathbb{R}^{p+1}: \|y\|^* \leq c/(1 - \rho)\}$ and note that the mapping $y \mapsto A(S'y) y + b(S'y)$ ($y \in \mathbb{R}^{p+1}$) is continuous. Furthermore, when $y \in K$, the range of this mapping is contained in $K$ because, for $y \in K$,

$$
\|A(S'y) y + b(S'y)\|^* \leq \|A(S'y)\|^* \|y\|^* + \|b(S'y)\|^* \\
\leq \rho \|y\|^* + c/2 \\
\leq \rho c/(1 - \rho) + c/2 \\
= c(1 + \rho)/2(1 - \rho) \\
\leq c/(1 - \rho).
$$

Thus, it follows from Schauder’s fixed point theorem (see e.g. Simmons (1963, Appendix 1)) that there exists a state $y^* \in K$ such that $y^* = A(S'y^*) y^* + b(S'y^*)$.

We shall now demonstrate that, from any $z_0 \in Z$, it is possible for the associated control model to reach a state $z^*$ whose first $p + 1$ components are $y_0, \ldots, y_{p+1}$, the components of the vector $y^*$, and the last component is $h^*$. Let $z_0 = [y_0, h_0]^T \in Z$.
where \( y_0 = [y_{0,1} \cdots y_{0,p+1}]' \). From the first step of the associated control model one then obtains

\[
\begin{align*}
y_1 &= \left[ a(S'y_0)' : 0 \right] y_0 + b(S'y_0) + h(z_0)^{1/2} e_1 \\
h_1 &= h(z_0)
\end{align*}
\]

and with \( e_1 = h(z_0)^{-1/2} (y_{p+1}' - [a(S'y_0)' : 0] y_0 - b(S'y_0)) \) we get \( y_1 = y_{p+1}' \). Next, setting \( \bar{y}_1 = [y_{p+1}' y_{0,1} \cdots y_{0,p}]' \) and \( z_1^* = [\bar{y}_1' \ h_1]' \) the second step of the associated control model gives

\[
\begin{align*}
y_2 &= \left[ a(S'ar{y}_1')' : 0 \right] \bar{y}_1 + b(S'ar{y}_1') + h(z_1^*)^{1/2} e_2 \\
h_2 &= h(z_1^*)
\end{align*}
\]

which with \( e_2 = h(z_1^*)^{-1/2} (\bar{y}_1' - [a(S'ar{y}_1')' : 0] \bar{y}_1 - b(S'ar{y}_1')) \) yields \( y_2 = y_{p+1}' \). The next step is to set \( \bar{y}_2 = [y_{p+1}' y_{p+1}' y_{0,1} \cdots y_{0,p-1}]' \) and \( z_2^* = [\bar{y}_2' \ h_2]' \) and choose \( e_3 = h(z_2^*)^{-1/2} \times (y_{p-1}' - [a(S'ar{y}_2')' : 0] \bar{y}_2 - b(S'ar{y}_2')) \). This gives \( y_3 = y_{p+1}' \) and \( z_3^* = [\bar{y}_3' \ h_3]' \) defined in an obvious way. Continuing in this way we reach the state \( z_{p+1}^* = [y_1' \cdots y_{p+1}' h_{p+1}'] \) in \( p + 1 \) steps.

Next form \( z_1^* \) with \( e_t = 0, t = p+2, p+3, \ldots \). Because \( y^* = A(S'y^*) + b(S'y^*) \) the first \( p + 1 \) components of \( z_1^* \) will be the components of \( y^* \) for all \( t \geq p + 2 \). Thus, \( z_1^* = [y^* \ h_1']' \) \( (t \geq p + 2) \) where the last component satisfies \( h_1^* = g(0, h_{p+1}^-) \) for \( t \geq p + 3 \). Because Assumption 4(c) implies that \( h_1^* \rightarrow h^* \) as \( t \rightarrow \infty \) we can conclude that \( z^* \) is a globally attracting state for the associated control model.

Now suppose that the convergence in Assumption 4(c) holds for all \( h_0 \geq h^* \). By Assumption 4(b) we can first choose an \( e_1 \) such that \( h_2 = g(h_1^{1/2} e_1, h_1) > h^* \). As seen above, we can next choose \( e_2, \ldots, e_{p+2} \) to reach a state whose first \( p + 1 \) components are \( y_1', \ldots, y_{p+1}' \), the components of the vector \( y^* \). This can be done regardless of the initial value \( z_0 \). Because \( h_2 > h^* \), the relevant part of Assumption 4(c) implies \( h_3 = g(h_2^{1/2} e_2, h_2) > h^* \) and similarly \( h_k > h^* \) for \( k = 4, \ldots, p + 2 \). Thus, after \( p + 2 \) steps we are in a state \( z_{p+2}^* = [y^* \ h_{p+2}]' \) and we continue by forming \( z_{t+1}^* \) with \( e_t = 0, t = p+3, p+4, \ldots \). Then the first \( p + 1 \) components of \( z_t^* \) will not change and, because \( h_{t+2} \geq h^* \), the last one tends to \( h^* \) as \( t \rightarrow \infty \). Thus we have again shown that a globally attracting state exists for the associated control model. \( \blacksquare \)

**Lemma 4** Under Assumptions 1–4 the set \( A_N = \{ z \in Z : \|y\|^{2r} \leq N, h^*(z) \leq N \} \) is small for any vector norm and for all positive \( r \) and \( N \) such that \( g^* < N \).

**Proof.** Writing equation (5) as \( Z_t = F_0 (Z_{t-1}) + t h (Z_{t-1})^{1/2} \varepsilon_t \) we have

\[
E \|Z_t\| | Z_{t-1} = z = E \|F_0 (z) + t h (z)^{1/2} \varepsilon_t\| \\
\leq \|F_0 (z)\| + t h (z)^{1/2} E \|\varepsilon_t\|
\]
and, since the functions $F_0$ and $h$ are bounded on the set $A_N$, we can find an $M_N < \infty$ such that

$$\sup_{z \in A_N} E[\|Z_t\| \mid Z_{t-1} = z] < M_N. \quad (14)$$

Now define the set $B_N = \{ z \in \mathbb{Z} : \|z\| \leq M_N, h \geq g \}$ (where $h$ is the last component of $z$). This set is small because it is compact, as noted above. We have

$$\inf_{z \in A_N} \Pr(Z_t \in B_N \mid Z_{t-1} = z) = 1 - \sup_{z \in A_N} \Pr(Z_t \notin B_N \mid Z_{t-1} = z) \geq 1 - \sup_{z \in A_N} \Pr(\|Z_t\| \geq M_N \mid Z_{t-1} = z) \geq 1 - \sup_{z \in A_N} E[\|Z_t\| \mid Z_{t-1} = z]/M_N > 0.$$ 

Here the first inequality is justified by the fact that, for all $z$, $\Pr(Z_t \notin B_N \mid Z_{t-1} = z) = \Pr(\|Z_t\| > M_N$ or $h(z) < g \mid Z_{t-1} = z)$ but $h(z) < g$ is impossible by Assumption 4(a). The second inequality is Markov’s and the third one is due to (14). That the set $A_N$ is small can now be concluded from Proposition 5.2.4 of Meyn and Tweedie (1993).

Proof of Theorem 1. First note that, by Lemma 2, $Z_t$ is irreducible and aperiodic and the set $A_N$ is small. Let $\|\cdot\|$ be any vector norm, and let $\|\cdot\|^*$ be an induced matrix norm with properties described in Assumption 3, i.e., a norm that satisfies $\|A\|^* \leq \rho$ for all $A \in \mathcal{A}$ and with $\rho \in (0, 1)$. Since all vector norms are equivalent in finite-dimensional real (or complex) vector spaces, there exists a finite positive constant $C$ such that $\|y\| \leq C^{1/2r}\|y\|^*$ for all $y \in \mathbb{R}^{p+1}$ (see e.g. Horn and Johnson (1985, Section 5.4)). Denote $V_*(z) = 1 + C\|y\|^{2r} + h^r(z)$. In Lemma 5 the conditional expectation $E[V_*(Z_t) \mid Z_{t-m} = z]$ is examined and it is demonstrated that it satisfies an $m$–step ahead drift criterion (for a large $m$ chosen in the proof of the lemma). More precisely, in this lemma it is shown that condition (19.15) of Meyn and Tweedie (1993) holds for the function $V_*(z)$ (with the choice $n(z) \equiv m$). Finally an application of our Lemma 6 below establishes that $Z_t$ is $V_*$–geometrically ergodic in the sense of Liebscher (2005).

Thus, the following two lemmas complete the proof of Theorem 1. □

Lemma 5 Suppose the assumptions of Theorem 1 are satisfied and define the function $V_*(z) = 1 + C\|y\|^{2r} + h^r(z)$. Then, there exist a small set $K$, a positive integer $m$, and positive real numbers $\lambda < 1$ and $b$ such that

$$E[V_*(Z_t) \mid Z_{t-m} = z] \leq \lambda^{1/2} \left( 1 + C\|y\|^{2r} + h^r(z) + 1_K(z) \right). \quad (15)$$

In other words, the drift condition (19.15) of Theorem 19.1.3 of Meyn and Tweedie (1993) holds (with the choice $n(z) \equiv m$).

Proof. First note that, by Hölder’s inequality,

$$\left( \sum_{i=1}^{n} x_i \right)^r \leq \sum_{i=1}^{n} x_i^r \cdot n^{r-1} \quad (16)$$
for any positive $x_i$, $1 \leq i \leq n$, $n \in \mathbb{Z}_+$, and $r > 1$ (and this trivially holds also for $r = 1$).

To analyze the conditional expectation in the lemma we first consider the quantity $h(Z_{t-1})$. From equations (1), (4), and (5) we obtain $h(Z_{t-m+1}) = g(h^{1/2}(Z_{t-m})\xi_{t-m+1}, h(Z_{t-m}))$, and, by using Assumption 4(d) with the notation $c_{t-1} = a + \phi(\xi_{t-1})$, this quantity can be bounded from above with $h(Z_{t-m})\xi_{t-m+1} + c$. Therefore we have, for $k \geq 1$, and interpreting that an empty summation equals zero,

$$h(Z_{t-m+k}) \leq \prod_{j=1}^{k} c_{t-m+j} \cdot h(Z_{t-m}) + c \left(1 + \sum_{j=0}^{k-2} \prod_{i=0}^{j} c_{t-m+k-i}\right).$$

Using (16) we obtain

$$(k + 1)^{1-r} h^r(Z_{t-m+k}) \leq \prod_{j=1}^{k} c_{t-m+j} \cdot h^r(Z_{t-m}) + c^r \left(1 + \sum_{j=0}^{k-2} \prod_{i=0}^{j} c_{t-m+k-i}\right).$$

By Assumption 4(d), $E[c_i^r] < 1$ and we denote this expectation by $\delta$.

Next note that (trivially) $E[h^r(Z_{t-m}) \mid Z_{t-m} = z] = h^r(z)$. Furthermore, using the notation $d = c^r/(1 - \delta)$ and the independence of the $c_i$’s,

$$(k + 1)^{1-r} E[h^r(Z_{t-m+k}) \mid Z_{t-m} = z] \leq h^r(z)\delta^k + c^r \left(1 + \sum_{j=0}^{k-2} \delta^{j+1}\right) \leq h^r(z)\delta^k + d.$$

In particular, for $k = 1, \ldots, m - 1$,

$$E[h^r(Z_{t-m+k}) \mid Z_{t-m} = z] \leq (k + 1)^{r-1} (h^r(z)\delta^k + d) \leq m^{r-1} h^r(z)\delta^k + d'.$$

where $d' = m^{r-1}d$.

Now consider $Y_t$ which we wish to express in terms of past values of the process $Z_t$ until $t - m$. Recall that $\|\cdot\|^*$ and $\rho$ are as in Assumption 3. Repeated substitution and usual properties of vector and matrix norms yield

$$\|Y_t\|^* \leq \prod_{j=0}^{m-1} \|A(S'Y_{t-1-j})\|^* \|Y_{t-m}\|^* + \|ib(S'Y_{t-1})\|^* + \sum_{j=0}^{m-2} \prod_{i=0}^{j} \|A(S'Y_{t-1-i})\|^* \|ib(S'Y_{t-2-j})\|^* + \|ih(Z_{t-1})^{1/2} \xi_t\|^*$$

$$+ \sum_{j=0}^{m-2} \prod_{i=0}^{j} \|A(S'Y_{t-1-i})\|^* \|ih(Z_{t-2-j})^{1/2} \xi_{t-1-j}\|^*.$$
In the summation above there are \(2m + 1\) terms, and hence using (16)

\[
(2m + 1)^{1 - 2r} \| Y_t \|^{2r} \leq \prod_{j=0}^{m-1} \| A(S'Y_{t-1-j}) \|^{2r} \| Y_{t-m} \|^{2r} + \| \varepsilon \|^{2r} \\
+ \sum_{j=0}^{m-2} \prod_{i=0}^{j} \| A(S'Y_{t-1-i}) \|^{2r} \| \varepsilon \|^{2r} \\
+ \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \\
+ \sum_{j=0}^{m-2} \prod_{i=0}^{j} \| A(S'Y_{t-1-i}) \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r}.
\]

Denote \(\| \varepsilon \|^{2r} = \nu^*\) and note that \(\| A(\cdot) \|^{2r} \leq \rho^2 \| \cdot \| \| \varepsilon \|^{2r} \leq \nu^* B\) for some finite \(B\) (because \(b(\cdot)\) is bounded), \(\| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \leq \nu^* \rho^r \| \cdot \| \| \varepsilon \|^{2r}\), and \(\| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r} \| \varepsilon \|^{2r}\).

Thus,

\[
(2m + 1)^{1 - 2r} E \| Y_t \|^{2r} | Z_{t-m} = z \] \\
\leq \prod_{j=0}^{m-1} \rho^2 \| y \|^{2r} + \nu^* B + \sum_{j=0}^{m-2} \left( \prod_{i=0}^{j} \rho^2 \right) \nu^* B \\
+ \nu^* \rho^{2r} \| y \|^{2r} + \sum_{j=0}^{m-2} \left( \prod_{i=0}^{j} \rho^2 \right) \nu^* E [\nu^* \rho^r (Z_{t-2-j}) | Z_{t-m} = z] \gamma_{2r} \\
= \rho^{2r} \| y \|^{2r} + \nu^* B \left( 1 + \sum_{j=0}^{m-2} \rho^{2r(j+1)} \right) \\
+ \nu^* \gamma_{2r} E [\nu^* \rho^r (Z_{t-1}) | Z_{t-m} = z] + \nu^* \gamma_{2r} \rho^{2r(j+1)} E [\nu^* \rho^r (Z_{t-2-j}) | Z_{t-m} = z] \\
\leq \rho^{2r} \| y \|^{2r} + \nu^* B \left( 1 + \sum_{j=0}^{m-2} \rho^{2r(j+1)} \right) \\
+ \nu^* \gamma_{2r} \left( m^{r-1} \delta^{m-1} \rho^r (z) + d' \right) \\
+ \nu^* \gamma_{2r} \left( \sum_{j=0}^{m-3} \rho^{2r(j+1)} (m^{r-1} \delta^{m-2-j} \rho^r (z) + d') + \rho^{2r(m-1)} \rho^r (z) \right),
\]

where the last inequality makes use of (18) and the fact that \(E [\nu^* \rho^r (Z_{t-m}) | Z_{t-m} = z] = \)
\( h^r(z) \). Defining \( \phi = \max\{\rho^{2r}, \delta\} < 1 \) and \( \phi' = \frac{1}{1-\phi} \) we get

\[
(2m+1)^{1-2r} E \left[ \|Y_t\|^{2r} | Z_{t-m} = z \right] 
\leq \phi'' \|y\|^{2r} + \phi'' B \left( 1 + \sum_{j=0}^{m-2} \phi''^{j+1} \right) + \phi'' \gamma_{2r} \left( m^{r-1} \phi''^{m-1} h^r(z) + d' \right)
\]

\[
+ \phi'' \gamma_{2r} \left( \sum_{j=0}^{m-3} \phi''^{j+1} \left( m^{r-1} \phi''^{m-2-j} h^r(z) + d' \right) + \phi''^{m-1} h^r(z) \right)
\]

\[
\leq \phi'' \|y\|^{2r} + \phi'' B \phi' + \phi'' \gamma_{2r} \left( m^{r-1} \phi''^{m-1} h^r(z) + d' \right)
\]

\[
+ \phi'' \gamma_{2r} \left( \sum_{j=0}^{m-3} \phi''^{m-1} m^{r-1} h^r(z) + \sum_{j=0}^{m-3} \phi''^{j+1} d' + \phi''^{m-1} h^r(z) \right)
\]

\[
\leq \phi'' \|y\|^{2r} + \phi'' B \phi' + m \cdot \phi'' \gamma_{2r} m^{r-1} \phi''^{m-1} h^r(z) + \phi'' \gamma_{2r} d' \phi'
\]

\[
= \phi'' \|y\|^{2r} + \phi'' \gamma_{2r} m^{r-1} \phi''^{m-1} h^r(z) + \phi'' \phi'(B + \gamma_{2r} d'). \tag{19}
\]

Combining the inequalities (17) (with \( k = m \)) and (19) yields

\[
E [V_s(Z_t) \mid Z_{t-m} = z] 
= E \left[ 1 + C \|Y_t\|^{2r} + h^r(Z_t) \mid Z_{t-m} = z \right]
\leq 1 + C(2m+1)^{2r-1} \left( \phi'' \|y\|^{2r} + \phi'' \gamma_{2r} m^{r-1} \phi''^{m-1} h^r(z) + \phi'' \phi'(B + \gamma_{2r} d') \right)
\]

\[
+ (m+1)^{r-1} (h^r(z)\delta^m + d)
\]

\[
= 1 + C \left[ (2m+1)^{2r-1} \phi'' \right] \|y\|^{2r} + \left[ C(2m+1)^{2r-1} \phi'' \gamma_{2r} m^{r-1} \phi''^{m-1} (m+1)^{r-1} \delta^m \right] h^r(z)
\]

\[
+ \left[ C(2m+1)^{2r-1} \phi'' \phi'(B + \gamma_{2r} d') + (m+1)^{r-1} d \right]. \tag{20}
\]

Since \( 0 < \delta \leq \phi < 1 \), we can clearly choose an \( m \) large enough so that both of the expressions in square brackets in (20) are smaller than some \( \lambda < 1 \). The expression in curly brackets in (20) is clearly finite, and thus for some \( L < \infty \) we have

\[
E [V_s(Z_t) \mid Z_{t-m} = z] \leq \lambda \left( 1 + C \|y\|^{2r} + h^r(z) \right) + L. \tag{21}
\]

What remains to be examined is the behaviour of (21) on and off a small set. To this end, write the right-hand-side of (21) as

\[
\lambda^{1/2} \left( 1 + C \|y\|^{2r} + h^r(z) \right) \cdot \lambda^{1/2} \left( 1 + \frac{L}{\lambda (1 + C \|y\|^{2r} + h^r(z))} \right). \tag{22}
\]

By Lemma 4 the set \( A_N = \{ z \in \mathcal{Z} : \|y\|^{2r} \leq N, h^r(z) \leq N \} \) is small. Off this set either \( \|y\|^{2r} > N \) or \( h^r(z) > N \), and the ratio in (22) can clearly be made arbitrarily small by choosing \( N \) large enough. Therefore for a large enough \( N \)

\[
\lambda^{1/2} \left( 1 + \frac{L}{\lambda (1 + C \|y\|^{2r} + h^r(z))} \right) < 1
\]
and hence
\[ E[V(Z_t) | Z_{t-m} = z] \leq \lambda^{1/2} \left( 1 + C\|y\|^2 + h^r(z) \right) \]
on off the set \( A_N \). On the other hand, the right hand side of (21) is clearly bounded on the set \( A_N \) (the smoothness requirement in Assumption 4(a) implies that \( h^r(z) \) is bounded on \( A_N \)). Therefore, condition (15) is satisfied.

**Lemma 6** Let \( X_t \) be an irreducible and aperiodic Markov chain on a state space \( \mathcal{X} \), and let \( m \) be a positive integer. Suppose that for a small set \( K \), a function \( V : \mathcal{X} \to [1, \infty) \) bounded on \( K \), and positive constants \( \lambda < 1 \) and \( b < \infty \)
\[ E[V(X_t) | X_{t-m} = x] \leq \lambda^m (V(x) + b1_K(x)) \] (23)
for all \( x \in \mathcal{X} \). Then \( X_t \) is \( V \)-geometrically ergodic in the sense of Liebscher (2005).

**Proof.** If \( m = 1 \) then \( X_t \) is \( V \)-geometrically ergodic in the sense of Meyn and Tweedie (1993) by their Theorem 15.0.1, and hence the stated weaker form of geometric ergodicity also follows. Suppose now that \( m > 1 \). It immediately follows from Theorem 19.1.3 of Meyn and Tweedie (1993) that \( X_t \) is geometrically ergodic and for some \( g < 1 \) and \( R < \infty \)
\[ \|P^n_X(x, \cdot) - \pi_X(\cdot)\| \leq \rho^g RV(x), \]
where \( \|\cdot\| \) signifies the total variation norm, and \( P^n_X(x, \cdot) \) and \( \pi_X(\cdot) \) are the \( n \)-step transition probability measure and stationary measure of \( X_t \), respectively. What remains to be proven is that the expectation \( \int_{\mathcal{X}} \pi_X(dy)V(y) \) is finite. To this end, we will first establish that \( X_{tm} \), the \( m \)-skeleton of \( X_t \), is \( V \)-geometrically ergodic in the sense of Meyn and Tweedie (1993). By Proposition 5.4.5(iii) of Meyn and Tweedie (1993) and the assumptions of the lemma, the \( m \)-skeleton is irreducible and aperiodic, and satisfies the drift criterion (23) where the set \( K \) is small for the original chain \( X_t \) (but not necessarily for the \( m \)-skeleton). To establish a drift criterion with a set that is small for the \( m \)-skeleton, first choose a \( \lambda \) such that \( \lambda^m < \lambda < 1 \). By Lemma 14.2.8 of Meyn and Tweedie (1993) we can find a set \( K_m \) which is small for the \( m \)-skeleton and such that
\[ 1_K(x) \leq \sum_{i=0}^{m-1} \int_{\mathcal{X}} P^i_X(x, dy) 1_K(y) \leq m1_{K_m}(x) + (\lambda - \lambda^m)/\lambda^mb. \]
Therefore
\[ E[V(X_t) | X_{t-m} = x] \leq \lambda^m V(x) + \lambda^mb \left[ m1_{K_m}(x) + (\lambda - \lambda^m)/\lambda^mb \right] \leq \lambda^m V(x) + (\lambda - \lambda^m) + \lambda^mbm1_{K_m}(x) \leq \lambda V(x) + \lambda^mb1_{K_m}(x) \]
because \( 1 \leq V(x) \). Thus the \( m \)-skeleton satisfies a drift criterion with a set \( K_m \), which is small for the \( m \)-skeleton. Therefore by Theorem 15.0.1 of Meyn and Tweedie (1993) the \( m \)-skeleton is \( V \)-geometrically ergodic in the sense of Meyn and Tweedie (1993).

To complete the proof, note that by Theorem 10.4.5 of Meyn and Tweedie (1993) the stationary distributions of the \( m \)-skeleton of \( X_t \) and \( X_t \) itself are the same and, by the \( V \)-geometric ergodicity of the \( m \)-skeleton, the expectation \( \int_{\mathcal{X}} \pi_X(dy)V(y) \) is finite. □
References


REFERENCES


REFERENCES


EFI, The Economic Research Institute

Reports since 2001
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title.

2006
Books

Dissertations
Gaspar, Raquel M. Credit Risk and Forward Price Models.

2005
Books
Andersson, Per, Susanne Hertz and Susanne Sweet (eds). Perspectives on market networks - boundaries and new connections.
Charpentier, Claes. IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.
Dembrower, Maria. Entreprenörskap i industriella nätverk.
Samuelson, Lars A. Organizational governance and control - a summary of research in the Swedish society.

Dissertations
Andersson, Martin. Making a Difference - Project Result Improvement in Organizations.
Arvidsson, Per. Styrning med belöningssystem - Två fallstudier om effekter av belöningsystem som styrmedel.
Bems, Rudolfs. Essays in International Macroeconomics.
Björckman, Hans. Learning from members - Tools for strategic positioning and service innovation in trade unions.
Clapham, Eric. Essays in Real Estate Finance.
Dareblom, Jeanette. Prat, politik och praktik - Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.
Langenskiöld, Sophie. Peer Influence on Smoking: Causation or Correlation?
Lindahl, Therese. Strategic and Environmental Uncertainty in Social Dilemmas.
Madestam, Andreas. Developing Credit Markets.
Nordin, Fredrik. Externalising Services - Walking a Tightrope between Industrial and Service Logics.
Tolis, Christofer. Framing the Business - Business Modelling for Business Development.
Östberg, Per. Corporate Disclosure and Investor Recognition.

2004

Books
Ahrne, Göran och Nils Brunsson (red). Regelexplosionen.

Dissertations
Blix, Magnus. Essays in Mathematical Finance - Modelling the Futures Price.
Ivarsson Westerberg, Anders. Papperspolisen - varför ökar administrationen i moderna organisationer.
Jutterström, Mats. Att påverka beslut - företag i EU:s regelsättande.
Larsson, Pär. Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.

Lagerwall, Björn. Empirical Studies of Portfolio Choice and Asset Prices.


Nilsson, Hans. Medborgaren i styrsystemet - beskrivning av VAD och HUR i styrning av kommunal verksamhet.


Pajuste, Anete. Corporate Governance and Controlling Shareholders.


Skallsjö, Sven. Essays on Term Structure and Monetary Policy.


Talia, Krim. The Scandinavian Currency Union, 1873-1924 - Studies in Monetary Integration and Disintegration.

2003

Books


Dissertations

Andersson, Henrik. Valuation and Hedging of Long-Term Asset-Linked Contracts.

Bergman, Malin. Essays on Human Capital and Wage Formation.

Damsgaard, Niclas. Deregulation and Regulation of Electricity Markets.

Eklund, Bruno. Four Contributions to Statistical Inference in Econometrics.

Hakkala, Katarina. Essays on Restructuring and Production Decisions in Multi-Plant Firms.


Lange, Fredrik. Brand Choice in Goal-derived Categories - What are the Determinants?
Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*
Nilsson, Göran. *Processororientering och styrning - Regler, mål eller värderingar?*
Tillberg, Ulrika. *Ledskap och samarbete - En jämförande fallstudie i tre skolor.*
Waldenström, Daniel. *Essays in Historical Finance.*
Wällén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

**2002 Books**

**Dissertations**
Berglund, Johan. *De otillräckliga - En studie av personalspecialisternas kamp för erkännande och status.*
Bolander, Pernilla. *Anställningsbilder och rekryteringsbeslut.*
Kallifatides, Markus. *Modern företagsledning och omoderna företagsledare.*
Kaplan, Michael. *Acquisition of Electronic Commerce Capability - The Cases of Compaq and Dell in Sweden.*
Mähring, Magnus. *IT Project Governance.*
Schenkel, Andrew. *Communities of Practice or Communities of Discipline - Managing Deviations at the resund Bridge.*
Skogsvik, Stina. *Redovisningsmätt, värdererelevans och informationseffektivitet.*
Sundén, David. *The Dynamics of Pension Reform.*


Westling, Gunnar. *Balancing Innovation and Control - The Role of Face-to-face Meetings in Complex Product Development Projects.*


2001

**Books**

Sjöstrand, Sven-Erik. *Center for Management and Organization 50 (1951-2001).*

Charpentier, Claes. *Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämnardsreformen.*

Hedlund, Andreas. *Konsumentens erfarenhet - och dess inverkan på livsmedelsinkp på Internet.*

Hvenmark, Johan. *Värför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.*


Ljunggren, Ulrika. *Nyckeltal i grundskolan i Stockholms stad före och efter stadsdelsnämnardsreformen.*

Thorén, Bertil. *Stadsdelsnämnardsreformen och det ekonomiska styrsystemet - Om budgetavvikelse.*

**Dissertations**


Björklund, Christina. *Work Motivation - Studies of its Determinants and Outcomes.*


Engström, Stefan. *Success Factors in Asset Management.*


Eriksson, Rickard. *Price Responses to Changes in Costs and Demand.*

Frisell, Lars. *Information and Politics.*


Hägglund, Peter B. *Företaget som investeringsobjekt - Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.*

Höök, Pia. *Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.*

Johansson, Christer. *Styrning för samordning.*
Josephson, Jens. Evolution and Learning in Games.
Lindkvist, Bjöorn. Kunskapsöverföring mellan produktutvecklingsprojekt.
Löf, Mårten. On Seasonality and Cointegration.
Matros, Alekandr. Stochastic Stability and Equilibrium Selection in Games.
Mårtensson, Pr. Management Processes - An Information Perspective on Managerial Work.
Norberg, Peter. Finansmarknadens amoralitet och det kalvinska kyrkorummet - En studie i ekonomisk mentalitet och etik.
Persson, Bjöorn. Essays on Altruism and Health Care Markets.
Skoglund, Jimmy. Essays on Random Effects Models and GARCH.