

BAYESIAN TIME SERIES AND PANEL MODELS: UNIT ROOTS, DYNAMICS AND RANDOM EFFECTS

Mickael Salabasis

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
Fredagen den 30 Januari 2004
kl. 13.15 i sal Ragnar
Handelshögskolan, Sveavägen 65



Bayesian Time Series and Panel Models: Unit Roots, Dynamics and Random Effects



STOCKHOLM SCHOOL OF ECONOMICS

EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in 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.

EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization

The research activities are organized in twenty Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT

Management and Organisation; (A)

Center for Ethics and Economics; (CEE)

Center for Entrepreneurship and Business Creation; (E)

Public Management; (F)

Information Management; (I)

Center for People and Organization; (PMO)

Center for Innovation and Operations Management; (T)

Prof Sven-Erik Sjöstrand

Adj Prof Hans de Geer

Prof Carin Holmquist

Prof Nils Brunsson

Prof Mats Lundeberg

Prof Jan Löwstedt

Prof Christer Karlsson

ECONOMIC PSYCHOLOGY

Center for Risk Research; (CFR)

Economic Psychology; (P)

Prof Lennart Sjöberg

Prof Guje Sevón

MARKETING

Center for Consumer Marketing; (CCM)

Center for Information and Communication

Research; (CIC)

Marketing, Distribution and Industrial

Dynamics; (ID)

Acting Prof Magnus Söderlund

Adj Prof Bertil Thorngren

Prof Björn Axelsson

ACCOUNTING, CONTROL AND CORPORATE FINANCE

Accounting and Managerial Finance; (B)

Managerial Economics; (C)

Prof Lars Östman

Prof Peter Jennergren

FINANCE

Finance; (FI)

Prof Clas Bergström

ECONOMICS

Center for Health Economics; (CHE)

International Economics and Geography; (IEG)

Economics; (S)

Prof Bengt Jönsson

Prof Mats Lundahl

Prof Lars Bergman

ECONOMIC STATISTICS

Economic Statistics; (ES)

Prof Anders Westlund

LAW

Law; (RV)

Prof Erik Nerep

Chairman of the Board: Prof Håkan Lindgren. *Director:* Associate Prof Bo Sellstedt.

Address

EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/

Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

BAYESIAN TIME SERIES AND PANEL MODELS: UNIT ROOTS, DYNAMICS AND RANDOM EFFECTS

Mickael Salabasis



STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute



Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2004

©EFI and the author
ISBN 91-7258-632-X

Keywords: Bayesian analysis; time series; panel data; unit roots; seasonality; random effects;
Markov chain Monte Carlo

Printed by:
Elanders Gotab, Stockholm 2004

Distributed by:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm, Sweden

To my family
Καὶ το ὅλον τοῦ μέρους μείζον [ἐστίν].

Contents

| | |
|---|------------|
| Acknowledgments | vii |
| I Summary of Thesis | 1 |
| II The papers | 15 |
| 1 Seasonality, Cycles and Unit Roots | 17 |
| 1.1 Introduction | 19 |
| 1.2 Modelling the features of interest | 20 |
| 1.2.1 Stochastic and deterministic trends | 21 |
| 1.2.2 Dynamics | 22 |
| 1.2.3 The likelihood | 23 |
| 1.3 The Prior Structure | 24 |
| 1.3.1 Stochastic and deterministic trends | 24 |
| 1.3.2 Lag polynomial prior | 25 |
| 1.3.3 Polynomial trend prior | 27 |
| 1.3.4 The innovation variance | 28 |
| 1.4 The Posterior Simulation | 28 |
| 1.4.1 Implementation | 28 |
| 1.4.2 Updating the trend properties | 30 |
| 1.4.3 Updating the lag polynomial | 32 |
| 1.4.4 Updating the innovation variance | 37 |
| 1.4.5 Sampling latent initial values | 38 |
| 1.5 The Swedish GDP | 38 |
| 1.6 Final remarks | 42 |

| | | |
|----------|--|-----------|
| 2 | Bayesian Panel Unit Root Analysis | 45 |
| 2.1 | Introduction | 47 |
| 2.2 | The model | 48 |
| 2.2.1 | A panel of time series | 48 |
| 2.2.2 | Conditional likelihoods | 49 |
| 2.3 | Prior structure | 51 |
| 2.3.1 | A natural hierarchical structure | 51 |
| 2.3.2 | Real roots | 52 |
| 2.3.3 | Complex conjugate pairs | 52 |
| 2.3.4 | Weights | 53 |
| 2.3.5 | Covariance matrix prior | 55 |
| 2.4 | Posterior Analysis | 55 |
| 2.4.1 | Real roots | 56 |
| 2.4.2 | Complex conjugate pairs | 57 |
| 2.4.3 | Weights | 57 |
| 2.4.4 | Covariance matrix | 58 |
| 2.4.5 | Latent initial values | 58 |
| 2.5 | Purchasing power parity | 58 |
| 2.6 | Final remarks | 64 |
| 3 | Parametric Covariance Matrix Modelling in Bayesian Panel Regression | 69 |
| 3.1 | Introduction | 71 |
| 3.2 | Parametric Covariance Matrix Modelling | 72 |
| 3.2.1 | An illustrative example | 72 |
| 3.2.2 | A general direct approach | 75 |
| 3.3 | Posterior Simulation | 76 |
| 3.3.1 | The model | 76 |
| 3.3.2 | A generic algorithm | 77 |
| 3.4 | Relative Performance | 78 |
| 3.4.1 | Determinants of corporate investment | 84 |
| 3.4.2 | Firm size effect | 86 |
| 3.5 | Model Selection | 87 |
| 3.5.1 | Extending the prior | 87 |
| 3.5.2 | Adjusting the posterior simulation | 91 |
| 3.5.3 | A simple illustration | 93 |
| 3.6 | Final Remarks | 95 |

| | | |
|----------|---|------------|
| A | Two algorithms | 99 |
| A.1 | The simple multiple block algorithm | 99 |
| A.2 | An improved multiple block algorithm | 100 |
| | | |
| 4 | Serial Correlation and Random Effects in Bayesian Panel Regression | 101 |
| 4.1 | Introduction | 103 |
| 4.2 | Panel regression with serial correlation | 103 |
| 4.3 | The mixed effects model with serial correlation | 107 |
| 4.4 | Model Selection | 112 |
| 4.5 | Discussion | 113 |

Acknowledgments

Looking back I am able to pick out two major themes in my life. The first one is patience. The second is surprise.

I would like to thank the department for their patience and for letting me nourish my inquisitive side. I have learned many things that I honestly believe make me a better econometrician. Unfortunately it has mostly been about what works badly, not at all or, perhaps worst of all, not as expected. Sadly, to the best of my knowledge the demand for a thesis about things that did not work out is negligible. Otherwise I feel quite confident I would have been done a long time ago. I would also like to thank my parents and sister who have been paragons of patience; always there, always giving me their unconditional support, and always accepting my decisions. But most of all I am thankful for the infinite patience of my beloved wife Susanne. Together with our darling daughter Ellen, you define my world.

Besides patience, the other main theme is surprise. It was an unlikely sequence of events that eventually brought me to the Stockholm School of Economics. And when I got here I knew one thing with absolute certainty: time series is not for me. To make a long story short, with the exception of the first chapter which is about a single time series, this thesis is mostly about many of them. What happened? I have no clue. And of course, the biggest surprise of them all is that I am actually done. The latter I largely owe to my advisor Sune Karlsson who contributed in a major way towards the completion of this thesis, not only as the coauthor of the two first papers.

I would also like to take this opportunity to thank a few special persons that in one way or another have made a difference in my life. First of all the Fat Food Society and in particular Patric Dahlquist for being such a good friend. Also I am indebted to Sotirios Tserepis; your perseverance in the face of adversity has served as an inspiration to me. All the people from the department of statistics at Stockholm University, where it all started. In particular Michael Carlson, with whom I wrote my first published paper, and Mattias Villani, who is the source of inspiration for the second half of this

thesis. John Ekberg, my long-time late-lunch buddy, and the 602 all star team. Finally, my special thanks go to Jana Eklund without whom this thesis would most likely be just a bunch of scrappy notes and papers tied together by a string. Tempting as it is to blame her for all the mistakes in it, they do remain all mine.

Stockholm, December 2003

Mickael Salabasis

Part I

Summary of Thesis

Introduction

In the literature, the autoregressive (AR) model appears in two contexts. The first is as a model in its own right, within the framework introduced by Box and Jenkins (1970), and the second is as a regression model with autocorrelated errors, as in Savage (1972) and Zellner and Tiao (1964). In economics, the idea of a contemporaneous shock having an effect that lasts over a period of time is considered not only plausible but also probable. That the AR model features prominently in time series econometrics is then only natural. Its proven track record, flexibility and versatility has kept it a favorite tool of the applied economist. Quite remarkably, even though aspects of the model have been scrutinized continuously, original work is still being added to an ever growing body of literature.

Collecting a number of time series they may be organized in a panel. The potential benefits usually stated include more degrees of freedom and increased efficiency, the possibility of controlling for individual heterogeneity, and also the ability to detect and measure effects that are not observable in either a time series or a cross section alone. It is important to note how to realize any potential benefit it is necessary that the series of the panel share some characteristics; that they are in some broad sense similar. A popular class of models are the random effects models where this dependence often is implicitly introduced through a covariance matrix common across individuals. Much of the influential work in this area has recently been collected in Nerlove (2002). Vintage Bayesian treatment of this class of models, in a non-econometric context, can be found in Box and Tiao (1973). Closer to home, Hsiao, Appelbe and Dineen (1993) is a good starting point.

This thesis consists of four papers and the main theme present is dependence, through time as in serial correlation, and across individuals, as in random effects. The individual papers may be grouped in many different ways. As is, the first two are concerned with autoregressive dynamics in a single time series and then a panel context, while the subject of the final two papers is parametric covariance modelling. Though set in a panel context the results in

the latter are generally applicable.

The approach taken is Bayesian. This choice is prompted by the coherent framework that the Bayesian principle offers for quantifying uncertainty and subsequently making inference in the presence of it. Recent advances in numerical methods have also made the Bayesian choice simpler and Gamerman (1997) is a nice introductory text on the subject. For a convincing presentation of the paradigm see Bernardo and Smith (1994). Other good sources of general information on Bayesianism are Howson and Urbach (1993) and the classical text of Savage (1972). For the application to econometrics, an introductory text that has managed to stand the test of time is Zellner (1971). A more recent presentation is available in Bauwens, Lubrano and Richard (1999). I am convinced that we are all Bayesians, only that some of you have not discovered it yet. As there are 46656 interpretations of Bayesianism according to Good (1983), there is bound to be a flavor for everyone.

Summary and the main results of the papers

Representing the AR model there are a number of options available. One possibility is in terms of the reciprocal roots to the characteristic polynomial. The interest in this representation stems mainly from the fact that these roots may be given a physical interpretation in terms of interesting features. Examples of features that may be associated with the physical state of roots are persistence, cycles and seasons. While all have a long history in the literature, the issue of persistence, unit roots, the question of stochastic versus deterministic trends or permanent versus transitory shocks has attracted most of the attention in recent years. Introducing a singularity in the likelihood, unit roots are problematic regardless of approach taken. In the classical framework, with unit root tests exhibiting numerous primarily power related problems, considerable effort has been expended on generalizing the methods to panels of time series. Two major concerns then serve to complicate matters, the first being model uncertainty and the second being the treatment of cross-section correlation. In this context the superiority of the Bayesian approach is confirmed as it allows a seamless and coherent introduction of the former and simple means to handle the latter.

In the AR model the source of dependence is the propagation of errors through time. Contrasting, in the random effects models the source is typically the measurement of the same individual across time. Ignoring this heterogeneity can have profound effects and lead to considerable loss in efficiency if inference on the individual level is required. While the modern Bayesian solution uses the powerful device of data augmentation, a direct approach is always possible. Not necessarily always practical, the implementation can be described by a simple and generally applicable template. The standard simple random effects models are examples of models where the parametrization of the covariance matrix is simple, treating the covariance matrix explicitly is practical, and the computations can be kept at a minimum.

Seasonality, Cycles and Unit Roots

Featuring prominently in the literature, seasonality, cycles, trends and unit roots are time series characteristics of particular interest. Huerta and West (1999a, 1999b) introduce a flexible model to conduct inference directly on the roots of the autoregressive polynomial. One of the things that make their approach attractive is that it allows immediate inference on at least two of features of interest, namely cycles and unit roots. Another appealing feature is the simplicity with which uncertainty about the lag length and uncertainty about the number and nature of unit roots is introduced. It is achieved by letting individual roots be either null or non-null and thus contribute to an effective rather than fix lag length.

This paper extends their model in three directions. First, we offer a new interpretation of the component mixture priors used. Second, we incorporate seasonal components which makes it possible to analyze the presence of seasonal unit roots and stationary seasonal features. Third we generalize the model to allow for a polynomial trend of arbitrary degree. The resulting highly flexible model robustifies against misspecification by implicitly averaging over different lag lengths, number of unit roots and specifications for the deterministic trend.

An application to the Swedish real GDP illustrates the rich set of information about the dynamics of a time series that can be extracted using this modelling framework.

Bayesian Panel Unit Root Analysis

While Dickey and Fuller (1978) is commonly held as the first modern reference on the topic of unit roots, it is the study by Nelson and Plosser (1982) that signals the start of an ever expanding literature on the subject. From a Bayesian perspective, a now slightly dated overview is found in Sims and Uhlig (1991). Other rich sources of information are the special issues of the *Journal of Applied Econometrics* (volume 6(4) of 1991) and *Econometric theory* (volume 10(3-4) of 1994). As can be expected, much of the focus of the literature is on the role of the prior. The development and application of panel unit root tests is partly prompted by the frequent failure of the standard ADF test, attributed to its low power over short spans of data, persistent but mean-reverting dynamics under local AR consistent with near unit root alternatives, as well as fractional and trend-break alternatives.

Exploiting cross-section variation to improve the power properties of classical tests has been proposed by, among others, Quah (1994), Levin, Lin and

Chu (2002), and Im, Pesaran and Shin (2003). To a varying degree, the most popular panel unit root tests impose some homogeneity restrictions and there is also some confusion on the issue of what the null and alternative should be. A possible source of this confusion is the inability, in the classical framework, to formulate a hypothesis that actually corresponds to the composite question of main interest. Can the panel be modeled jointly, either as stationary or difference stationary, and if not how do the series sort themselves into the categories of interest? Attempting to come to terms with the issues of panel composition, model uncertainty and correlation, we propose a multivariate extension of the model in the first paper of this thesis. Limiting the scope, but at the same time simplifying matters considerably, the mean model is dropped restricting the applicability to non-trending panels. This is not a problem in the example used to illustrate the performance.

Adjusting the tools developed in the previous paper, the approach is tested on a small panel of real exchange rates. Overall, the results are encouraging. As expected, cross-section correlation is present but not homogenous. On a side note, this means that the centering over time, advocated by some as a solution to the cross-correlation issue, will not work properly. The sampler works reliably and apparently converges quickly and cleanly. One of the more important improvements that our panel approach offers is in terms of more distinct inference with respect to unit root composition in the collection of time series. The suggested approach will, in the presence of correlation, typically have a double effect. First, the marginal probabilities of unit roots associated with each series is affected. Second, the joint outcomes will not be independent.

Experimenting with the model raises some issues that are worth considering. One is that the default prior choices tend to favor large models. This in itself is not a problem, but indicates that care should be taken when specifying the prior.

Parametric Covariance Modelling

Many interesting panel models imply a structure on the covariance matrix in terms of a small number of parameters. Exploiting this structure it is demonstrated how common panel data models lend themselves to direct sampling of the variance parameters. This is of interest since it makes it possible to avoid the use of data augmentation and the poor convergence properties sometimes associated with data augmentation. For the method to be practical, simple to program and quick to execute, it is essential that the inverse of the covariance matrix can be written as a reasonably simple function of the parameters of

interest. Also preferable but in no way necessary is the availability of a computationally convenient expression for the determinant as well as a bounded support for the parameter. The latter is truly only a matter of convenience as it simplifies the approximation of the full conditional posterior subsequently used either in a Metropolis type update, applying the gridgy Gibbs sampler of Ritter and Tanner (1992), or used to form the basis of any resampling method.

Using the described generic template, the computations involved in direct sampling are illustrated in the context of a one-way random effects model. The modern standard using data augmentation is used as a benchmark to evaluate the relative performance of the suggested approach. The comparison is made for one panel with small cross-section and a modest number of dates, the typical situations in applied work involving macroeconomics panels, and one panel with a large cross-section and a small number of dates, the typical situation in microeconomics panels. Two standard algorithms are used for the comparison. The first, a 3-block algorithm known to be prone to poor mixing, and one improved algorithm due to Chib and Carlin (1999). The results are very encouraging, direct sampling performing at least as well as the better of the two.

One of the main motivations of the direct approach in this particular case is the simplicity with which model selection can be administrated. With a single parameter governing the presence or absence of the effect, the status of this single parameter is the only one that needs to be toggled. Compared with any data augmentation approach, needing to keep track of a number of parameters equalling the cross-section size, the necessary additional computations are minimal. Effect selection is then illustrated on the smaller of the panels above in the context of a two-way random effects model.

Passing an overall judgement, the direct sampling approach for parametric covariance models constitutes a viable alternative to the standard solutions. In the context of error component models, this is true in particular as the dimensions of the panel increases as the number of latent effects grows with them.

Serial Correlation and Random Effects

Having established direct sampling as a viable alternative in the previous paper, the generic template is applied to panel models with serial correlation. In this case direct sampling may arguably represent the only risk free alternative, in particular when the time series dimension is small as the direct approach does not need to consider any initial conditions. In the case of pure serial correlation, with no random effects present, applying the template and using

a Jeffreys type prior leads to very simple computations. The posterior simulation can then actually be performed by means of a sequence of simple Gibbs steps. This simplifies models selection should we opt for it and makes the sampler very fast.

In the very general setting of a mixed effects model with autocorrelated errors direct sampling of all variance parameters does not appear to be possible or at least not obviously practical. Making progress, one possibility is a conditional approach as in Chib (1993) and Chib and Greenberg (1994). The initial condition is either assumed given, that is the series is truncated, or it is modelled as a latent variable. In either case inference is conducted within the standard regression framework. The other possibility is an exact approach which circumvents the issue of the initial condition. However, inference is typically complicated unless data is augmented with the random effects. Then, conditioning on the random effects, the correlation parameter can be sampled using the methods developed for the fixed effects model. There is however at least one important special case when direct sampling is possible and that is the random individual effects model with autocorrelation. In this case the simple structure of the variance-covariance matrix make direct sampling of both the random effects variance and the correlation parameter possible.

Bibliography

- Bauwens, L., Lubrano, M. and Richard, J.-F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.
- Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, Wiley.
- Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, Holden Day, San Fransisco.
- Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Wiley Classics, New York.
- Chib, S. (1993), ‘Bayes regression with autocorrelated errors’, *Journal of Econometrics* **58**, 275–294.
- Chib, S. and Carlin, B. P. (1999), ‘On MCMC sampling in hierarchical longitudinal models’, *Statistics and Computing* **9**, 17–26.
- Chib, S. and Greenberg, E. (1994), ‘Bayes inference in regression models with ARMA(p,q) errors’, *Journal of Econometrics* **64**, 183–206.
- Dickey, D. A. and Fuller, W. A. (1978), ‘Distribution of the estimator for autoregressive time series with a unit root’, *Journal of the American Statistical Association* **74**, 427–431.
- Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman Hall, London.
- Good, I. J. (1983), *Good Thinking : The Foundations of Probability and its Applications*, University of Minnesota Press, Minneapolis.
- Green, P. J. (1995), ‘Reversible jump markov chain monte carlo computation and bayesian model determination’, *Biometrika* pp. 711–732.

- Howson, C. and Urbach, P. (1993), *Scientific Reasoning. The Bayesian Approach*, La Salle, Chicago.
- Hsiao, C., Appelbe, T. W. and Dineen, C. R. (1993), 'A general framework for panel data models with an application to canadian customer-dialed long distance telephone service', *Journal of Econometrics* **59**, 63–86.
- Huerta, G. and West, M. (1999a), 'Bayesian inference on periodicities and component spectral structures in time series', *Journal of Time Series Analysis* pp. 401–416.
- Huerta, G. and West, M. (1999b), 'Priors and component structures in autoregressive time series models', *Journal of the Royal Statistical Society. Series B (Methodological)* .
- Im, K. S., Pesaran, M. H. and Shin, Y. (2003), 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics* **115**, 53–74.
- Levin, A., Lin, C. F. and Chu, C. J. (2002), 'Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties', *Journal of Econometrics* **108**, 1–24.
- Nelson, C. R. and Plosser, C. I. (1982), 'Trends and random walks in macroeconomic time series: Some evidence and implications', *Journal of Monetary Economics* **10**, 139–162.
- Nerlove, M. (2002), *Essays in Panel Data Econometrics*, Cambridge University Press, Cambridge.
- Quah, D. (1994), 'Exploiting cross-section variation for unit root inference in dynamic data', *Economics Letters* **44**, 9–19.
- Ritter, C. and Tanner, M. A. (1992), 'Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler', *Journal of the American Statistical Association* **87**, 861–868.
- Savage, L. (1972), *Foundations of Statistics*, Dover, New York.
- Sims, C. A. and Uhlig, H. (1991), 'Understanding unit rooters: A helicopter tour', *Econometrica* **59**, 1591–1600.
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley Classics, New York.

Zellner, A. and Tiao, G. C. (1964), 'Bayesian analysis of the regression model with autocorrelated errors', *Journal of the American Statistical Association* **59**, 763–778.

Part II

The papers

Paper 1

Seasonality, Cycles and Unit Roots

1.1 Introduction

Seasons, cycles, and unit roots are dominant themes of time series econometrics. Having both important policy implications and a major impact on inference, these features are typically analyzed using the work horse autoregressive model. While inference on the autoregressive parameters is straightforward inference on the features of interest - seasons, cycles and unit roots - is more complicated since they depend on the roots of the autoregressive polynomial. To overcome this difficulty we utilize a novel model formulation and follow Huerta and West (1999*a*, 1999*b*) in writing the model in terms of and conduct inference directly on the roots. Adopting a Bayesian approach, the analysis is facilitated by assigning point mass priors on root configurations corresponding to the features of interest.

In addition to focusing directly on the roots of the process we adopt a flexible modelling framework. In particular we allow for uncertainty about the lag length, uncertainty about the number and nature of unit roots while allowing for polynomial trends in the data. The lag length is allowed to vary by letting individual roots be either null or non-null and thus contribute to the effective lag length. Allowing the number of positive unit roots to vary may cause trend parameters to become unidentified. We circumvent this problem by introducing an element of model selection, treating different combinations of positive unit roots and trend polynomial degrees as distinct models.

This flexibility and ease of interpretation of the results comes at a cost. The main complication arises because the model is nonlinear in the roots of the characteristic polynomial and the posterior distribution is intractable. We use MCMC methods to analyze the posterior distribution with the reversible jump Markov Chain Monte Carlo algorithm of Green (1995) as a unifying computational framework.

The output of the Markov Chain contains a rich set of information on the posterior distribution. The posterior probabilities for the presence of particular features as well as the posterior distribution of the roots or the ordinary autoregressive parameters are easily obtained. In doing this we can either condition on specific features being present in the model (e.g. an AR(3) with one unit root and a constant term) or average over the different combination of features. In the latter case we are in effect performing Bayesian model averaging and the results are robust to model misspecification within the class of models considered here.

The present paper extends the work of Huerta and West in three directions. Firstly we add seasonal components which makes it possible to analyze the presence of seasonal unit roots and stationary seasonal features. Secondly, and

perhaps most importantly, we generalize the model to allow for a polynomial trend of arbitrary degree. This is in contrast to Huerta and West who demean the data and exclude the constant term from the model. Their modelling framework is thus in effect only applicable to non-trending data. Thirdly, we give a different interpretation of the prior on the features of interest which makes it easier to specify the prior and avoids some potential pitfalls.

In related work Franses, Hoek and Paap (1997) propose a Bayesian analysis of seasonal unit roots based on the model formulation of Hylleberg, Engle, Granger and Yoo (1990). While computationally more demanding our approach is much more general, providing a richer set of information about the appropriate model and offering robustness against misspecification through model averaging. The approach of DeJong and Whiteman ((1991*b*), (1991*a*)) to unit root inference is similar in spirit to our work in that they consider the posterior distribution of the largest root of the characteristic polynomial. An important difference is that DeJong and Whiteman do not model the roots directly, instead they model the AR parameters and solve for the roots given these parameters.

The organization of this paper is as follows. The modelling framework is introduced in Section 1.2 and a suitable prior is given in Section 1.3. The necessary tools for posterior analysis by MCMC are developed in Section 1.4. Section 1.5 applies the model to data on the Swedish GDP and section 1.6 offers some closing remarks.

1.2 Modelling the features of interest

In most cases the features of interest correspond to single points in the parameter space. We will thus classify the roots accordingly and later on introduce a prior adapted to this classification. At this stage it suffices to note that a Bayesian treatment of features that correspond to a single point in the parameter space requires a prior that assigns point mass probabilities to each of them. Excluding explosive behavior, the roots of the autoregressive polynomial or, equivalently, the characteristic polynomial can be null, stationary or on the unit circle. By allowing for null roots the effective lag length can be varied and different lag lengths explored. For the active roots the classification into stationary roots and unit roots can be further refined by in each case considering roots at frequencies of special interest (typically seasonal) and roots at arbitrary frequencies. In each case the prior assigns a point mass to the particular feature thus permitting the calculation of posterior probabilities of the feature.

There is one feature of interest that does not fit directly into this framework and that is cycles of a general nature. An example is business cycles where the exact length of the cycle is unknown or may vary over time. Instead of considering a single frequency we need to focus on a range of frequencies. In this case it is not necessary to assign a point mass to the feature and it can be accommodated in the prior for roots at arbitrary frequencies. Still, modelling the roots directly makes inference on cycles of arbitrary length straightforward.

1.2.1 Stochastic and deterministic trends

Consider the autoregression of order p with a general polynomial trend of degree g

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=0}^g \delta_i t^i + \varepsilon_t, \quad (1.1)$$

where ε_t are iid $\mathcal{N}(0, \sigma_\varepsilon^2)$ innovations. Given (p, g) and an adequate number of initial values, this model may be analyzed within the standard framework of multiple regression.

The model formulation (1.1) is problematic in the sense that the interpretation, in terms of the implied trend behavior, depends on the presence or absence of a real positive unit root in the lag polynomial; see Schotman and van Dijk (1991) and Lubrano (1995) for details. For each positive real unit root introduced in the lag polynomial, the interpretation of the trend parameters change. Alternatively, the model can be parameterized as

$$\phi(L) \left(y_t - \sum_{i=0}^g \beta_i t^i \right) = \varepsilon_t$$

where the order of the dynamics does not depend on the number of positive unit roots. Unfortunately, the underlying problem remains and is now manifest as local non-identification of trend parameters as a root approaches unity.

In line with our approach of modelling the features of interest directly we address the problem by explicitly treating each combination of the differencing order, d , and degree of the trend polynomial, g , as giving rise to distinct models with different interpretations. Decomposing $\phi(L)$ into $(1-L)^d \Psi(L)$ where the positive unit roots are excluded from $\Psi(L)$ we can rewrite the model as

$$\begin{aligned} (1-L)^d y_t &= \tau'_t \beta_{d,g} + v_t, \\ \Psi(L) v_t &= \varepsilon_t, \end{aligned} \quad (1.2)$$

where $\tau_t = (t^0, t^1, \dots, t^g)'$ and g is implicitly redefined by subtracting d . This allow us to treat the stochastic and deterministic trends separately. In the posterior analysis we can then condition on a specific choice of d and g or average over them for a more robust analysis of the other features of interest.

While this solves the identification problem it is still the case that the data will be less and less informative about the constant term as the largest root of $\Psi(L)$ approaches unity. There is still the possibility of near non-identification and we will need to account for this in the prior specification.

1.2.2 Dynamics

Factoring the seasonal differencing polynomial

$$1 - L^S = (1 - L)(1 + L + \dots + L^{S-1})$$

we obtain seasonal unit roots as -1 and $\exp\left(\pm \frac{ik\pi}{S/2}\right)$, $k = 1, \dots, S/2 - 1$ for S even. For the less common case of S odd the seasonal roots are instead given by $\exp\left(\pm \frac{ik\pi}{S/2}\right)$, $k = 1, \dots, (S - 1)/2$. In either case, we can add one real root and a set of complex roots to our collection of model features. Stationary seasonal dynamics are also of interest and we allow for complex roots at the seasonal frequencies with modulus less than unity. The set of roots is completed by adding stationary real roots and stationary complex roots with arbitrary modulus and frequency.

Since the complex roots must appear in conjugate pairs the modelling exercise will be simplified if we treat real and complex roots separately. The lag polynomial $\Psi(L)$ will thus be taken to consist of a maximum of $\bar{\tau}$ real roots and \bar{c} complex root pairs for a total maximum order of $\bar{p} = \bar{\tau} + 2\bar{c}$. Each of the $\bar{\tau}$ real roots may be either a null root, a stationary root or a negative unit root. Similarly each of the \bar{c} complex pairs may be null, have an arbitrary frequency with arbitrary stationary modulus, seasonal frequency with stationary modulus or be a seasonal unit root.

It is convenient to parameterize the stationary lag polynomial $\Psi(L)$ in terms of the roots of the characteristic polynomials,

$$\Psi(L) = \prod_{j=1}^{\bar{\tau}} (1 - \rho_j L) \prod_{j=1}^{\bar{c}} (1 - 2m_j a_j L + m_j^2 L^2).$$

For the real roots the possible states then include null, $\rho_j = 0$, stationary, $0 < |\rho_j| < 1$ or a negative unit root, $\rho_j = -1$. For the complex pairs of roots, $m_j \exp(\pm i\omega)$, we have null roots, $m_j = 0$, stationary roots, $0 < m_j < 1$ or

unit roots, $m_j = 1$. For the non-null roots the modulus are combined with the seasonal frequencies, $\omega_j \in \{2k\pi/S, k = 1, \dots, S/2 - 1\}$ or an arbitrary frequency, $0 < \omega_j < \pi$ to obtain the full set of complex roots. For ease of notation we let $a_j = \cos \omega_j$.

1.2.3 The likelihood

The parametrization in terms of the roots simplifies the identification, interpretation, and introduction of interesting features. Unfortunately, while the general form of the likelihood does not change, the root parameters appear in a non-trivial multiplicative way. Conducting joint inference on the full set of parameters is thus complicated. On the other hand, the likelihood is amenable to conditional inference and we can divide the problem into smaller, simple, parts that – provided this structure is maintained by the prior – naturally leads to a MCMC algorithm for analyzing the posterior distribution. In this process, filtering plays an important role.

Given the differencing order, d , and the degree of the polynomial trend, g , define the trend vector $\tau_t = (t^0, t^1, \dots, t^g)$ and $\tilde{y}_t = (1 - L)^d y_t$. Then, for inference on the trend coefficient vector, conditional on the roots of $\Psi(L)$ the model in (1.2) reduces to

$$\begin{aligned} y_t^* &= \tau_t^{*'} \beta_{d,g} + \varepsilon_t, \\ y_t^* &= \Psi(L) \tilde{y}_t, \quad \tau_t^{*'} = \Psi(L) \tau_t' \end{aligned} \quad (1.3)$$

a standard linear regression model.

Conversely, for inference on the roots of the lag polynomial $\Psi(L)$, conditional on $\beta_{d,g}$, let $u_t = (\tilde{y}_t - \tau_t' \beta_{d,g})$ be the centered appropriately differenced data. For a real root ρ_i , conditional on all else, define

$$z_t = \prod_{j \neq i} (1 - \rho_j L) \prod_{j=1}^{\bar{c}} (1 - 2m_j a_j L + m_j^2 L^2) u_t \quad (1.4)$$

and conditional inference on ρ_i can be conducted in

$$(1 - \rho_i L) z_t = \varepsilon_t,$$

an autoregressive process of order one with known variance σ_ε^2 . Similarly for a complex pair, $m_i \exp(\pm i\omega_i)$, define

$$z_t = \prod_{j=1}^{\bar{r}} (1 - \rho_j L) \prod_{j \neq i} (1 - 2m_j a_j L + m_j^2 L^2) u_t. \quad (1.5)$$

Then, conditional on all but the complex pair the model reduces to

$$(1 - \phi_1 L - \phi_2 L^2) z_t = \varepsilon_t, \quad \phi_1 = 2m_i a_i = 2m_i \cos \omega_i, \quad \phi_2 = -m_i^2,$$

an autoregressive process of order two with parameter restrictions.

A remaining issue is the treatment of the initial values. We can either take them as given and use the likelihood for $y_{\bar{p}+d+1}, \dots, y_T$ conditional on $y_1, \dots, y_{\bar{p}+d}$ or use the full likelihood for y_1, \dots, y_T . The former has the advantage of simplicity but can lead to a substantial loss of information if $\bar{p} = 2\bar{c} + \bar{r}$ is large relative to T . Instead, we prefer the latter approach and marginalize out the unobserved initial values $y_{1-\bar{p}-d}, \dots, y_0$, by treating them as latent variables to be simulated together with the parameters in the Markov chain.

1.3 The Prior Structure

The overall prior structure follows from a few reasonable assumptions of independence. We take the seasonal and cyclical behavior to be a priori independent of the trend properties. In addition, the innovation variance, σ_ε^2 , plays a different role in autoregressive models than in standard regression models. In particular it does not influence how informative the data is about the parameters in the autoregressive polynomial and the usual conditioning on σ_ε^2 does not simplify the prior specification. Decomposing the prior on the trend parameters β , d and g as $\pi(\beta_{d,g} | \sigma_\varepsilon^2, d, g) \pi(d, g)$ we obtain the prior as

$$\pi(\theta, \beta_{d,g}, d, g, \sigma_\varepsilon^2) = \pi(\theta) \pi(\beta_{d,g} | \sigma_\varepsilon^2, d, g) \pi(d, g) \pi(\sigma_\varepsilon^2) \quad (1.6)$$

where the characteristic roots of $\Psi(L)$ are collected in

$$\theta = (\rho_1, \dots, \rho_{\bar{r}}, m_1, a_1, \dots, m_{\bar{c}}, a_{\bar{c}}).$$

1.3.1 Stochastic and deterministic trends

In most cases only a few possible values for d and g are considered, i.e. at most two unit roots or at most a quadratic trend. This leads to a limited number of combinations $(d, g) \in (0, \dots, \bar{d}) \times (-1, \dots, \bar{g})$ where then $g = -1$ corresponds to a model with no constant term or trends. In addition, some of these might be ruled out a priori as unreasonable. Assigning prior probabilities is thus fairly straightforward and if nothing else a uniform prior on the possible combinations can be used.

1.3.2 Lag polynomial prior

Turning to the characteristic roots of $\Psi(L)$ we note that the polynomial is invariant to permutations of the roots. Consequently we treat the roots symmetrically and specify identical and independent priors for the the real and complex roots respectively. We write the prior on θ as

$$\pi(\theta) = \prod_{j=1}^{\bar{r}} \pi(\rho_j) \prod_{k=1}^{\bar{c}} \pi(m_k, a_k) \quad (1.7)$$

where $\pi(\rho_j)$ is the prior on an arbitrary real root and $\pi(m_k, a_k)$ the prior on an arbitrary complex root. This prior structure is close to the one used by Huerta and West (1999a, 1999b) with obvious modifications to cater for the differences in model set up.

Real roots

For real roots we distinguish between three states, null roots, stationary roots and negative unit roots. Corresponding to this the prior is a mixture over the three states,

$$\pi(\rho) \sim w_0^r I_{\rho=0} + w_s^r f_{|\rho|<1}(\rho) + w_{-1}^r I_{\rho=-1}, \quad (1.8)$$

where the weights $\{w_i^r\}$ sum to one, $I_{\rho=k}$ is the usual indicator function, and $f_{|\rho|<1}(\rho_j)$ is a continuous distribution with support on the stationarity region.

When interpreting the prior the primary distinction is between null roots and non-null roots which contribute to the effective lag length. The non-null roots can in turn be either stationary or a unit root. This hierarchical structure is useful when thinking about the weights. We suggest choosing the weights as

$$w_0^r = p_0, \quad w_s^r = (1 - p_0) p_s, \quad w_{-1}^r = (1 - p_0) (1 - p_s)$$

where p_0 is the a priori expected proportion of null roots and p_s is the a priori expected proportion of stationary roots given that the root is active. A natural uninformative prior specification given this structure is $p_0 = 1/2$, $p_s = 1/2$. Note that this differs from the uninformative prior, $w_1^r = w_2^r = w_3^r = 1/3$, obtained by considering (1.8) directly. The latter is essentially the prior used by Huerta and West (1999b). Given \bar{r} , the number of real roots, the choice of weights induces a multinomial prior on the number of roots of the different types. In particular, the induced prior on the contribution of the real roots to the overall lag length is binomial, $\text{Bin}(\bar{r}, 1 - w_1^r)$, and it is important to keep this in mind when choosing \bar{r} .

To complete the specification of the prior we need to specify $f_{|\rho|<1}(\rho)$, the prior on stationary roots. Convenient choices are a uniform prior on $(-1, 1)$ and a truncated normal distribution. Both these priors are problematic in the sense that the posterior for the constant term in the deterministic trend will be dominated by the prior as $\rho \rightarrow 1$. In the context of a model with only a constant term Schotman and van Dijk (1991) propose dealing with this problem by making the prior for the constant more concentrated around the initial condition, see Schotman (1994) for a detailed discussion of these issues. The prior of Schotman and van Dijk has the disadvantage that it is data dependent and it is unclear how this should be generalized to trend models of higher order. Instead we take the simple route of specifying a uniform prior on the restricted range $(-1, 1 - \delta)$ for a small value of δ . In connection with this it is important to note that this does not rule out unit roots as these are modeled together with the deterministic trend. Instead we view this as driving a wedge between trend stationary and unit root dynamics and forcing the model to clearly distinguish between them.

Complex roots

The prior for the complex conjugate pairs can be constructed in a similar manner. A root is null with probability p_0 or otherwise non-null. If it is non-null, it may be stationary with probability p_s or on the unit circle. In addition, non-null roots may have a frequency of special interest with probability p_{ω^*} or have an arbitrary frequency with probability $1 - \sum p_{\omega^*}$. In the case of a single frequency of special interest, ω^* , we would have

$$\begin{aligned} \pi(m, a = \cos \omega) \sim & w_0^c I_{m=0} + w_{s,\omega^*}^c f_{m,\omega=\omega^*} + w_{s,\omega}^c f_{m,\omega} \\ & + w_{1,\omega^*}^c I_{m=1,\omega=\omega^*} + w_{1,\omega}^c f_{m=1,\omega} \quad (1.9) \end{aligned}$$

with

$$\begin{aligned} w_0^c &= p_0, \quad w_{s,\omega^*}^c = (1 - p_0) p_s p_{\omega^*}, \quad w_{s,\omega}^c = (1 - p_0) p_s (1 - p_{\omega^*}) \\ w_{1,\omega^*}^c &= (1 - p_0) (1 - p_s) p_{\omega^*}, \quad w_{1,\omega}^c = (1 - p_0) (1 - p_s) (1 - p_{\omega^*}) \end{aligned}$$

where p_0 is the a priori expected proportion of null roots, p_s the proportion of stationary roots given that the root is non-null and p_{ω^*} the proportion of non-null roots with the special frequency ω^* .

Apart from the structure imposed on the prior weights, this is a straightforward extension of the prior used by Huerta and West (1999b). It should be clear that the structure plays a larger role with the complex roots than with the

real roots. It greatly reduces the number of prior parameters to specify and implies a more reasonable default, uninformative, prior with $p_0 = 1/2$, $p_s = 1/2$ and $p_{\omega^*} = \frac{1}{n^*+1}$ where n^* is the number of special frequencies. Recall that the implied prior for the number of non-null complex roots is $Bin(\bar{c}, 1 - w_0^c)$. It is clear that the choice $w_0^c = \frac{1}{2(n^*+1)+1}$ inspired by considering (1.9) directly can lead to an implausible large number of complex roots if, as prudence might dictate, a large value of \bar{c} is used.

To complete the prior specification we need to specify the two continuous components. Following Huerta and West (1999b) a joint uniform prior on the implied autoregressive parameters in (1.5) is used. For the component corresponding to a stationary cycle at an arbitrary frequency, $f_{m,\omega}$, this results in

$$m \sim Beta(3, 1), a \equiv \cos \omega \sim U(-1, \bar{a}).$$

where $\bar{a} = \cos \underline{\omega}$. The upper bound depends on the minimum observable frequency $\underline{\omega}$ which in turn depends on the length of the time series in a trivial way. This leads to a truncated bivariate normal posterior and potential simplifications in the posterior simulation for the implied parameters in the time domain. For the other continuous components added by the presence of special frequencies, a similar argument leads to the prior $f_{m,\omega=\omega^*} \sim Beta(2, 1)$ for the modulus with ω fixed at a special frequency. Finally, for unit roots with arbitrary frequency, the same uniform prior $a \equiv \cos \omega \sim U(-1, \bar{a})$ is used for the frequency.

1.3.3 Polynomial trend prior

By isolating the positive unit roots and then rewriting the model as (1.2) a singularity in the conditional likelihood (1.3) for $\beta_{d,g}$ is avoided at the cost of making the interpretation of $\beta_{d,g}$ depend on the differencing order, d . Specifying a general prior for $\beta_{d,g}$ is thus difficult unless we make it uninformative or vague. A uniform, improper, prior is in fact possible after the removal of the singularity but this choice is problematic if we view the choice of d and g as a model selection exercise. Instead we use a g -prior with the scaling factor selected to make the prior vague. That is,

$$\beta_{d,g} | d, g \sim \mathcal{N}_{g+1}(0, \sigma_\varepsilon^2 M^{-1})$$

where $M = gX'X$ and X the trend matrix with rows $\tau'_t = (t^0, t^1, \dots, t^g)$.

1.3.4 The innovation variance

For the innovation variance, σ_ε^2 , we use the standard inverse gamma prior,

$$\sigma_\varepsilon^2 \sim \mathcal{IG}(a_0, b_0),$$

with a_0 and b_0 selected to make the prior proper but vague. This leads to an inverse gamma full conditional posteriors, simplifying the MCMC computations.

1.4 The Posterior Simulation

1.4.1 Implementation

Exploration of the model space requires occasional changes in the dimension or content of the parameter vector. The dimension changing moves are handled using the reversible jump MCMC algorithm of Green (1995). This entails the construction of moves between the possible states together with their balancing counterparts. Some additional notation is needed when discussing the moves. Let the subscripts c and p denote current and proposed quantities. Collecting all current parameters in $\xi_c = [\theta_c, \beta_c, d_c, g_c, \sigma_c^2]$, the prior of the current state is summarized by $\pi_c = \pi(\xi_c)$. When appropriate, let j_{cp} be the probability of selecting a move that attempts a transition from the current state to the proposed state

The reversible jump algorithm solves the problem of changing dimensions by completing the parameter spaces. Let u_c and u_p be any two parameter vectors satisfying the dimension matching requirement

$$\dim(\xi_p) + \dim(u_c) = \dim(\xi_c) + \dim(u_p)$$

and

$$(u_c, \xi_p) = g(\xi_c, u_p)$$

a bijection. A proposal parameter vector, ξ_p , for the proposed state is then obtained by drawing u_p from a suitable distribution q_p and applying the transformation. As this is a Metropolis-Hastings step the corresponding, imaginary, operations are conducted for the balancing move from the proposed to the current state when forming the acceptance probability,

$$\alpha = \min \left\{ 1, \frac{L(\mathbf{y}|\xi_p)}{L(\mathbf{y}|\xi_c)} \frac{\pi_p}{\pi_c} \frac{j_{pc}q_c(u_c)}{j_{cp}q_p(u_p)} \left| \frac{\partial g(\xi_c, u_p)}{\partial(\xi, u)} \right| \right\},$$

where j_{cp} and j_{pc} are the probabilities of proposing a move that attempts a transition from the current to the proposed state and vice versa.

When $\dim(\xi_c) > \dim(\xi_p)$ it is common to take u_p to be empty and $\dim(u_c) = \dim(\xi_c) - \dim(\xi_p)$. The proposal would not involve any sampling although the balancing move is still conducted as if u_c was a random quantity. In contrast with this we will find it convenient to complete both parameter vectors in some cases. In practice, the proposals involve only a subset of the parameters and the acceptance probabilities simplify considerably by using the appropriate conditional likelihood and exploiting the structure of the prior.

For the particular problem of interest, the structure of the model suggests two major subsets, one containing the trend parameters (d, g) and another containing the collection of all roots. We use four types of moves to explore the (d, g) space. The first (T1) is a local move that holds (d, g) constant and only updates the coefficients of the trend polynomial. The second move (T2) changes the degree of the trend polynomial, $g_p = g_c \pm 1$, while holding the differencing order constant. The third move (T3) is the corresponding move for the differencing order. The final move (T4) change both quantities in such a way as to keep the total order of the trend constant, $d_p + g_p = d_c + g_c$. With the exception of T1, all moves require the execution of a reversible jump.

When exploring the dynamic properties, real and complex roots are treated separately. For real roots, as discussed in Section 1.3.2, a uniform prior over the stationary region for the continuous component leads to updates with a Gibbs step. For the complex roots the updates are more complicated and four different moves are used. Corresponding to the major states of the roots, M1 moves between null and stationary roots and M2 moves between stationary roots and roots on the unit circle. Complementing this, M3 explores the modulus and frequency of stationary roots and M4 explores the frequency of the complex unit roots. Both major moves are reversible jumps. While M3 in theory could be updated with a Gibbs step, the resulting complicated truncated bivariate normal prompts the use of a Metropolis-Hastings step. Contrasting, in M4 it is possible to update the frequency of a pair at the boundary using a procedure similar to that for real roots.

The Markov chain proceeds by iterating the steps in Algorithm 1. In what follows let Y be the vector of original time series measurements, including the appropriate number of latent initial values, and X the matrix of appropriately defined trend variables.

Algorithm 1 Structure of the Markov chain

1. Update the trend properties. Select a new state (d_p, g_p) and attempt a transition to the state using the relevant move in Section 1.4.2
 2. Update the lag polynomial (Section 1.4.3). Establish a cycle of length $l = \bar{r} + \bar{c}$. In each iteration sample a root index at random
 - (a) If the selected index is associated with a real root, update it using the full conditional posterior in Section 1.4.3.
 - (b) If the selected index is associated with a complex pair, query the current status and attempt to update it using one of four moves detailed in Section 1.4.3.
 3. Update the variance σ_ε^2 using the full conditional posterior in Section 1.4.4.
 4. Sample new initial values using the procedure detailed in Section 1.4.5.
-

1.4.2 Updating the trend properties

Propose a new state (d_p, g_p) with equal probability from $\{(d, g) : (d_c, g_c), (d_c \pm 1, g_c), (d_c, g_c \pm 1), (d_c \pm 1, g_c - d + d_c), 0 \leq d \leq \bar{d}, -1 \leq g \leq \bar{g}\}$. Due to the endpoint restrictions the available states vary and the proposal probabilities are $j = 1/3$ for $(d_c, g_c) = (0, -1)$ or (\bar{d}, \bar{g}) , $j = 1/4$ for $(d_c, g_c) = (0, \bar{g})$ or $(\bar{d}, 0)$, $j = 1/5$ for $d_c \in \{0, \bar{d}\}$ and $-1 < g_c < \bar{g}$ or $g_c \in \{-1, \bar{g}\}$ and $0 < d_c < \bar{d}$ and $j = 1/7$ otherwise. Given the proposed state a transition is attempted using one of the moves T1 - T4. Throughout this section we condition on the current lag polynomial and innovation variance and for simplicity of notation we omit θ_c and σ_c^2 from the conditioning variables.

Although we consider four different moves they all fit in a common generic template. The key to the simplicity of the updates is that we draw the proposal for β directly from the full conditional posterior and treat the current value as a draw from the full conditional posterior. That is $q_p(\beta_p) = p(\beta|Y, d_p, g_p)$ and $q_c(\beta_c) = p(\beta|Y, d_c, g_c)$ and $g(\cdot, \cdot)$ is the identity function with unit Jacobian.

To illustrate, for the current model, let

$$E_c = \Psi^c(L) \tilde{Y}_c, \quad \tilde{Y}_c = (1 - L)^{d_c} Y,$$

be the filtered appropriately differenced series, and filter the relevant section

of the trend matrix X to obtain

$$D_c = \Psi^c(L) X_c.$$

With the conditional normal structure, the normal prior leads to the standard full conditional posterior

$$\beta_c \sim \mathcal{N}_{g_c+1}(\bar{\beta}_c, \Omega_c)$$

where

$$\bar{\beta}_c = \Omega_c [\sigma_\varepsilon^{-2} D'_c E_c], \quad \Omega_c = \sigma_\varepsilon^2 [D'_c D_c + M_c]^{-1}.$$

Collecting all factors in the acceptance probability that are associated with the current value, an expression for the current contribution simplifies to

$$\frac{j_{pc} q_c(\beta_c)}{L(\mathbf{y}|\xi_c) \pi_c} = j_{pc} \kappa_c \sqrt{\frac{1}{|\sigma_\varepsilon^2 M_c| |\Omega_c|}} \exp \left\{ \frac{1}{2} \left[\sigma_\varepsilon^{-2} E'_c E_c - \bar{\beta}'_c \Omega_c^{-1} \bar{\beta}_c \right] \right\},$$

where κ_c is used to collect the scale factors from the normal likelihood, prior and proposal. From a similar exercise for the proposed state we get the analogous proposal quantities E_p , D_p , $\bar{\beta}_p$, and Ω_p . Some matrix algebra yields a simplified generic expression for the acceptance probability as

$$\alpha_T = \min \left\{ 1, \frac{j_{pc}}{j_{cp}} \sqrt{\frac{|\Omega_p| |\sigma_\varepsilon^2 M_p|}{|\Omega_c| |\sigma_\varepsilon^2 M_c|}} \frac{\exp \left\{ -\frac{1}{2} \left[\sigma_\varepsilon^{-2} E'_p E_p - \bar{\beta}'_p \Omega_p^{-1} \bar{\beta}_p \right] \right\}}{\exp \left\{ -\frac{1}{2} \left[\sigma_\varepsilon^{-2} E'_c E_c - \bar{\beta}'_c \Omega_c^{-1} \bar{\beta}_c \right] \right\}} \right\}. \quad (1.10)$$

Depending on the selected move, this expression may simplify further.

Move T1 (Updating the polynomial trend parameters)

When the differencing order and the trend degree are the same, the current and proposal quantities in (1.10) are all equal. The acceptance probability simplifies to $\alpha_{T1} = 1$ and this is equivalent to a Gibbs step. Note the special case $g_c = g_p = -1$ where no computations at all are necessary.

Move T2 (Updating the trend polynomial degree)

With $(d_p, g_p) = (d_c, g_c \pm 1)$ the current and proposal models share the same centered differenced series and the acceptance probability simplifies to

$$\alpha_{T2} = \min \left\{ 1, \frac{j_{pc}}{j_{cp}} \sqrt{\frac{|\Omega_p| |\sigma_\varepsilon^2 M_p|}{|\Omega_c| |\sigma_\varepsilon^2 M_c|}} \frac{\exp \left\{ \frac{1}{2} \bar{\beta}'_p \Omega_p^{-1} \bar{\beta}_p \right\}}{\exp \left\{ \frac{1}{2} \bar{\beta}'_c \Omega_c^{-1} \bar{\beta}_c \right\}} \right\}.$$

The expression simplifies further if either the current or proposed model implies a trend free model. For instance, assuming $g_c = -1$ we get

$$\alpha_{T2} = \min \left\{ 1, \frac{j_{pc}}{j_{cp}} \sqrt{|\Omega_p| |\sigma_\varepsilon^2 M_p|} \exp \left\{ \frac{1}{2} \bar{\beta}'_p \Omega_p^{-1} \bar{\beta}_p \right\} \right\}.$$

Move T3 (Updating the differencing order)

With $(d_p, g_p) = (d_c \pm 1, g_c)$ the filtered trend matrix, D , and the prior precision are unaffected with $\Omega_c = \Omega_p$ and the acceptance probability simplifies to

$$\alpha_{T3} = \min \left\{ 1, \frac{j_{pc}}{j_{cp}} \frac{\exp \left\{ -\frac{1}{2} \left[\sigma_\varepsilon^{-2} E'_p E_p - \bar{\beta}'_p \Omega_p^{-1} \bar{\beta}_p \right] \right\}}{\exp \left\{ -\frac{1}{2} \left[\sigma_\varepsilon^{-2} E'_c E_c - \bar{\beta}'_c \Omega_c^{-1} \bar{\beta}_c \right] \right\}} \right\}.$$

Note how the major simplification is the cancelling of the ratio of scale factors. If there is currently no active trend model, that is $g_c = g_p = -1$, the expression simplifies to

$$\alpha_{T3} = \min \left\{ 1, \frac{j_{pc}}{j_{cp}} \exp \left\{ -\frac{1}{2} \left[\sigma_\varepsilon^{-2} (E'_p E_p - E'_c E_c) \right] \right\} \right\}$$

Move T4 (Updating the trend degree and the differencing order)

Setting $(d_p, g_p) = (d_c \pm 1, g_c - d_p + d_c)$ no further simplifications are achieved and the generic expression in (1.10) is used so that $\alpha_{T4} = \alpha_T$.

Note that the acceptance probabilities do not depend on the current or proposed value of β . This suggests the following strategy to improve the mixing properties of the chain. Unless Move T1 is actually selected, sample the next trend state using the appropriate acceptance probability. Then, regardless of the result, sample a new coefficient vector from the relevant full conditional posterior necessary for updating the lag polynomial. Essentially, the updating sequence in Step 1 of Algorithm 1 then always ends with the execution of Move T1.

1.4.3 Updating the lag polynomial

Conditional on all else, and in particular the trend properties d, g and $\beta_{d,g}$, the data can be filtered as outlined in Section 1.2.3 to isolate the contribution of a single root. It is then straightforward to obtain the full conditional posterior for the root up to a scaling factor. To improve the mixing properties of the chain we update the roots in random order. Draw $l = \bar{r} + \bar{c}$ integers with replacement from $1, \dots, \bar{r} + \bar{c}$ and update the roots in this order.

Sampling real roots

Assuming a real root ρ_i is selected for updating, define

$$U = \tilde{Y} - X_c \beta_c, \quad \tilde{Y} = (1 - L)^{d_c} Y,$$

the demeaned appropriately differenced series. Using (1.4), filter U to obtain the dependent variable Z and explanatory variable Z_{-1} for the AR(1). With a uniform prior for the continuous component in (1.8), the full conditional posterior is

$$\rho_i | Y, \xi_c \setminus \rho_i \propto \left(w_0^r I_{\rho=0} + w_s^r \frac{1}{2} + w_{-1}^r I_{\rho=-1} \right) \exp \left\{ -\frac{1}{2s^2} (\hat{r} - \rho_i)^2 \right\},$$

where

$$\hat{r} = (Z'_{-1} Z_{-1})^{-1} Z'_{-1} Z \quad s^2 = \sigma_\varepsilon^2 (Z'_{-1} Z_{-1})^{-1},$$

are the usual least squares quantities. Computing the scale factor, the full conditional posterior state probabilities for a null, stationary, and real negative unit root are given by

$$p_i = \frac{t_i}{\sum_{k=1}^3 t_k},$$

where

$$\begin{aligned} \rho_i = 0 : \quad t_1 &= w_0^r \exp \left\{ -\frac{1}{2s^2} (\hat{r} - 0)^2 \right\}, \\ |\rho_i| < 1 : \quad t_2 &= \frac{w_s^r}{2} \int_\rho \exp \left\{ -\frac{1}{2s^2} (\hat{r} - \rho)^2 \right\} d\rho = \frac{w_s^r \cdot \kappa}{2} (2\pi s^2)^{1/2} \\ \rho_i = -1 : \quad t_3 &= w_{-1}^r \exp \left\{ -\frac{1}{2s^2} (\hat{r} - (-1))^2 \right\}. \end{aligned}$$

In the expression for t_2 ,

$$\kappa = \Phi \left(\frac{1 - \hat{r}}{s} \right) - \Phi \left(\frac{-1 - \hat{r}}{s} \right)$$

is a scale factor due to the restriction $\rho_i \in (-1, 1)$. Sampling a new state, if the next state corresponds to a stationary root the actual value is sampled from

$$\rho \sim \mathcal{TN} \left(\rho | \hat{r} = (Z'_{-1} Z_{-1})^{-1} Z'_{-1} Z \quad s^2 = \sigma_\varepsilon^2 (Z'_{-1} Z_{-1})^{-1}, -1, 1 \right),$$

a normal distribution truncated to the stationary region.

Sampling complex pairs

Assuming a complex pair $m_c \exp(\pm i\omega_c)$ is selected for updating, define

$$U = \tilde{Y} - X_c \beta_c, \quad \tilde{Y} = (1 - L)^{d_c} Y,$$

the demeaned appropriately differenced series. Using (1.5), filter U to obtain the filtered sequence Z which is an AR(2) with non-linear parameter restrictions. While possible, a Gibbs step would be complicated and time consuming to execute. Instead, we use Metropolis-Hastings steps to move between the possible states for the complex roots.

Applying the reversible jump, move probabilities must be selected. As the suggested moves are tailored for transitions between particular root states, the jump probabilities will the very least depend on the current state. In particular, if the root selected for updating is at the origin, with probability j_{ns} attempt to introduce a stationary root using Move M1 and otherwise do not update the root. If the root selected for updating is instead stationary we may attempt Moves M1, M2, or M3, with probability j_{sn} , j_{su} , and $j_{ss} = 1 - j_{sn} - j_{su}$ respectively. In turn, this amounts to attempting to deactivate the root by placing it at the origin, moving it to the boundary and introducing a persistent cycle or seasonal component, or simply updating it. Finally, if the root selected is currently placed at the unit circle, an attempt to move it into the stationary region using Move M2 is performed with probability j_{us} . Otherwise, an attempt to update the frequency using Move M4 is made with probability $j_{uu} = 1 - j_{us}$.

Throughout we will, without loss of generality, assume that there is a single special frequency, $a^* = \cos \omega^*$.

Move M1 (Transitions between null and stationary roots)

Assume the complex pair is currently at the origin. To sample a proposal, from the filtered sequence obtain the dependent variable Z and the explanatory variables Z_{-1} and Z_{-2} for the AR(2). Calculate ordinary least squares quantities

$$W = [Z_{-1}, Z_{-2}], \quad \hat{\phi} = (W'W)^{-1} W'Z, \quad \Omega = \sigma_\varepsilon^2 (W'W)^{-1}.$$

Begin by sampling a proposal for the modulus by sampling ϕ_2 from

$$\phi_2 \sim \mathcal{TN}(\phi_2 | \hat{\phi}_2, \omega_{22}, -1, 0), \quad (1.11)$$

a normal distribution truncated with lower bound -1 and upper bound 0 . Set the proposed modulus to $m_p = \sqrt{-\phi_2}$. Proposing a frequency, we want to

propose either a special or arbitrary frequency. This is achieved by introducing a discrete distribution over all possible frequency states; in this case just two. While the probabilities may be fixed at any values, the mixing of the chain is improved if a more elaborate strategy is used. Calculate a proposal probability for the special frequency as

$$p_s = \frac{w_{s,\omega^*}^c \exp \left\{ -\frac{1}{2} \sigma_1^{-2} (\mu_1 - 2\sqrt{-\phi_2} a^*)^2 \right\}}{w_{s,\omega^*}^c \exp \left\{ -\frac{1}{2} \sigma_1^{-2} (\mu_1 - 2\sqrt{-\phi_2} a^*)^2 \right\} + w_{s,\omega}^c \frac{\kappa}{\bar{a}+1} (2\pi\sigma_1^2)^{1/2}}, \quad (1.12)$$

where

$$\begin{aligned} \mu_1 &= \hat{\phi}_1 + \omega_{12} \omega_{22}^{-1} (\phi_2 - \hat{\phi}_2), \\ \sigma_1^2 &= \omega_{11} - \omega_{12} \omega_{22}^{-1} \omega_{21}, \sigma_2^2 = \omega_{22} \\ \kappa &= \Phi \left(\frac{2\sqrt{-\phi_2} \bar{a} - \mu_1}{\sigma_1} \right) - \Phi \left(\frac{2\sqrt{-\phi_2} - \mu_1}{\sigma_1} \right). \end{aligned}$$

Then, with probability p_s set $a_p = a_s$ and with probability $1 - p_s$ sample ϕ_1 from

$$\phi_1 \sim \mathcal{TN}(\phi_1 | m_1, \sigma_1^2, -2\sqrt{-\phi_2}, 2\sqrt{-\phi_2} \bar{a}), \quad (1.13)$$

to propose either $(m_p, a_p) = (\sqrt{-\phi_2}, a_s)$ or $(\sqrt{-\phi_2}, \phi_1/2\sqrt{-\phi_2})$.

To calculate the relevant acceptance probability, and sample the next state, define the current and proposed root contribution to the lag polynomial and filter Z to get respective residuals,

$$\begin{aligned} E_c &= (1 - 2m_c a_c L + m_c^2 L^2) Z \\ E_p &= (1 - 2m_p a_p L + m_p^2 L^2) Z \end{aligned}$$

Then, if a stationary pair with arbitrary frequency is proposed, the acceptance probability is calculated with

$$\begin{aligned} \alpha_{M1} &= \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\varepsilon^{-2} (E_p' E_p - Z' Z) \right] \kappa \right\}, \\ \kappa &= \frac{w_{s,\omega}^c}{w_0^c} \frac{3}{4(\bar{a}+1)} \frac{j_{sn}}{j_{ns} (1-p_s) f_1(\phi_1) f_2(\phi_2)}. \end{aligned}$$

where f_1 and f_2 the truncated normal distributions in (1.11) and (1.13). If instead a special frequency is proposed the acceptance probability is

$$\begin{aligned} \alpha_{M1} &= \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\varepsilon^{-2} (E_p' E_p - Z' Z) \right] \kappa \right\}, \\ \kappa &= \frac{w_{s,\omega^*}^c}{w_0^c} \frac{j_{sn}}{j_{ns} p_s f_2(\phi_2)}. \end{aligned}$$

For the reciprocal move attempting to delete a stationary root, the proposal does not involve any sampling. Instead, the balancing move treats $\phi_1 = -m_c^2$ as if it had been sampled from (1.11), and depending on the current type of frequency $\phi_2 = 2m_c a_c$ from the appropriate part of (1.12)-(1.13). The correct acceptance probabilities are then simply the inverse of the relevant expression for α_{M1} after the obvious substitutions.

Move M2 (Transitions between the stationary roots and unit roots)

Assume the root is currently stationary. In an attempt to move the pair to the boundary of the stationary region the proposal is simply $(1, a_c)$ and leaves the frequency unchanged. It does not involve any sampling and is trivial to extend to an arbitrary number of special frequencies. When balancing, the current modulus is treated as if it had been sampled from the truncated normal in (1.11). Computing the appropriate residuals E_c and E_p , if the current root is stationary with an arbitrary frequency, the acceptance probability is

$$\alpha_{M2} = \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\epsilon^{-2} \left(E'_p E_p - E'_c E_c \right) \right] \kappa \right\},$$

$$\kappa = \frac{w_{1,\omega}^c}{w_{s,\omega}^c} \frac{2}{3m_c} \frac{j_{us} f_2(-m_c^2)}{j_{su}}.$$

If the root currently has a special frequency, the transition to a unit root at the special frequency is instead accepted with probability

$$\alpha_{M2} = \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\epsilon^{-2} \left(E'_p E_p - E'_c E_c \right) \right] \kappa \right\},$$

$$\kappa = \frac{w_{1,\omega^*}^c}{w_{s,\omega^*}^c} \frac{j_{us} f_2(-m_c^2)}{j_{su}}.$$

If instead the root is currently at the unit circle, a modulus is sampled from the truncated normal in (1.11). The balancing move involves no sampling. Depending on the status of the current frequency the acceptance probability is the inverse of the appropriate version of α_{M2} after substituting the current value of the modulus $m_c = 1$ with the proposed value $m_p = \sqrt{-\phi_2}$.

Move M3 (Updating stationary roots)

The first local move explores the stationary region and is only available to roots that are currently stationary. A proposal is generated using the same

procedure as in Move M1 and the balancing move treats the current value as if it had been sampled with the same procedure. The acceptance probability depends on the status of the current and proposed frequency. If both the current and proposed pairs are stationary with arbitrary frequency the acceptance probability is

$$\alpha_{M3} = \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\varepsilon^{-2} \left(E'_p E_p - E'_c E_c \right) \right] \kappa \right\},$$

$$\kappa = \frac{(1 - p_s^c) f_{1c} (2m_c a_c) f_2 (-m_c^2)}{(1 - p_s^p) f_{1p} (\phi_1) f_2 (\phi_2)}.$$

where p_s^c and p_s^p are the probability of a special frequency (1.12) for the current and proposed root respectively. If the current root has an arbitrary frequency and a special frequency is proposed, the acceptance probability is

$$\alpha_{M3} = \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\varepsilon^{-2} \left(E'_p E_p - E'_c E_c \right) \right] \kappa \right\},$$

$$\kappa = \frac{w_{s,\omega}^c 4 (1 - p_s^c) f_{1c} (2m_c a_c) f_2 (-m_c^2)}{w_{s,\omega}^c 3 (\bar{a} + 1) p_s^p f_2 (\phi_2)}.$$

If both the current and proposed pairs are stationary with special frequency the acceptance probability is just

$$\alpha_{M3} = \min \left\{ 1, \exp \left[-\frac{1}{2} \sigma_\varepsilon^{-2} \left(E'_p E_p - E'_c E_c \right) \right] \kappa \right\},$$

$$\kappa = \frac{w_{s,\omega}^c}{w_{s,\omega}^c} \frac{p_s^c f_2 (-m_c^2)}{p_s^p f_2 (\phi_2)}.$$

Move M4 (Updating the frequency of unit roots)

The second local move explores the frequencies of unit roots. Keeping the modulus constant, a new frequency can be sampled with a Gibbs step, similar to that used for updating real roots. This move is executed using (1.12) and (1.13) with $\phi_2 = -1$ and setting $a_p = \phi_1/2$.

1.4.4 Updating the innovation variance

With the selected model structure, the treatment of the innovation variance may be performed in a standard fashion. Having completed a full cycle of model selection, and updating of trend and lag polynomials compute the observed residuals,

$$E = \Psi^c(L) \left(\tilde{Y} - X_c \beta_c \right), \quad \tilde{Y} = (1 - L)^{d_c} Y.$$

Assuming normally distributed errors, the conditional likelihood of σ_ε^2 has an inverse gamma kernel. With the conjugated inverse gamma prior on the precision this leads to

$$\sigma_\varepsilon^{-2} | Y, \xi_c \setminus \sigma_\varepsilon^2 \sim \mathcal{Gamma}(\sigma_\varepsilon^2 | \alpha_0 + T/2, \beta_0 + E'E/2),$$

where T the number of observations.

1.4.5 Sampling latent initial values

The suggested approach circumvents the initial value problem by simply modelling the initial values as latent variables. As we may for the purpose of updating them condition on all else, three simple steps will generate the required initial values.

1. Conditional on the differencing order, the trend and the lag polynomial, define

$$v_t = (1 - L)^{d_c} y_t - \tau'_t \beta_c,$$

for $t = d_c + 1, \dots, T$

2. Using $\Psi^c(L) v_s = \varepsilon_s$ and reversing the time arrow, back-cast the required number of initial values v_s^0 for $s = d_c, d_c - 1, \dots, -\bar{r} - 2\bar{c} - \bar{d} + d_c + 1$ iteratively while sampling an innovation, conditional on the variance. Note how this is possible even when the polynomial includes unit roots, see Huerta and West (1999a) for a proof.
3. Extract initial values on the original scale using

$$(1 - F)^{d_c} y_s = (-1)^{d_c} (\tau'_{s+d_c} \beta_c + v_{s+d_c}^0)$$

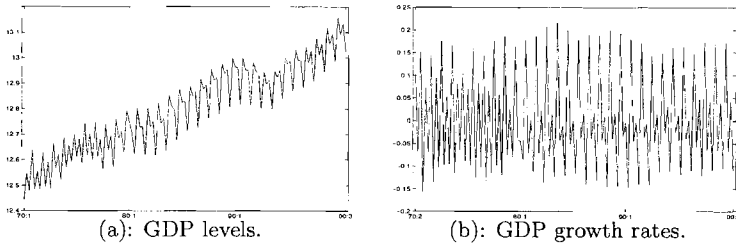
for $s = 0, \dots, -\bar{r} - 2\bar{c} - \bar{d} + 1$.

1.5 The Swedish GDP

Figure 1.1 shows the log of Swedish real GDP, in levels as well as after taking a first difference. The quarterly data span the period from the first quarter in 1970 to the third quarter in 2000. The data is dominated by two features, a strong seasonal pattern which appears to be stable from the beginning of the eighties and a fall in GDP in the early nineties.

We specify the prior parameters as follows. For the trend component we set the maximum differencing order to $\bar{d} = 2$ and the maximum degree of the

Figure 1.1 The log of Swedish real GDP, quarterly from 1970:1 to 2000:3; levels (left) and differenced (right).



trend polynomial to $\bar{g} = 3$, i.e. a quadratic trend, with a uniform prior over all possible combinations. The scale factor in the g -prior for the parameters in the deterministic trend is set to $g = 1/T$, i.e. the prior information corresponds to the average information in one observation. We allow for a maximum of $\bar{r} = 8$ real roots and use the proposed default choice for the weights in the mixture prior, that is $w_0^r = 1/2$ and $w_s^r = w_{-1}^r = 1/4$. The expected number of active real roots is four with two stationary roots and and two negative unit roots. The uniform prior for stationary roots is truncated above at 0.99, that is $\delta = 0.01$. For the complex roots we set $\bar{c} = 4$ so the real and complex roots can contribute in equal amounts to the effective lag length. The weights in the mixture prior are set to $w_0^c = 1/2$ $w_{s,\omega}^c = w_{s,\omega}^c = w_{1,\omega}^c = w_{1,\omega}^c = 1/8$. That is, we expect two complex roots with equal probability of these stationary or on the unit circle and have a seasonal or arbitrary frequency. The precision hyperparameters are calibrated using results based on available data from 1950 to 1970.

We obtain the posterior distribution from one run of the Markov chain with 250 000 replicates after having discarded 5000 replicates as burn in. The chain was monitored for convergence by studying the sample path of the model indexes, running means and the Geweke (1992) z -statistic for the constant term and error variance. To reduce autocorrelation and simplify postprocessing the chain was thinned by only retaining every fifth replicate.

Simple univariate measures of the posterior probabilities of different model configurations are shown in Table 1.1. It is evident that the dominant model is one with a constant term, one positive and one negative unit root and one seasonal unit root. The seasonal difference $(1 - L^4)$ that might be expected from viewing the data is thus completed. With probability 0.24 a second complex root, a stationary root with arbitrary or seasonal frequency or possibly

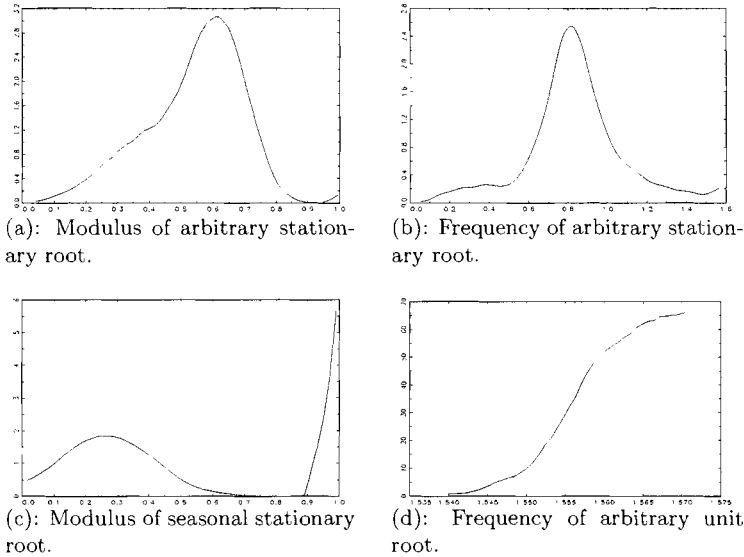
Table 1.1 Univariate posterior probabilities for model configurations

| Number of roots | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Real roots | 0.000 | 0.001 | 0.032 | 0.526 | 0.307 | 0.111 | 0.021 |
| Stationary real | 0.001 | 0.032 | 0.520 | 0.310 | 0.113 | 0.022 | 0.002 |
| Neg. unit root | 0.010 | 0.990 | | | | | |
| Complex roots | 0.000 | 0.723 | 0.242 | 0.033 | 0.002 | | |
| Arbitrary stationary | 0.788 | 0.194 | 0.018 | 0.000 | | | |
| Seasonal stationary | 0.883 | 0.111 | 0.006 | 0.000 | | | |
| Unit arbitrary | 0.986 | 0.014 | 0.000 | 0.000 | | | |
| Unit seasonal | 0.054 | 0.946 | 0.000 | 0.000 | | | |
| Stochastic and deterministic trends | | | | | | | |
| Unit roots | 0.000 | 1.000 | | | | | |
| Degree of trend polynomial | 0.183 | 0.766 | 0.046 | 0.004 | 0.001 | | |

a unit root with arbitrary frequency, is present in the model. The posterior distribution of these roots are displayed in Figure 1.2. It is clear that they all, to some extent mimic a seasonal unit root. This is, in particular the case for the unit root at arbitrary frequency and the stationary seasonal root. The arbitrary unit root only appears when the seasonal unit root is absent and all the probability mass above a modulus of 0.9 for the stationary seasonal root correspond to replicates when the seasonal unit root is absent. The frequency for the arbitrary stationary root has a mode close to $\pi/4$ and this root appear to be aliasing for the seasonal frequency $\pi/2$. The posterior probability of a third complex root is only 0.033 and there is no evidence of a business cycle.

Turning to the real roots, the posterior distribution of the largest negative root, the smallest root and the largest positive root are displayed in Figure 1.3. For the negative root we, again, see a tendency for it to proxy for the negative unit root when this is absent from the model. The posterior for the positive real root is well behaved does not show any adverse affects of the truncation at 0.99 enforced by the prior distribution. The posterior for the smallest root is a mixture over models with varying number of real roots and the posterior mode close to zero correspond to models with a large number of real roots.

Figure 1.3 also shows the posterior distribution of the constant term in models with only a constant. The posterior mean is 0.005 corresponding to a quarterly growth rate of 0.5%. The posterior for the constant is more dispersed in models with a linear trend and has a slightly lower mean of 0.0039. The posterior mean for the trend coefficient is 0.00006 with a posterior variance of

Figure 1.2 Posterior distribution of complex roots.**Table 1.2** Posterior distribution of coefficients in $\Psi(L)$

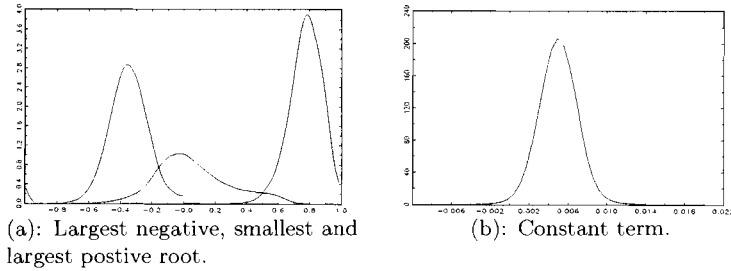
| | ψ_1 | ψ_2 | ψ_3 | ψ_4 | ψ_5 | ψ_6 | p |
|----------------|----------|----------|----------|----------|----------|----------|-------|
| $r = 3, c = 1$ | -0.549 | -0.277 | -0.278 | 0.721 | 0.272 | | 0.408 |
| $r = 4, c = 1$ | -0.556 | -0.272 | -0.286 | 0.713 | 0.270 | -0.014 | 0.217 |
| $r = 3, c = 2$ | -0.559 | -0.295 | -0.205 | 0.736 | 0.296 | 0.033 | 0.105 |
| Model average | -0.553 | -0.278 | -0.261 | 0.718 | 0.272 | -0.002 | |

6.8×10^{-9} , effectively removing the trend from the model.

The Markov chain explores a large range of lag length configurations as is evident from the posterior distribution for the number of real and complex roots in Table 1.1. Still the posterior distribution of the lag lengths is fairly concentrated with 90% of the posterior mass concentrated on lags 5 to 8 with posterior probabilities 0.408, 0.231, 0.178 and 0.097. In addition to this the unit root in the trend component is always present leading to effective lag lengths of 6 to 9.

Finally, Table 1.2 show the posterior mean of the coefficients in the stationary polynomial $\Psi(L)$ for the combinations of number of real and complex roots with the highest posterior probabilities. The posterior means for the

Figure 1.3 Posterior distribution of real roots and constant term.



first five lags are relatively insensitive to the root configuration. It might thus be difficult to infer the root configuration from the AR-coefficients if these are estimated directly. The coefficients for the higher lags are quite small, this – together with the small posterior probabilities for higher lag orders – is reflected in the posterior mean for the model averaged polynomial.

1.6 Final remarks

Extending the work of Huerta and West this paper demonstrates how the fundamental features of autoregressive processes, the roots of the characteristic polynomial, can be modelled directly. Focusing on the roots we can extract information about the dynamics which is otherwise not readily available.

The modelling framework is highly flexible and automatically robustifies against misspecification by implicitly averaging over different lag lengths, number of unit roots and specifications for the deterministic trend. This is accomplished at the same time as inference on these issues is straightforward, the posterior probability of any given model specification can be estimated directly from the output of the Markov chain.

The flexibility comes at a cost, mainly in terms of the specification of the prior distribution. The choice of weights for the mixture prior is not always straightforward. The default choices discussed in section 1.3.2 are useful starting points but can lead to counter intuitive results and should not be adopted automatically. The restriction of the prior on stationary real roots to a subset of the stationary region might also be unpalatable in some contexts. We plan to address both the specification of the prior on the components of the polynomial and alternative ways of resolving the near non-identification of the constant term in future research.

Bibliography

- DeJong, D. N. and Whiteman, C. H. (1991*a*), 'Reconsidering 'trends and random walks in macroeconomic time series'', *Journal of Monetary Economics* **28**, 221–254.
- Dejong, D. N. and Whiteman, C. H. (1991*b*), 'The temporal stability of dividends and stock prices: Evidence from the likelihood function', *American Economic Review* **81**, 600–617.
- Franses, P. H., Hoek, H. and Paap, R. (1997), 'Bayesian analysis of seasonal unit roots and seasonal mean shifts', *Journal of econometrics* **78**, 359–380.
- Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, Oxford.
- Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* pp. 711–732.
- Huerta, G. and West, M. (1999*a*), 'Bayesian inference on periodicities and component spectral structures in time series', *Journal of Time Series Analysis* pp. 401–416.
- Huerta, G. and West, M. (1999*b*), 'Priors and component structures in autoregressive time series models', *Journal of the Royal Statistical Society. Series B (Methodological)* .
- Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990), 'Seasonal integration and cointegration', *Journal of Econometrics* **44**, 215–238.
- Lubrano, M. (1995), 'Testing for unit roots in a bayesian framework', *Journal of Econometrics* **69**, 479–507.

- Schotman, P. C. (1994), 'Priors for the AR(1) model: Parameterization issues and time series considerations', *Econometric Theory* **10**, 579–595.
- Schotman, P. C. and van Dijk, H. K. (1991), 'A bayesian analysis of the unit root hypothesis in real exchange rates', *Journal of Econometrics* **49**, 195–238.

Paper 2

Bayesian Panel Unit Root Analysis

2.1 Introduction

The idea of exploiting cross-section variation has in the context of unit root inference been proposed by, among others, Quah (1994), Levin et al. (2002), and Im et al. (2003). In general, tests are implemented under the specification of some cross-country homogeneity in dynamics, such as lag structures and other model specification particulars, though the IPS procedure allows for complete heterogeneity between individuals. While the null hypothesis in all the above tests is that each series contain a unit root, that is the panel is difference stationary, the alternative differs. Earlier approaches of Quah (1994) and Levin et al. (2002) have all series being stationary as the alternative. As this implicitly presupposes that the individual series of a panel are homogenous with respect to the presence or absence of a unit root, great care must be exercised when interpreting results as the tests may be skewed towards rejection. Rejecting the unit root null in a panel with a mix of stationary and difference stationary series does not imply joint stationarity of the series. This problem is partly recognized and addressed in Im et al. (2003) where the alternative is stated as at least one of the individual series being stationary. Though this acknowledges the potentially mixed nature of the panel, the problem is not solved; tests may suffer from high power when a small fraction of the series is stationary and conversely lack power when a large fraction is stationary, see Karlsson and Löthgren (2000).

One possible explanation to the difficulties above is that we are generally interested in the composite question whether the panel can be modeled jointly as stationary or difference stationary and if not so what fraction of the series can? As none of the above tests are designed for that purpose it should not come as a surprise when they fail. Further, as argued by O'Connell (1998), the assumption of cross-sectional independence is often unrealistic and is one possible source of contrary results. The idea of common shocks giving rise to cross sectional dependency is well represented in the econometric literature on panel data analysis. But with respect to panel unit root inference the issue of cross-section correlation has not been completely resolved. Attempting to come to terms with the issues of panel composition, model uncertainty and correlation, we propose a multivariate extension of the model and methods in Huerta and West (1999a) and Huerta and West (1999b), further developed in the first paper of this thesis.

The approach is Bayesian and relies on a mixed component prior for the reciprocal of the roots to the characteristic polynomial. In section 2.2 we discuss how we may use the model for multivariate autoregressive models with correlated innovations after only some slight modifications. Section 2.3

discuss prior specification and section 2.4 give the tools for analyzing the posterior distribution. In section 2.5 the procedure is used to investigate the support in the data for purchasing power parity between major industrialized countries and section 2.6 concludes.

2.2 The model

2.2.1 A panel of time series

For a cross section of N time series observed over T time periods we consider the basic model

$$y_{it} = \mu_{it} + \sum_{j=1}^{p_i} \phi_{ij} y_{i,t-i} + \varepsilon_{it} \quad (2.1)$$

with $V(\varepsilon_t) = \Sigma$. That is, we allow for considerable heterogeneity with cross section specific parameters and lag lengths as well as for correlation between the time series. The set up is thus similar to that used in Im et al. (2003) with the addition of a general covariance structure as in O'Connell (1998). We depart from this, by now, fairly standard framework for panel unit root analysis in two important directions. Firstly, instead of treating the problem as a linear regression we focus directly on the time series properties. We do this by modelling the roots of the autoregressive polynomials directly as in Huerta and West (1999a), Huerta and West (1999b) and the previous paper in this thesis. Secondly, we adopt a Bayesian viewpoint on both practical and philosophical grounds. This brings with it two advantages compared to any frequentist approach. We obtain exact small sample results and simulating the posterior distribution using MCMC methods is considerably more straightforward than solving the highly non-linear maximization problems that results from considering maximum likelihood or other frequentist estimation methods.

Modelling the roots we rewrite the autoregressive polynomials

$$\phi_i(L) = 1 - \sum_{j=1}^{p_i} \phi_{ij} = \prod_{j=1}^{\bar{r}} (1 - \rho_{ij}L) \prod_{j=1}^{\bar{c}} (1 - 2m_{ij}a_{ij}L + m_{ij}^2L^2) \quad (2.2)$$

in terms of the real roots of the characteristic polynomial, ρ_{ij} , and the complex roots $m_{ij} \exp(\pm i\omega_{ij})$ with $a_{ij} = \cos \omega_{ij}$. The representation (2.2) immediately leads to the possibility of identifying interesting root configurations. In the current context positive unit roots, $\rho_{ij} = 1$, are of primary interest. In addition we will also pay attention to negative unit roots, $\rho_{ij} = -1$ and, crucially, null roots, $\rho_{ij} = 0$. It is the device of allowing individual roots to take the

value zero that allow the lag lengths to vary up to a prespecified maximum of $\bar{p} = \bar{r} + 2\bar{c}$. In addition, real roots may be exchanged for complex roots while maintaining the overall lag order by setting two real roots to zero and activating one complex pair. Completing the possibilities for the real roots we may also have arbitrary stationary roots, $-1 < \rho_{ij} < 1$.

Similar considerations for the complex roots leads to null roots, $m_{ij} = 0$, roots on the unit circle, $m_{ij} = 1$ and unrestricted stationary complex roots. For seasonal data we have, in addition, for each seasonal frequency defined $\omega_{ij} \in \{2k\pi/S, k = 1, \dots, S/2 - 1\}$ a set of stationary, $0 < m_{ij} < 1$, and unit roots, $m_{ij} = 1$.

2.2.2 Conditional likelihoods

While the model is highly non-linear in the characteristic roots, the contribution of a single roots – holding all other parameters constant – is within a simple AR(1) or AR(2) model. Filtering out all but one of the real roots,

$$z_{it} = \prod_{j \neq i} (1 - \rho_{ij}L) \prod_{j=1}^{\bar{c}} (1 - 2m_{ij}a_{ij}L + m_{ij}^2L^2) y_{it}, \quad (2.3)$$

we obtain the simple AR(1) model

$$(1 - \rho_{ij}L) z_{it} = \varepsilon_{it}. \quad (2.4)$$

Similarly for a complex pair, $m_{ij} \exp(\pm i\omega_{ij})$, we can filter out all other roots,

$$z_{it} = \prod_{j=1}^{\bar{r}} (1 - \rho_{ij}L) \prod_{j \neq i} (1 - 2m_{ij}a_{ij}L + m_{ij}^2L^2) y_{it} \quad (2.5)$$

to obtain an AR(2) model,

$$(1 - \phi_1L - \phi_2L^2) z_{it} = \varepsilon_{it}, \quad (2.6)$$

subject to the non-linear restrictions,

$$\phi_1 = 2m_{ij}a_{ij} = 2m_{ij} \cos \omega_{ij}, \quad \phi_2 = -m_{ij}^2.$$

With a normality assumption, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma)$, the simple conditional structure leads directly to a MCMC algorithm for exploring the posterior distribution. See Huerta and West (1999b) and the previous paper in this thesis for the univariate case.

To simplify the discussion of the multivariate case we will take the deterministic trend component, μ_{it} , to be zero. Allowing for a general deterministic trend is straightforward but would distract from the main points of interest. In addition, in the application to purchasing power parity (PPP) considered here the data is not trending and we remove the constant by subtracting the mean from each time series.

For the vector of innovations at time t , ε_t , the conditional distributions follow immediately from standard normal theory

$$\begin{aligned}\varepsilon_{1,t}|\varepsilon_{-1,t} &\sim \mathcal{N}(\mu_{1,t}, \sigma_1^2) \\ \mu_{1,t} &= \Sigma_{12}\Sigma_{22}^{-1}\varepsilon_{-1,t} \\ \sigma_1^2 &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\end{aligned}\tag{2.7}$$

where $\varepsilon_{-1,t}$ is ε_t with ε_{1t} deleted and Σ is partitioned conformably. Note how this conditioning is meaningful since $\varepsilon_{it} = \phi_i(L)y_{it}$ is observable conditional on the collection of roots, θ_i , of the polynomial. Using (2.3) and (2.4) we can thus write, conditional on Σ and all roots except the real root ρ_{ij} ,

$$\begin{aligned}(1 - \rho_{ij}L)z_{it} &= \varepsilon_{it} = \mu_{it} + \eta_{it} \\ \tilde{z}_{it} &= \rho_{ij}z_{i,t-1} + \eta_{it}\end{aligned}\tag{2.8}$$

for $\tilde{z}_{it} = z_{it} - \mu_{it}$ and $\eta_{it} \sim \mathcal{N}(0, \sigma_i^2)$. That is the root ρ_{ij} is, conditionally, the parameter of a linear regression with normal error and known variance.

Similarly, for a complex root we have, conditional on all but the root $m_{ij} \exp(\pm i\omega_{ij})$ and using (2.5) and (2.6),

$$\begin{aligned}(1 - \phi_1L - \phi_2L^2)z_{it} &= \varepsilon_{it} = \mu_{it} + \eta_{it} \\ \tilde{z}_{it} &= \phi_1z_{i,t-1} + \phi_2z_{i,t-2} + \eta_{it}.\end{aligned}\tag{2.9}$$

A final point is our treatment of initial values. Instead of conditioning on the first observations, $y_{i1}, \dots, y_{i\bar{p}}$ we use the full likelihood for y_{i1} to y_{iT} . Formally this requires marginalizing out the unobserved $y_{i,1-\bar{p}}, \dots, y_{i0}$ which complicates the form of the likelihood considerably. Instead we use numerical methods for this by treating the starting values as latent variables and simulate them together with the parameters in the Markov chain. This allow us to use the simpler likelihood conditional on the simulated $y_{i,1-\bar{p}}, \dots, y_{i0}$ and the marginalization is achieved by averaging over the output from the Markov chain.

2.3 Prior structure

2.3.1 A natural hierarchical structure

While the basic model (2.1) allows for considerable cross sectional heterogeneity, the motivation for a panel approach is the prior belief that there are some properties that are common to the time series. The belief that the time series are similar is naturally expressed using a hierarchical prior for the roots, θ_i ,

$$\begin{aligned}\theta_i &\sim \pi(\theta_i|w) \\ w &\sim \pi(w).\end{aligned}$$

That is, the roots for the different time series have the same prior distribution, reflecting similarity, but there is also prior uncertainty about the parameters of the prior, reflecting uncertainty about the exact nature of the similarity. A case in point is the issue of unit roots, w could then represent the common prior probability that a time series contains a unit root and $\pi(w)$ represents prior uncertainty about this probability. Note how the hierarchical nature of the prior neatly separates the issue of degree of similarity, governed by the tightness of $\pi(\theta_i|w)$, from the issue of the nature of the similarity and the uncertainty about this which is embodied in $\pi(w)$. Another important feature of the hierarchical prior is borrowing of strength, to the extent that the time series actually are similar the structure allows for pooling of information about θ across time series through the posterior for w .

Writing the prior as

$$\pi(\theta, \Sigma, w) = \pi(\Sigma) \pi(w) \prod_{i=1}^N \pi(\theta_i|w)$$

we proceed by considering the different parts of the prior. The most care has to be exercised in specifying $\pi(\theta_i|w)$ since this should reflect both our prior notions about the prevalence of the different root states discussed in section 2.2 and, for a convenient MCMC algorithm, conform to the conditional structure for the likelihood outlined in section 2.2.2.

The latter requirement is naturally fulfilled by considering the implications of the individual roots being unidentified in the sense that the lag polynomial is invariant to permutations of the roots. The roots are, in other words exchangeable, and we are lead to specifying identical and independent priors for the real and complex roots respectively. That is,

$$\pi(\theta_i|w) = \prod_{j=1}^{\bar{r}} \pi(\rho_{ij}|w) \prod_{j=1}^{\bar{c}} \pi(m_{ij}, \omega_{ij}|w).$$

Another important feature of the prior on the roots is that we must assign point mass priors to states of interest that correspond to a single point in the parameter space, e.g. unit roots, in order to obtain posterior probabilities for the presence of the state. The priors will thus be mixtures over the different states for the roots discussed in section 2.2.

2.3.2 Real roots

The status of a real root may be characterized as either null or non-null. If non-null we can classify it according to the position relative to the unit circle as either a stationary root, a positive unit root, or a negative unit root. The four possible states are made operative in the prior by using point masses for the three singular states and a continuous component, properly restricted, for the stationary state. Dropping the complicated root index, the mixture prior is

$$\pi(\rho|w) = w_0^r I_{\rho=0} + w_1^r I_{\rho=1} + w_{-1}^r I_{\rho=-1} + w_s^r f_{|\rho|<1}(\rho) \quad (2.10)$$

where $\sum w_i^r = 1$, $I_{\rho=k}$ the indicator function, and $f_{|\rho|<1}(\rho)$ a uniform distribution on the stationarity region $(-1, 1)$. This is exactly the prior suggested in Huerta and West (1999b). This prior differs from the prior on the real roots used in the previous paper of this thesis where the positive unit root is handled together with the deterministic trend.

For the weights w^r , non-zero probabilities w_1^r and w_{-1}^r allow inference about unit roots. A non-zero probability w_0^r at the origin caters for model uncertainty, allowing the contribution of real roots to the total order to fall below the stipulated maximum. It is important to note how this prior induces complicated priors on the parameters in the time domain, as well as on the total autoregressive order and the existence of unit roots.

2.3.3 Complex conjugate pairs

For the complex roots we use our extended prior to allow for inference on the seasonal components. Adopting a similar structure for complex roots, a complex pair is either at the origin or not. If not, it may have a stationary modulus or a modulus at the boundary of the stationary region. Regardless of modulus type, the frequency may be arbitrary or at a seasonal frequency where an arbitrary frequency is taken to model general cyclical components. Again, this structure can be made operational with a component mixture prior. The number of components depend on the presence and number of seasonal frequencies. Thus, for quarterly data with an annual season we have a single seasonal frequency while for monthly data we get an ensemble of five seasonal

frequencies. With quarterly data, the seasonal frequency is at $\omega^* = \pi/2$ and we have five components

$$\begin{aligned} \pi(m, \omega) = & w_0^c I_{m=0} + w_{s, \omega^*}^c f_{m, \omega=\omega^*} + w_{s, \omega}^c f_{m, \omega} + \\ & + w_{1, \omega^*}^c I_{m=1, \omega=\omega^*} + w_{1, \omega}^c f_{m=1, \omega} \end{aligned} \quad (2.11)$$

where the indices on m and ω have been dropped for notational convenience. The indicator functions $I_{m=0}$ and $I_{m=1, \omega=\omega^*}$ represent complex conjugate root pairs located at the origin and the boundary at frequency ω^* respectively. The continuous distributions $f_{m, \omega=\omega^*}$, $f_{m, \omega}$ and $f_{m=1, \omega}$, correspond to a stationary root at the seasonal frequency, a stationary root at an arbitrary frequency and a unit root at an arbitrary frequency. As noted above, the prior generalizes to an arbitrary collection of special frequencies by the simple inclusion of the appropriate terms on the form $w_{1, \omega_k^*}^c I_{m=1, \omega=\omega_k^*}$, and $w_{s, \omega_k^*}^c f_{m, \omega=\omega_k^*}$. A non-zero probability w_0^c plays a similar role to w_0^r for the real roots.

The continuous part, $f_{m, \omega}$, and indirectly $f_{m=1, \omega}$ and $f_{m, \omega=\omega^*}$, need to be selected to complete the specification of the prior for complex roots. As before we use a Beta prior for the modulus and a uniform prior for $\cos \omega$. For $f_{m, \omega}$, we have

$$m \sim Be(3, 1), a \equiv \cos \omega \sim U(-1, \bar{a}).$$

with prior independence between m and a where $\bar{a} = \cos \underline{\omega}$ and $\underline{\omega}$ is the minimum observable frequency. For the seasonal frequencies, fixing ω at ω^* we have $f_{m, \omega=\omega^*} \sim Beta(2, 1)$. Finally, for unit roots with arbitrary frequency, the same uniform prior $a \equiv \cos \omega \sim U(-1, \bar{a})$ is used for the frequency.

2.3.4 Weights

The weights w^r and w^c play a crucial role in the prior specification. In particular, if we condition on the weights, the induced priors on the number of roots in a particular state are multinomial or marginal binomial distributions. That is, the number of non-null real roots, r , is distributed as $Bin(\bar{r}, 1 - w_0^r)$ and the number of non-null complex roots, c , as $Bin(\bar{c}, 1 - w_0^c)$. The implied prior on the effective lag length is thus the distribution of $r + 2c$. While it is, in general, desirable to specify relatively large values for \bar{r} and \bar{c} this needs to be counteracted by not choosing too small values for w_0^r and w_0^c in order to avoid putting too much prior probability on excessively long lag lengths.

A similar logic applies to the other states, the implied prior for the number of positive unit roots in a single time series is $Bin(\bar{r}, w_1^r)$. In most economic time series we would expect at most two unit roots and with a large value

for \bar{r} we need a small value for w_1^r in order to avoid putting excessive prior probabilities on a large number of unit roots.

Dirichlet distributions are a natural choice for parameterizing the prior for w^r and w^c and we use

$$\pi(w) = \text{Dir}(w^r | k_r \delta^r) \text{Dir}(w^c | k_c \delta^c)$$

where δ^r and δ^c are positive vectors with elements summing to one and k_r and k_c are scale factors determining how informative the prior is. That is, $E(w_i^r) = \delta_i^r$, $V(w_i^r) = \delta_i^r(1 - \delta_i^r) / (1 + k_r)$ and similarly for the complex weights. The implied marginal prior for the number of roots in a particular state is of the binomial-dirichlet form with δ playing a similar role to the weights in the examples above.

For guidelines on default choices for the weights when they are specified directly as prior parameters see Paper 1. Adapting these guidelines to the choice of δ suggests the following default choices that can be used in absence of other prior information. Reflecting the two stage nature of the root states where a root is either null or non-null and is in a particular state if non-null, a natural way to structure this is to take

$$\delta_0^r = p_0, \delta_s^r = (1 - p_0) p_s, \delta_1^r = (1 - p_0) p_1, \delta_{-1}^r = (1 - p_0) (1 - p_s - p_1)$$

for the real roots with $p_0 = 1/2$ and $p_s = p_1 = 1/3$. Similar reasoning applies to the complex roots with the added feature that we now have to consider combinations of stationary modulus and modulus on the unit circle with arbitrary frequencies and seasonal frequencies. A reasonable structure is then

$$\begin{aligned} \delta_0^c &= p_0, \delta_{s,\omega^*}^c = (1 - p_0) p_s p_{\omega^*}, \delta_{s,\omega}^c = (1 - p_0) p_s (1 - p_{\omega^*}) \\ \delta_{1,\omega^*}^c &= (1 - p_0) (1 - p_s) p_{\omega^*}, \delta_{1,\omega}^c = (1 - p_0) (1 - p_s) (1 - p_{\omega^*}) \end{aligned}$$

with $p_0 = p_s = 1/2$ and $p_{\omega^*} = 1/(n^* + 1)$ where n^* is the number of special frequencies.

Finally k_r and k_c can be thought of as measuring how many observations the prior is worth while keeping in mind that the number of "observations" on the state of the real and complex roots is $N\bar{r}$ and $N\bar{c}$, respectively. It turns out that k_r and k_c are important tuning parameters for the Markov chain, not selecting them to small in relation to $N\bar{r}$ and $N\bar{c}$ will decrease the probability of introducing near trapping states.

2.3.5 Covariance matrix prior

To complete the model a prior for the covariance matrix Σ must be selected. A convenient conjugate choice is,

$$\pi(\Sigma) = \mathcal{W}^{-1}(\Sigma|B, d),$$

that is an inverse Wishart, with scale matrix B and d degrees of freedom.

2.4 Posterior Analysis

The structure of the likelihood and the prior leads to a posterior where full conditional posteriors are available for all parameters. The Markov chain is thus essentially implemented as a Gibbs sampler where the parameters of each root are updated one at a time. The matter is, however, complicated by the need to explore the different possible root states with associated changes in the dimension of the parameter space. For the complex roots this leads us to embed reversible jump Metropolis-Hastings steps (see Green (1995)) within the Gibbs sampler.

The overall structure of the Markov chain is as follows:

1. For each time series $i = 1, \dots, N$, in turn
 - (a) Calculate the residuals for the other time series, $1, \dots, i-1, i+1, \dots, N$ and obtain the parameters, μ_{it} and σ_i^2 of the conditional distribution of ε_{it} as in (2.7)
 - (b) For $j = 1, \dots, \bar{r}$ update the real roots ρ_{ij} using the algorithm in section 2.4.1
 - (c) For $j = 1, \dots, \bar{c}$ update the complex roots $(m_{ij}, \exp(\pm i\omega_{ij}))$ using the algorithm in section 2.4.2
2. Update the weights as in section 2.4.3.
3. Update the variance covariance matrix Σ using the algorithm in section 2.4.4.
4. Sample the latent initial values y_{it} , for $t = 0, \dots, 1 - \bar{r} - 2\bar{c}$, $i = 1, \dots, N$ as outlined in section 2.4.5.

The strategy of updating one root at a time may have a detrimental effect on the mixing of the chain and lead to slow convergence. On the other hand it

leads to simple computations and we expect this to make up for the possible slow convergence.

With respect to unit root inference, each iteration produces vectors of indices for real and complex unit roots. These vectors may be used to estimate the probability of any series having a unit root of some or any kind as well as estimate a probability distribution for the number of unit root series in the panel. We can then use these estimates for qualitative or quantitative inference on panel composition or individual series properties.

2.4.1 Real roots

For the real roots we have a simple Gibbs step which mixes over the possible states for the root. When updating ρ_{ij} , first filter out the remaining roots in the polynomial using (2.3). Then let \tilde{Z} and Z_{-1} be the dependent and explanatory variables in the regression (2.8) for ρ_{ij} and

$$\hat{r}_{ij} = (Z'_{-1}Z_{-1})^{-1} Z'_{-1}\tilde{Z}, \quad s_{ij}^2 = \sigma_i^2 (Z'_{-1}Z_{-1})^{-1} \quad (2.12)$$

the least squares estimates. The full conditional posterior is then

$$p(\rho_{ij}|\cdot) = \frac{\pi(\rho_{ij}) \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - \rho_{ij})^2\right\}}{\int_{\rho_{i,j}} \pi(\rho_{ij}) \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - \rho_{ij})^2\right\}}, \quad (2.13)$$

where the conditioning is on all quantities except ρ_{ij} and $\pi(\rho_{ij})$ the mixture prior as defined in (2.10).

The conditional posterior probabilities of the root states are

$$p_i = \frac{t_i}{\sum_{k=1}^4 t_k},$$

where

$$\begin{aligned} \rho_{ij} = 0 : \quad t_1 &= w_0^r \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - 0)^2\right\}, \\ |\rho_{ij}| < 1 : \quad t_2 &= \frac{w_s^r}{2} \int_{-1}^1 \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - \rho)^2\right\} d\rho = \frac{w_s^r \kappa}{2} \left(2\pi s_{ij}^2\right)^{1/2} \\ \rho_{ij} = -1 : \quad t_3 &= w_{-1}^r \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - (-1))^2\right\}, \\ \rho_{ij} = 1 : \quad t_4 &= w_1^r \exp\left\{-\frac{1}{2s_{ij}^2} (\hat{r}_{ij} - 1)^2\right\} \end{aligned}$$

and $\kappa = \Phi[(1 - \hat{r}_{ij})/s_{ij}] - \Phi[(-1 - \hat{r}_{ij})/s_{ij}]$. A state for the root is drawn with probabilities p_i . If the continuous state is drawn a value for ρ_{ij} is drawn from a normal distribution with mean \hat{r}_{ij} and variance s_{ij}^2 truncated to the stationary region.

2.4.2 Complex conjugate pairs

While a Gibbs step along the lines of the updating for the real roots is possible in principle the computations would be expensive due to the non-linear restrictions on the parameters in the conditional regression (2.9). Instead a reversible jump algorithm is used. This requires designing moves between the states that will allow us to explore all possible states. Employing the same algorithm as in Paper 1 we use four move types. M1 moves between null roots and stationary roots, M2 moves between stationary roots and roots on the unit circle, M3 updates the frequency and modulus of stationary roots and M4 updates the frequency of roots on the unit circle. Note that moves M3 and M4 mixes over both arbitrary frequencies and the seasonal frequencies. See Paper 1 for details, the only modification needed for the current case is to use the dependent and explanatory variables from (2.9). In order to implement the algorithm we need to select move probabilities. If the root is currently a null root we attempt to move to a stationary root (move M1) with probability j_{ns} and leave it at zero with probability $1 - j_{ns}$. If the root is stationary we attempt to deactivate it (move M1) with probability j_{sn} , move it to the unit circle (move M2) with probability j_{su} or update the frequency and modulus (move M3) with probability $1 - j_{sn} - j_{su}$. Finally if the root is currently on the unit circle we attempt to move it to the stationary region (move M2) with probability j_{us} and attempt to update the frequency (move M4) with probability $1 - j_{us}$.

2.4.3 Weights

The Dirichlet prior on the weights results in simple full conditional posteriors. For the real roots, let n_{ik}^r be the number of roots in state k in series i . Summing over the series and forming the vector n^r with elements $n_k^r = \sum_{i=1}^N n_{ik}^r$ the full conditional posterior for the weights on the real roots is

$$w^r | \cdot \sim \text{Dir}(w^r | n^r + k_r \delta^r).$$

Similarly, the full conditional posterior for the weights on the complex roots is

$$w^c | \cdot \sim \text{Dir}(w^c | n^c + k_c \delta^c).$$

It is clear from the form of the full conditional posterior that the performance of the chain can suffer if the prior on the weights is uninformative with small values for k_r or k_c . In this case the Markov chain can get trapped in a state where essentially all the weight is assigned to a single root state.

2.4.4 Covariance matrix

With the innovations assumed to be multivariate normal, updating the covariance matrix is done within the normal Wishart conjugate framework. Calculating the residuals for each series

$$e_{i,t} = \prod_{j=1}^{\bar{r}} (1 - \rho_{ij}L) \prod_{j=1}^{\bar{c}} (1 - 2m_{ij}a_{ij}L + m_{ij}^2L^2) y_{it}$$

and forming

$$A = \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t'$$

The full conditional posterior is then inverse Wishart

$$\Sigma | \cdot \sim \mathcal{W}^{-1}(\Sigma | A + B, d + T).$$

2.4.5 Latent initial values

Conditional on the roots (or the autoregressive polynomials) and the variance covariance Σ , the latent initial values are easily generated as follows. For $s = 0, \dots, 1 - \bar{p}$, generate draws ε_s from a multivariate normal, $\mathcal{N}(\mathbf{0}, \Sigma)$. The latent initial values are then calculated from the autoregressions $\phi_i(L) y_{it} = \varepsilon_{it}$ by reversing the time index. That is

$$\begin{aligned} \phi_i(F) y_{it} &= \varepsilon_{it} \\ y_{is} &= \sum_{j=1}^{\bar{p}} \phi_j y_{j,s+j} + \varepsilon_{is} \end{aligned}$$

for $s = 0, \dots, 1 - \bar{p}$. Note that this is possible even when the polynomial includes unit roots, see Huerta and West (1999a) for a proof.

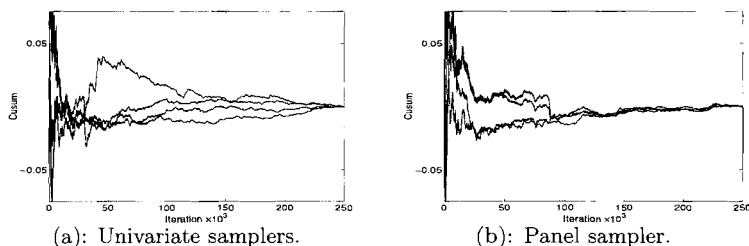
2.5 Purchasing power parity

As a consequence of the law of one price, purchasing power parity (PPP) is a fundamental assumption in economics with a large body of empirical

Table 2.1 Select results for the individual time series, univariate estimation and as a panel. The posterior probability of observing a unit root and the posterior mean of the first few coefficients of the autoregressive polynomial.

| | | $P(\rho_{\max} = 1)$ | ϕ_1 | ϕ_2 | ϕ_3 | ϕ_4 | ϕ_5 | ϕ_6 |
|-----|----|----------------------|----------|----------|----------|----------|----------|----------|
| (U) | CA | 0.70 | 1.307 | -0.307 | 0.348 | -0.490 | 0.150 | -0.017 |
| | DE | 0.49 | 1.176 | -0.217 | 0.021 | -0.011 | -0.006 | 0.001 |
| | JP | 0.56 | 1.275 | -0.318 | 0.039 | -0.008 | -0.013 | 0.004 |
| | UK | 0.46 | 1.059 | -0.106 | -0.003 | 0.001 | -0.003 | 0.001 |
| (P) | CA | 0.75 | 1.312 | -0.296 | 0.315 | -0.460 | 0.139 | -0.015 |
| | DE | 0.41 | 1.199 | -0.248 | 0.030 | -0.016 | -0.008 | 0.001 |
| | JP | 0.27 | 1.310 | -0.370 | 0.068 | -0.028 | -0.031 | 0.011 |
| | UK | 0.44 | 1.114 | -0.163 | -0.001 | 0.007 | -0.010 | 0.013 |

Figure 2.1 Assessing the convergence of the samplers with standardized CUSUM statistics.



work devoted to testing for the presence of PPP. To date the results have been mixed, see Rogoff (1996) for an overview of the literature. Implying stationary real exchange rates, PPP is a natural context for applying unit root tests. While unit root testing is a large subset of the empirical literature on PPP there has been relatively little use of Bayesian methods, one notable exception being Schotman and van Dijk (1991).

The lack of unambiguous results in the literature is usually attributed to data and model limitations. While we have no control over the data, the flexible model used here address two of the more common concerns, namely model uncertainty, with respect to lag length, and cross-sectional dependency. We use quarterly real exchange rates against the US dollar for a small number of industrialized countries to investigate the support for PPP. The observation period, from the first quarter of 1973 to the last quarter of 1998 is selected to

coincide with the introduction of the float and to avoid any issues connected with the introduction of the Euro currency. Our primary panel consists of Canada, Germany, Japan and the UK.

In specifying the prior, the maximum number of real roots and complex pairs are set to 6 and 3 respectively, partly in response to the fact that the series are short, with just 100 observations, and partly reflecting our prior belief that any dynamics present ought to be simple. For the real roots we select the default weights implying that we expect to find some dynamics. However, any dynamics present is expected to be simple, as in not cyclical and we set p_0 to $2/3$, so that the *a priori* probability of observing at least one complex pair is in the neighborhood of 50%. The remaining mass is mainly concentrated on the arbitrary frequency components, as we do not expect to find a seasonal pattern. Specifically

$$\delta^r = \{1/2, 1/6, 1/6, 1/6\}, \delta^c = \{2/3, 9/60, 1/60, 9/60, 1/60\},$$

and the scale factors in the weight prior are set to equal total number of real and complex roots, ensuring that the weights will avoid any near-trapping states and thus improving the convergence properties of the sampler. For the variance, we select a diagonal structure on the scale matrix with a single degree of freedom. The prior standard deviation is set to 1000, making the prior proper but vague.

To establish a benchmark we begin by fitting each series individually. For this, as in all other cases, Markov chains are run for 250 000 iterations after discarding a short initial transient of 1 000 samples as burn in. An issue of some concern is convergence. While the coefficients of the autoregressive polynomial converge fast, with the model exploring the root configuration space at a healthy rate, we are mainly interested in the probability of observing a unit root. While this kind of quantity can be hard to pin down with accuracy, studying the standardized CUSUM plots in Figure 2.2.(a) shows how the univariate samplers seem to be converging, if not smoothly at least fast. In the top panel of Table 2.1, the posterior probability of a unit root ranges from a modest 0.46 for the UK to a substantial 0.70 for Canada. Noting that the prior probability of at least one unit root is 0.55 in the univariate case the results are, with the possible exception of Canada, not supportive of the presence of a unit root. They are, on the other hand, not supportive of stationary dynamics either. The data is simply not able to distinguish between the two hypothesis.

As can be seen in Figure 2.2, with the exception of Canada the typical autoregressive orders are small and associated with real root dynamics though all samplers do explore the full support. This is reflected in Table 2.1 where

the contribution of the model averaged coefficients quickly tend to zero. For Canada the larger order is associated with one complex pair at the seasonal frequency and one more at a multiple of the seasonal frequency. With no obvious seasonal pattern in the data the interpretation of this is unclear.

Collecting the four series into a panel and treating them jointly makes a difference, as seen in the bottom panel of Table 2.1 and by studying the covariance structure in Table 2.2. Though more complicated the sampler seems to be doing a good job, the unit root probabilities still converging quickly. The strength of the correlation varies, from a low -0.02 between Canada and Germany to a respectable 0.44 between Japan and the UK. An implication of the heterogeneity in the correlation pattern is that centering across time is ineffective as a means of circumventing the problem of correlation. For the lag lengths there is a small shift in the posterior where we tend to get slightly higher orders in the panel.

Turning to the issue of unit roots, the effect of panel estimation on the posterior probabilities of unit roots is, at first sight not very large except for Japan. The change in scale factor for the Dirichlet prior for the weights from $\bar{\tau} = 6$ in the univariate setting to $N\bar{\tau} = 24$ in the panel affects the prior probability of unit roots. For the panel the prior probability of at least one unit root in a series is 0.63 . Calculating Bayes factors as $\frac{P/(1-P)}{\pi/(1-\pi)}$ where π is the prior probability of a unit root and P the posterior probability we find weak evidence in the data against a unit root for all series except Canada.

One of the more important improvements that a panel approach offers is in terms of more distinct inference with respect to unit root composition in a collection of time series. An attempt at illustrating this is made in Figure 2.3 where the left histogram shows the resulting joint posterior of the number of unit roots based on the output of the individual independent samplers. Just the sum of independent binomial quantities, the posteriors centers around the expected value which in this case is around three. Contrasting, the panel has a double effect. First, the marginal probabilities of a unit root associated with each series are affected. Second, the joint outcomes are not independent. The total effect is illustrated in the right histogram where the change in marginal probabilities shifts the distribution to the left towards the area associated with a jointly stationary panel and the dependence skews the distribution. That is unit roots, when detected, occur in slightly larger clusters than we would expect given independence. The appearance of the posterior is one of four possible characteristic outcomes where a truly independent panel will center on a simple expected value and otherwise the dependence may lead to clustering at either or both ends. For this particular example, the results

Figure 2.2 The posterior autoregression order for the univariate samplers (top) and the panel sampler (bottom).

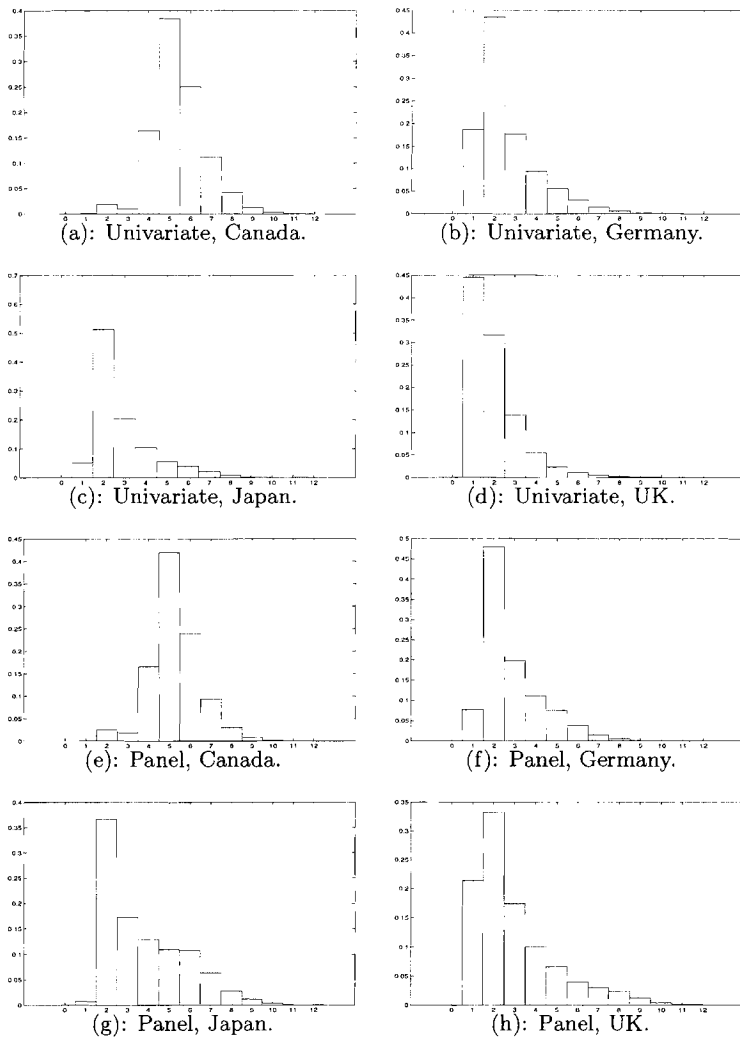


Figure 2.3 The effect of panel estimation on the total distribution of unit roots in the collection. The number of time series with a unit root, based on the independent output of univariate MCMC's (left) and on the output of a panel MCMC (right).

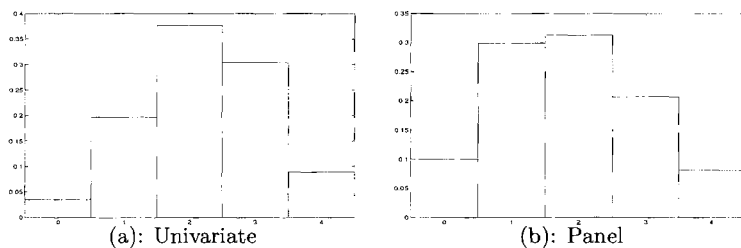


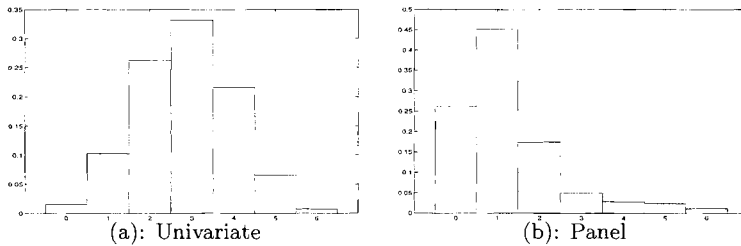
Table 2.2 Posterior estimated cross correlations estimated from the output of the panel sampler.

| | | DE | JP | UK |
|----|--------|--------|-------|-------|
| CA | 1.000 | -0.020 | 0.103 | 0.123 |
| DE | -0.020 | 1.000 | 0.069 | 0.110 |
| JP | 0.103 | 0.069 | 1.000 | 0.441 |
| UK | 0.123 | 0.110 | 0.441 | 1.000 |

are heavily influenced by the outcome for Canada. The difference between univariate and panel modelling is even more striking if we compare the priors for the number of unit root series with their respective posteriors. In each case this is a Binomial distribution, for the univariate approach we have the prior probabilities (0.04, 0.20, 0.37, 0.30, 0.09) for 0 to 4 unit root series and (0.01, 0.13, 0.33, 0.37, 0.16) in the panel setting. The shift in the posterior is thus even larger relative to the prior distribution, indicating considerably more support for stationary dynamics when the data is modelled as a panel.

We end with a cautionary note. In general we would expect to gain by adding more countries to the panel but, at least, in the case of PPP this is not necessarily the case. Adding two more countries, France and Italy, to the panel we obtain the univariate and panel results for the number of unit series in Figure 2.4 and the evidence in favour of stationarity appear to be even stronger. As indicated by the plot of the running means for the probability of a unit root in Figure 2.5 the sampler is experiencing convergence problems with this panel and the output is suspect. As both Italy and France, along

Figure 2.4 The G6 panel. The effect of panel estimation on the total distribution of unit roots in the collection. The number of time series with a unit root, based on the independent output of univariate MCMC's (left) and on the output of a panel MCMC (right).



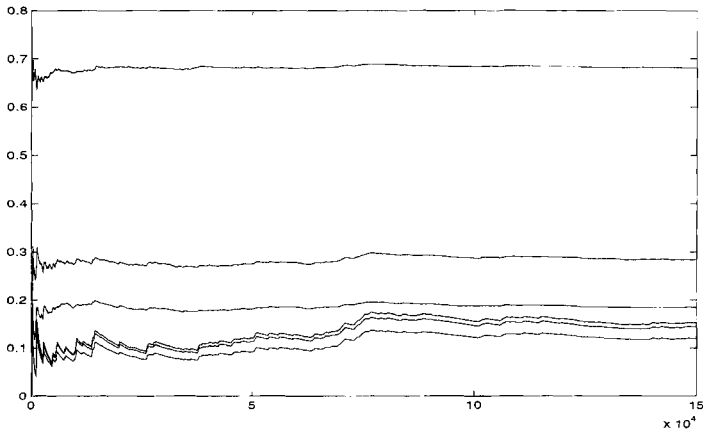
with Germany, were members of ERM and its precursors during the sample period adding these countries to the panel will contribute little independent variation. One indication of the problem is the high residual correlation of at least 0.95 between these countries. The high correlation causes the variance-covariance matrix to be near rank deficient and we believe that this is what causes the poor behavior of the sampler.

2.6 Final remarks

This paper presents a simple yet flexible autoregressive model for panel data. Building on the work in Huerta and West (1999a) and Huerta and West (1999b), the model addresses some major concerns in the field. First, with a general covariance matrix, cross-section correlation is explicitly modelled. This creates some practical difficulties when modeling the individual roots of each series as the necessary conditioning is deep and as such affects the mixing properties and the efficiency of the Markov chain. However, judging from the example, for the size of panels usually encountered in macroeconomic applications the performance can be expected to be fair. Second, using mixed priors on the roots and assigning a positive prior probability of a root at the origin, models of varying lag length are considered at least up to some pre-specified maximum. Further, with a lag polynomial fitted separately for each series, the model allows different orders between series. Thus, a simple form of heterogeneity is allowed.

The illustrations shows how, barring problematic panels, the model and

Figure 2.5 The G6 panel. Assessing the convergence of the samplers by inspection of the running means.



associated sampling strategy works well. The convergence is swift and clean, the sampler employed is quick and robust, and the inferential possibilities are rich. Further, contrary to standard methods, the suggested model offers simple means of monitoring the quality of the results. In the G6 example, one of the first tell tale signs of problems is the growing of long lag polynomials, a behavior associated with phenomena such as over differencing. The G6 example also serves as a reminder of a simple fact which is often forgotten. When confronted with slow, bad, or no convergence it is not necessarily due to poor implementation.

Experimenting with the model raises some issues that are worth considering. One is that the default prior choices tend to favor large models. This in itself is not a problem, but indicated that care should be taken when specifying the prior. This is directly connected to how the weights in the mixture priors are modelled and finding more flexible, yet easily interpretable, specifications is on our agenda for future research.

Bibliography

- Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* pp. 711–732.
- Huerta, G. and West, M. (1999a), 'Bayesian inference on periodicities and component spectral structures in time series', *Journal of Time Series Analysis* pp. 401–416.
- Huerta, G. and West, M. (1999b), 'Priors and component structures in autoregressive time series models', *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Im, K. S., Pesaran, M. H. and Shin, Y. (2003), 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics* **115**, 53–74.
- Karlsson, S. and Löthgren, M. (2000), 'On the power and interpretation of panel unit root tests', *Economics Letters* **66**, 249–255.
- Levin, A., Lin, C. F. and Chu, C. J. (2002), 'Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties', *Journal of Econometrics* **108**, 1–24.
- O'Connell, P. G. J. (1998), 'The overvaluation of purchasing power parity', *Journal of International Economics* **44**, 1–19.
- Quah, D. (1994), 'Exploiting cross-section variation for unit root inference in dynamic data', *Economics Letters* **44**, 9–19.
- Rogoff, K. (1996), 'The purchasing power parity puzzle', *Journal of Economic Literature* **34**, 647–668.
- Schotman, P. C. and van Dijk, H. K. (1991), 'A bayesian analysis of the unit root hypothesis in real exchange rates', *Journal of Econometrics* **49**, 195–238.

Paper 3

Parametric Covariance Matrix Modelling in Bayesian Panel Regression

3.1 Introduction

Many interesting panel regression models may be defined in terms of a simple covariance structure specified by just a few parameters. Examples include, but are in no way limited to, error component models and models having serially correlated errors. Parametrizing the mean and covariance structures separately, any difficulties encountered are mostly associated with the latter. In Bayesian analysis, problems usually stem from a lack of analytical results and the modern solutions typically involve posterior simulation together with Markov chain Monte Carlo methods.

Data augmentation, introduced by Gelfand and Smith (1990), is a frequently used tool where the ability to sample, augment and condition on latent variables may result in major simplifications for convenient prior structures. Although generally applicable, data augmentation is not necessarily efficient. Also, while elegant and simple, for covariance modeling it is never strictly necessary as parameterization of the covariance matrix and direct sampling is always possible, albeit not always very practical. This paper shows how direct sampling can be implemented for common panel data models and further demonstrates how direct sampling is a viable alternative to data augmentation.

For direct sampling to be practical and effective, the main requirement is that the simple structure of the covariance matrix, in a broad sense, carries over to its inverse. When this is the case, convenient expressions for the full conditional posterior are readily available. Complicating the posterior simulation, these will typically not be known densities. Often involving simple, but possibly nonlinear, polynomials in the covariance parameters they may instead be used to create arbitrarily good approximations. These may in turn be used to either sample candidate values for Metropolis type updates, apply the gridgy Gibbs sampler of Ritter and Tanner (1992), or form the basis of any resampling method. Whatever the preferred method, an expression for the determinant in terms of the parameters and bounded support for the parameters is helpful but not decisive; the alternative being brute force calculations.

In addition to providing a generic, if not always practical, solution to some of the problems associated with data augmentation, direct sampling brings other advantages. Many models imply restrictions on the covariance matrix, so that direct parametrizing will be efficient in terms of the number of unknowns in the model. For instance, in error component models the status of an effect as present or absent is typically governed by a single parameter, to be compared with many latent and augmented effects. This also simplifies the administration and implementation of model selection, keeping the necessary adjustments to the posterior simulation at a minimum. The changing

dimensionality of the parameter vector necessitates the use of more sophisticated sampling procedures such as the reversible jump Metropolis-Hastings developed by Green (1995).

The organization of this paper is as follows. In Section 3.2, direct covariance sampling is motivated and exemplified with the standard one-way random effects specification. The covariance matrix properties that make a direct approach practical are briefly discussed. Section 3.3 is concerned with elements of posterior simulation. A generic algorithm is detailed for the case with standard conjugated priors on all other quantities. Two panels are then used in Section 3.4 to illustrate the method in the context of the one-way error component model. The performance is compared to two standard algorithms of varying level of sophistication. Section 3.5 extends the prior to allow model selection and outlines a generic algorithm to include it. Effect selection is then illustrated in the context of the two-way error component model. Closing, Section 3.6 summarizes our experiences to date and offers some ideas for future research.

3.2 Parametric Covariance Matrix Modelling

3.2.1 An illustrative example

Consider the standard panel regression model where the mean response y_{it} of individual i at time t depends on a vector of p explanatory variables \mathbf{x}_{it} with associated fixed effects β , and is observed with error ϵ_{it}

$$y_{it} | \mathbf{x}_{it} = \mathbf{x}_{it}\beta + \epsilon_{it}, i = 1, \dots, N, t = 1, \dots, T. \quad (3.1)$$

Stacking observations associated with unit i into a $T \times 1$ vector \mathbf{y}_i , and a $p \times T$ matrix \mathbf{X}_i Bayesian inference may be conducted within the linear regression framework. Assuming $\epsilon_i \sim \mathcal{N}(0, \Sigma_\epsilon)$ and convenient prior choices, analytical results are available for the general covariance structure and the trivial model with marginally independent and homoscedastic errors.

Contrasting, many interesting models reduce the number of unique elements in Σ_ϵ or imply exact restrictions in terms of a small number of parameters. A good example is the simple one-way random effects model. Adding an individual, but time invariant, zero mean random effect, μ_i , to an idiosyncratic error, ν_{it} , results in the addition of a level to the model and a reduction of the number of unique elements of Σ_ϵ to just two, both being simple functions of the error component variances, σ_μ^2 and σ_ν^2 respectively. In particular

$$\Sigma_\epsilon = \sigma_\nu^2 I_T + \sigma_\mu^2 J_T, \quad (3.2)$$

where I_T the identity matrix and J_T a matrix of ones. Early examples of a full Bayesian treatment can be found in Tiao and Tan (1965) and Hill (1965). Box and Tiao (1973) contains a thorough presentation.

Assuming $\mu_i \sim \mathcal{N}_T(0, \sigma_\mu^2)$, $\nu_{ij} \sim \mathcal{N}(0, \sigma_\nu^2)$, μ_i independent of ν_{ij} , and \mathbf{x}_{ij} independent of both μ_i and ν_{ij} , and a conjugated prior that preserves the hierarchical structure, the modern full Bayesian treatment is performed by means of a sequence of Gibbs steps. In this context, the method of data augmentation introduced by Tanner and Wong (1987), is a crucial element, where inference is facilitated by the ability to sample, augment and condition on the latent random effects. Among others, Gelfand, Sahu and Carlin (1995), Vines, Gilks and Wild (1996), and Gilks and Roberts (1996) propose various refinements to the basic algorithm presented in Gelfand and Smith (1990). Chib and Carlin (1999) offers a brief overview and presents two of the best procedures available to date.

While representing a neat solution to the problem, data augmentation is not necessary when the collection of individual effects $\{\mu_i\}_{i=1}^N$ as such are of no particular interest. The unknown parameters of the model in (3.1) are $\{\beta, \sigma_\nu^2, \sigma_\mu^2\}$ and the joint distribution of all quantities, known and unknown, is described by

$$p\left(\{\mathbf{y}_i\}_{i=1}^N, \beta, \sigma_\nu^2, \sigma_\mu^2\right) = L\left(\{\mathbf{y}_i\}_{i=1}^N \mid \beta, \sigma_\nu^2, \sigma_\mu^2\right) \pi\left(\beta, \sigma_\nu^2, \sigma_\mu^2\right),$$

where the conditioning on the explanatory variables is implied. Ignoring a proportionality constant, given the assumptions the likelihood is

$$L\left(\{\mathbf{y}_i\}_{i=1}^N \mid \beta, \sigma_\nu^2, \sigma_\mu^2\right) \propto |\Sigma_\epsilon|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^N \left(\mathbf{y}_i - \mathbf{X}_i' \beta\right)' \Sigma_\epsilon^{-1} \left(\mathbf{y}_i - \mathbf{X}_i' \beta\right)\right\}$$

and a standard prior structure is

$$\pi\left(\beta, \sigma_\nu^2, \sigma_\mu^2\right) = \pi\left(\beta \mid \sigma_\nu^2, \sigma_\mu^2\right) \pi\left(\sigma_\mu^2 \mid \sigma_\nu^2\right) \pi\left(\sigma_\nu^2\right).$$

Ignoring the matter of β for the moment, the full conditional posterior of $(\sigma_\nu^2, \sigma_\mu^2)$ is then

$$p\left(\sigma_\nu^2, \sigma_\mu^2 \mid \beta, \{\mathbf{y}_i\}_{i=1}^N\right) \propto \left|\sigma_\nu^2 I_T + \sigma_\mu^2 J_T\right|^{-N/2} \times \exp\left\{-\frac{1}{2} \sum_{i=1}^N \mathbf{e}_i' \left(\sigma_\nu^2 I_T + \sigma_\mu^2 J_T\right)^{-1} \mathbf{e}_i\right\} \pi\left(\sigma_\mu^2 \mid \sigma_\nu^2\right) \pi\left(\sigma_\nu^2\right) \quad (3.3)$$

where

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i' \beta.$$

The main problem is the inverse in the exponent and the determinant. To make some progress, define the idempotent counterparts of J_t and I_t

$$\bar{J}_T \equiv T^{-1} J_T, E_T \equiv I_T - \bar{J}_T,$$

substitute into (3.2) and collect terms to get

$$\Sigma_\epsilon = \sigma_\nu^2 E_T + (T\sigma_\mu^2 + \sigma_\nu^2) \bar{J}_T = \sigma_\nu^2 [E_T + (T\sigma_\mu^2 \sigma_\nu^{-2} + 1) \bar{J}_T].$$

This suggests a useful reparametrization of the model. Define

$$\theta_\mu \equiv (T\sigma_\mu^2 \sigma_\nu^{-2} + 1)^{-1}, 0 < \theta_\mu < 1,$$

and write the covariance matrix as

$$\Sigma_\epsilon = \sigma_\nu^2 (E_T + \theta_\mu^{-1} \bar{J}_T) = \sigma_\nu^2 \Sigma_\theta.$$

Since θ_μ^{-1} and 1 are the two distinct characteristic roots of Σ_θ with multiplicity 1 and $T - 1$ we get

$$|\Sigma_\epsilon| = \sigma_\nu^{2T} |\Sigma_\theta| = \sigma_\nu^{2T} \theta_\mu^{-1}, \Sigma_\epsilon^{-1} = \sigma_\nu^{-2} \Sigma_\theta^{-1} = \sigma_\nu^{-2} (E_T + \theta_\mu \bar{J}_T).$$

Substituting into (3.3) and conditioning on θ_μ note how the full conditional posterior of σ_ν^{-2} has a gamma kernel

$$p(\sigma_\nu^2 | \beta, \theta_\mu, \{\mathbf{y}_i\}_{i=1}^N) \propto (\sigma_\nu^{-2})^{NT/2} \exp\{-S\sigma_\nu^{-2}\} \pi(\sigma_\nu^2),$$

where

$$S = \frac{1}{2} \sum_{i=1}^N \mathbf{e}_i' (E_T + \theta_\mu \bar{J}_T) \mathbf{e}_i.$$

Also, conditioning further on σ_ν^2 , the full conditional posterior of θ_μ can be written as

$$p(\theta_\mu | \beta, \sigma_\nu^2, \{\mathbf{y}_i\}_{i=1}^N) \propto \theta_\mu^{N/2} \exp\{c_1 + c_2 \theta_\mu\} \pi(\theta_\mu | \sigma_\nu^2), \quad (3.4)$$

where

$$c_1 = -\frac{\sigma_\nu^{-2}}{2} \sum_{i=1}^N \mathbf{e}_i' E_T \mathbf{e}_i, c_2 = -\frac{\sigma_\nu^{-2}}{2} \sum_{i=1}^N \mathbf{e}_i' \bar{J}_T \mathbf{e}_i.$$

While not a known density, being a simple function of the parameter of interest, the expression can be used for quick and efficient evaluation of the

posterior for select values. Doing so for a reasonably large number of values, an adequate first order approximation of the full conditional posterior may be constructed. This can then be used to sample the next value immediately, applying the griddy Gibbs sampler of Ritter and Tanner (1992), to sample a candidate value, executing a Metropolis-Hastings step, or to form the basis for the implementation of any resampling method such as rejection sampling. In this particular case, the contribution of the likelihood has the form of a truncated gamma for θ_μ which may be exploited in the posterior simulation.

3.2.2 A general direct approach

The results in Section 3.2.1 for the one-way error component model is just an example of a general form that is always possible to achieve albeit not necessarily practical. The crucial point is the ability to write the inverse of the covariance matrix as

$$\Sigma_\epsilon^{-1} = \sigma_\nu^{-2} \Sigma_\theta^{-1} = \sigma_\nu^{-2} \sum_{i=1}^m g_i(\theta) Q_i, \quad (3.5)$$

for g_i simple functions of θ and Q_i constant matrices with m ideally much smaller than the dimension of Σ_θ . In the example, from (3.4) we get

$$m = 2, g_1 = 1, Q_1 = E_T, g_2 = \theta_\mu, Q_2 = \bar{J}_T.$$

Contrasting, in the completely general case with no restrictions, the parameter vector includes the $T(T+1)/2$ unrestricted elements of the inverse, say $\{\lambda_i\}_{i=1}^{T(T+1)/2}$. The equal number of fixed matrices are zero everywhere except either at one position on the main diagonal or two positions symmetrically to the main diagonal.

Typically, every restriction introduced reduces the number of parameters, functions, and matrices needed. For direct sampling to be practical m should be a small number relative $T(T+1)/2$. Thus, assuming we can easily construct the inverse as above, the counterpart of the full conditional (3.4) for an arbitrary element θ_j of θ is

$$p\left(\theta_j | \beta, \sigma_\nu^2, \theta_{-j}, \{\mathbf{y}_i\}_{i=1}^N\right) \propto |\Sigma_\theta|^{-N/2} \exp\left\{-\frac{\sigma_\nu^{-2}}{2} \sum_{i=1}^N \mathbf{e}' \sum_{k=1}^m g_k(\theta) Q_k \mathbf{e}_i\right\} \pi\left(\theta_j | \sigma_\nu^2\right),$$

and θ_{-j} contains all parameters in θ except θ_j . Reorganizing this expression

slightly

$$p\left(\theta_j | \beta, \sigma_\nu^2, \theta_{-j}, \{\mathbf{y}_i\}_{i=1}^N\right) \propto |\Sigma_\theta|^{-N/2} \exp\left\{\sum_{k=1}^m c_k \cdot g_k(\theta)\right\} \pi\left(\theta_j | \sigma_\nu^2\right), \quad (3.6)$$

with the coefficients given by

$$c_k = -\frac{\sigma_\nu^{-2}}{2} \sum_{i=1}^N \mathbf{e}_i' Q_k \mathbf{e}_i.$$

In practice, using (3.6) we will only need to consider those $g_k(\theta)$ that actually reference θ_j . Updating more than one element of θ at a time is possible though constructing and administrating a multidimensional approximation increases the complexity of the implementation. On the other hand, if the dimension of θ is large it may be necessary to do so to counter the potentially high correlation and improve the mixing of the Markov chain.

In addition to an explicit structure for the inverse, computational efficiency of the direct approach described relies on the availability of an expression for the determinant as a function of θ , say $d(\theta)$. Also helpful, but in no way essential, is that the support of the parameters in θ is bounded as this can simplify the construction of an approximation to the full conditional posterior.

3.3 Posterior Simulation

3.3.1 The model

For the unknown parameters of the model in (3.1), $\{\beta, \sigma_\nu^2, \theta\}$, the joint likelihood is proportional to

$$L\left(\{\mathbf{y}_i\}_{i=1}^n | \beta, \sigma_\nu^2, \theta\right) \propto \sigma_\nu^{-NT} d(\theta)^{-N/2} \times \exp\left\{-\frac{\sigma_\nu^{-2}}{2} \sum_{k=1}^m g_k(\theta) \left[\sum_{i=1}^N \mathbf{e}_i' Q_k \mathbf{e}_i\right]\right\}.$$

The prior structure adopted is

$$\pi\left(\beta, \sigma_\nu^2, \theta\right) \sim \pi\left(\beta | \sigma_\nu^2, \theta\right) \prod_{k=1}^m \pi\left(\theta_k | \sigma_\nu^2\right) \pi\left(\sigma_\nu^2\right),$$

where the conditional independence structure for the covariance parameters is dictated by convenience. For the regression coefficients β and the idiosyncratic error precision σ_ν^{-2} we consider the standard conjugate choices

$$\pi\left(\beta | \sigma_\nu^2, \theta\right) \sim \mathcal{N}_p\left(b_0, \sigma_\nu^2 B_0\right), \pi\left(\sigma_\nu^{-2}\right) \sim \mathcal{G}\left(\nu_0, \nu_1\right).$$

This leads to simple updates in the form of Gibbs steps. The prior for θ is left unspecified at this point as its precise definition is problem dependent and also not essential for the general description of the posterior simulation.

3.3.2 A generic algorithm

For the model in Section 3.3.1, a simple three-step Markov-chain Monte Carlo algorithm suggests itself.

In Step 3.b of Algorithm 1 three widely used options are the griddy Gibbs sampler, Metropolis type updates and rejection sampling. Choosing the first option a single value is sampled for θ_k from $q(\theta_k)$ and it is accepted forthwith. Alternatively, the second option entails sampling a single value θ_k from $q(\theta_k)$, calculating an acceptance probability, and then sampling the next state. The expression for the acceptance probability is the familiar product of a likelihood, prior, and proposal ratio which in this case simplifies to

$$\begin{aligned}\alpha_{\theta_k} &= \min \left\{ 1, \kappa_1 \cdot \exp \left\{ \sum_{i=1}^m c_i \cdot \left[g_i \left(\theta_k^{(p)} \right) - g_i \left(\theta_k^{(c)} \right) \right] \right\} \cdot \kappa_2 \right\}, \\ \kappa_1 &= \left[\frac{d \left(\theta_k^{(p)} \right)}{d \left(\theta_k^{(c)} \right)} \right]^{N/2}, \quad \kappa_2 = \frac{\pi \left(\theta_k^{(p)} \mid \sigma_\nu^2 \right) q_{\theta_k} \left(\theta_k^{(c)} \right)}{\pi \left(\theta_k^{(c)} \mid \sigma_\nu^2 \right) q_{\theta_k} \left(\theta_k^{(p)} \right)}\end{aligned}$$

where $\theta_k^{(c)}$ is the current value and $\theta_k^{(p)}$ the proposed value. In the proposal ratio the current value is treated as if it had been sampled from the same density used to generate the proposed value. Depending on the quality of the approximation, α_θ is more or less close to one. Finally, implementing for instance a rejection sampler, values θ_k are sampled from $q(\theta_k)$ until a move to a new state is accepted. In this case we need to calculate an envelope constant A and the stopping rule is

$$A q_{\theta_k} \left(\theta_k^{(p)} \right) u \leq \kappa \cdot d \left(\theta_k^{(p)} \right) \exp \left\{ \sum_{i=1}^m c_i \cdot g_i \left(\theta_k^{(p)} \right) \right\} \pi \left(\theta_k^{(p)} \mid \sigma_\nu^2 \right),$$

where $u \sim U(0, 1)$ a uniform random number, $\theta_k^{(p)}$ the proposed value and κ a scale factor that is typically not known and needs to be computed numerically.

An important part of Algorithm 1 is the construction of and the sampling from the approximation $q(\theta_k)$. Assuming elements of θ are treated one by one, evaluating the full conditional posterior on a grid, is a simple task. For the quality of the approximation the issues of number and placement of nodes is critical. The classical conflict of precision and computational speed applies two ways. First, adding nodes will increase the precision of the estimated

cumulative density function at the cost of more function evaluations. Second, gains in precision from more advanced integration techniques or elaborate node placement strategies are often lost to loss of speed. Further, depending on the problem, more or less clever ways of constructing the approximation may be available.

3.4 Relative Performance

The merits of direct sampling are investigated in the context of the simple one-way error component model. Various approaches are used for direct sampling and the performance is compared with the two simpler algorithms in Chib and Carlin (1999); the standard 3-block algorithm, labeled A.1, and the improved 2-block algorithm, labeled A.2. Both are briefly described in Appendix A. A third, more complicated single block algorithm, is not used as it turned out to be very slow and sensitive to the choice of tuning parameters.

Two panels are used for the illustration. One is the Grunfeld panel which, having small unit and time dimensions, is a fairly typical example of the kind of panels appearing in macroeconomic applications. A second illustration uses a panel from the field of labor economics. With a cross-section size of 1571 it represents the kind of large panels that are becoming increasingly available.

In all examples a zero mean g-prior is selected for the regression parameters β and gamma priors for the precision of the error components. In direct sampling this implies a prior for the structural parameter θ_μ ,

$$\pi(\theta_\mu) \propto \theta_\mu^{\nu_3-1} (1 - \theta_\mu)^{-(\nu_3+1)} \exp\{-\nu_4 \theta_\mu T / (1 - \theta_\mu) / \sigma_\nu^2\},$$

where (ν_3, ν_4) are the prior parameters for the random effect precision. While perhaps not the most natural choice, it is used for comparability. As this choice results in a full conditional posterior which is very similar to a truncated gamma distribution, an attempt to use this is made in the context of Metropolis-Hastings updates by sampling proposals from

$$q(\theta_k) \sim \mathcal{G}(\nu_3 + N/2, \nu_4 T / \sigma_\nu^2 - c_2),$$

with c_2 defined (3.4).

Concentrating on the variance of the random effect and the intercept, various aspects of the resulting Markov chains are illustrated with the relative numerical efficiency as the main indicator of performance.

Algorithm 1 A generic algorithm.

1. Conditional on $\{\sigma_\nu^2, \theta\}$, sample and accept a proposal for β from its full conditional posterior. Due to the (conditional) Gaussian structure,

$$\mathbf{y}_i | \beta, \sigma_\nu^2, \theta \sim N_t(X_i \beta, \Sigma_\epsilon), i = 1, \dots, N$$

and combining for all units with the prior gives

$$\beta | \sigma_\nu^2, \theta \sim \mathcal{N}_p(b_1, \sigma_\nu^2 B_1), \quad (3.7)$$

where

$$B_1 = \left[B_0^{-1} + \sum_{j=1}^m g_j(\theta) \sum_{i=1}^N \mathbf{X}_i' Q_j \mathbf{X}_i \right]^{-1},$$

$$b_1 = B_1 \left(B_0^{-1} b_0 + \sum_{j=1}^m g_j(\theta) \sum_{i=1}^N \mathbf{X}_i' Q_j \mathbf{y}_i \right).$$

2. Conditional on $\{\beta, \theta\}$ sample and accept a proposal σ_ν^{-2} from its full conditional posterior. The conjugate structure results in

$$\sigma_\nu^{-2} | \beta, \theta \sim \mathcal{G}(\nu_0 + n_0, \nu_1 + S_1), \quad (3.8)$$

where

$$n_0 = NT/2, S_1 = \sum_{j=1}^m g_j(\theta) \sum_{i=1}^N \mathbf{e}_i' Q_j \mathbf{e}_i / 2.$$

3. Conditional on $\{\beta, \sigma_\nu^{-2}, \theta_{-k}\}$ cycle through θ updating θ_k .
 - (a) Selecting an appropriate number of nodes, construct an approximation to the full conditional posterior, $q(\theta_k)$,

$$\theta_k | \beta, \sigma_\nu^2, \theta_{-k} \propto d(\theta)^{-N/2} \exp \left\{ \sum_{i=1}^m c_i \cdot g_i(\theta) \right\} \pi(\theta_k | \sigma_\nu^2).$$

Note that we only need to evaluate the $c_i \cdot g_i(\theta)$ where θ_k is referenced.

- (b) Sample a new value for θ_k using any preferred method.
 4. Repeat steps 1-3 using the most recent values of all the conditioning variables.
-

Figure 3.1 Inference in the Grunfeld panel using data augmentation and the standard 3-block algorithm. Select Markov chain properties for the random effect variance σ_μ^2 and the intercept β_0 .

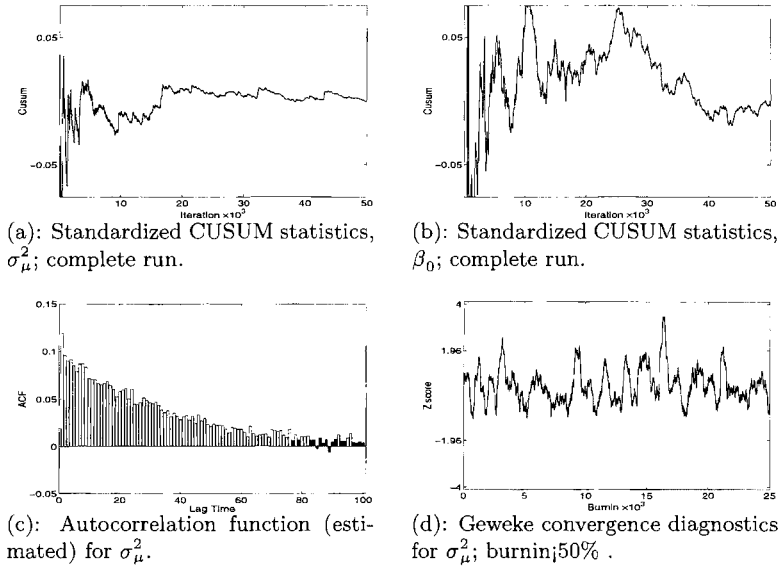


Table 3.1 Posterior simulation results for the Grunfeld panel; data augmentation and the simple 3-block algorithm.

| | Mean | S.d | Mode ^b | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|----------|----------|-------------------|----------|-----------------------|-------|--------------------|-------|
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -60.112 | 28.967 | -60.249 | 0.97 | 1.155 | 79.46 | 0.977 | 54.08 |
| β_1 | 0.109 | 0.010 | 0.110 | 0.90 | 0.217 ^c | 22.53 | 0.211 ^c | 20.28 |
| β_2 | 0.308 | 0.017 | 0.308 | 0.31 | 0.125 ^c | 2.66 | 0.127 ^c | 2.61 |
| σ_ν^2 | 2781.211 | 288.025 | 2721.409 | 0.05 | 1.355 | 1.11 | 1.337 | 1.02 |
| σ_μ^2 | 7306.572 | 4017.653 | 5065.369 | 0.12 | 48.732 | 7.36 | 44.982 | 5.95 |

^a Calculation based on estimated autocorrelation time.

^b Estimated from the kernel density estimate.

^c All values $\times 10^3$.

Figure 3.2 Inference in the Grunfeld panel using data augmentation and the improved 2-block algorithm. Select Markov chain properties for the random effect variance σ_μ^2 and the intercept β_0 .

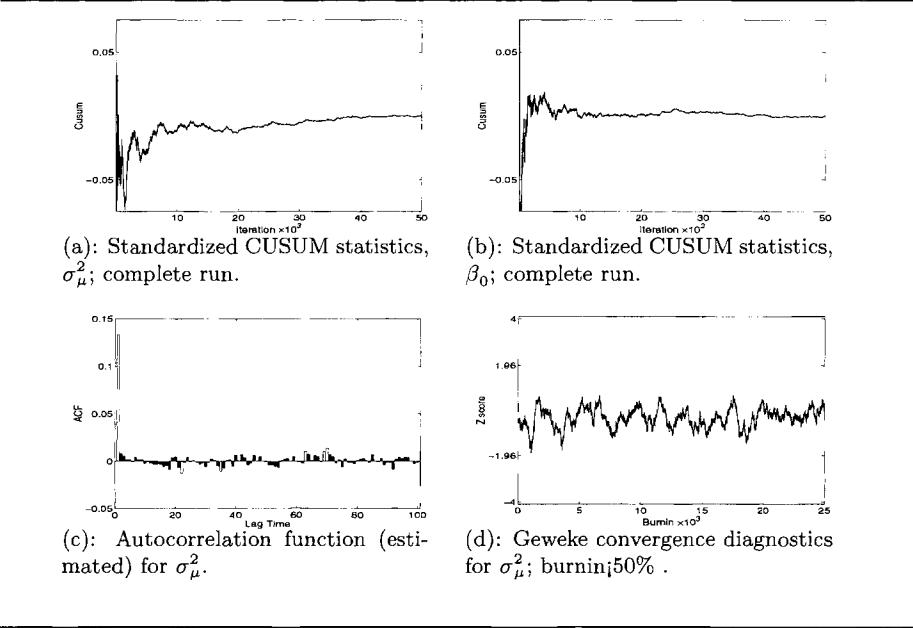


Table 3.2 Posterior simulation results for the Grunfeld panel; data augmentation and the simple 3-block algorithm.

| | Mean | S.d | Mode ^b | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|----------|----------|-------------------|----------|-----------------------|------|--------------------|------|
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -60.534 | 28.869 | -58.267 | 0.01 | 0.133 | 1.06 | 0.128 | 0.93 |
| β_1 | 0.109 | 0.010 | 0.109 | 0.01 | 0.046 ^c | 1.00 | 0.048 ^c | 1.02 |
| β_2 | 0.308 | 0.017 | 0.309 | 0.01 | 0.074 ^c | 0.94 | 0.078 ^c | 0.97 |
| σ_ν^2 | 2783.752 | 290.798 | 2725.108 | 0.06 | 1.371 | 1.11 | 1.385 | 1.08 |
| σ_μ^2 | 7319.079 | 4015.353 | 5106.032 | 0.13 | 20.183 | 1.26 | 20.426 | 1.23 |

^a Calculation based on estimated autocorrelation time.

^b Estimated from the kernel density estimate.

^c All values $\times 10^3$.

Figure 3.3 Inference in the Grunfeld panel using direct sampling with acceptance rejection. Select Markov chain properties for the random effect variance σ_μ^2 and the intercept β_0 .

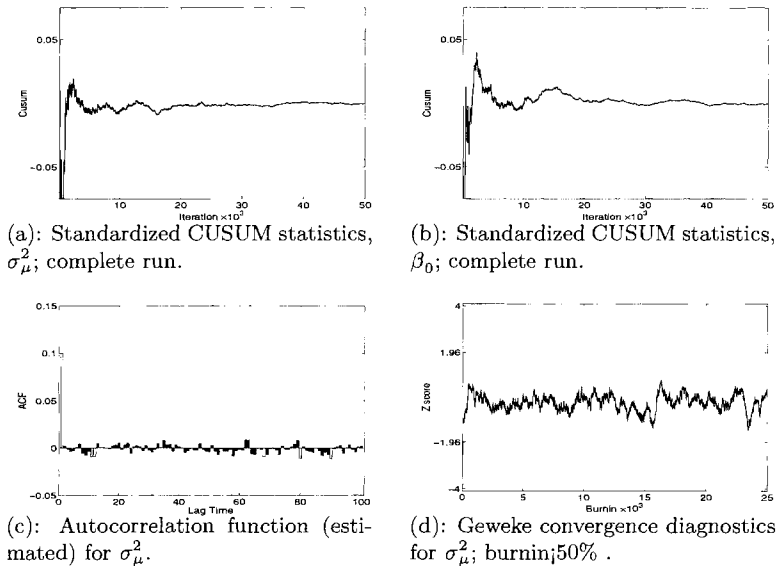


Table 3.3 Posterior simulation results for the Grunfeld panel; direct simulation using an equidistant grid on the truncated support and rejection resampling.

| | Mean | S.d | Mode ^b | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|----------|----------|-------------------|----------|-----------------------|------|--------------------|------|
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -60.507 | 28.737 | -58.536 | 0.00 | 0.130 | 1.02 | 0.127 | 0.98 |
| β_1 | 0.109 | 0.010 | 0.109 | -0.00 | 0.046 ^c | 0.98 | 0.049 ^c | 1.12 |
| β_2 | 0.308 | 0.017 | 0.309 | -0.00 | 0.075 ^c | 0.95 | 0.074 ^c | 0.92 |
| σ_ν^2 | 2812.870 | 293.840 | 2746.111 | 0.07 | 1.442 | 1.20 | 1.365 | 1.08 |
| σ_μ^2 | 7315.252 | 4017.892 | 5123.122 | 0.10 | 18.843 | 1.10 | 19.156 | 1.14 |
| θ_μ | 0.023 | 0.010 | 0.018 | 0.13 | 0.051 ^c | 1.26 | 0.051 ^c | 1.21 |

^a Calculation based on estimated autocorrelation time.

^b Estimated from the kernel density estimate.

^c All values $\times 10^3$.

Figure 3.4 Inference in the Grunfeld panel using direct sampling with Metropolis-Hastings and truncated gamma proposals. Select Markov chain properties for the random effect variance σ_μ^2 and the intercept β_0 .

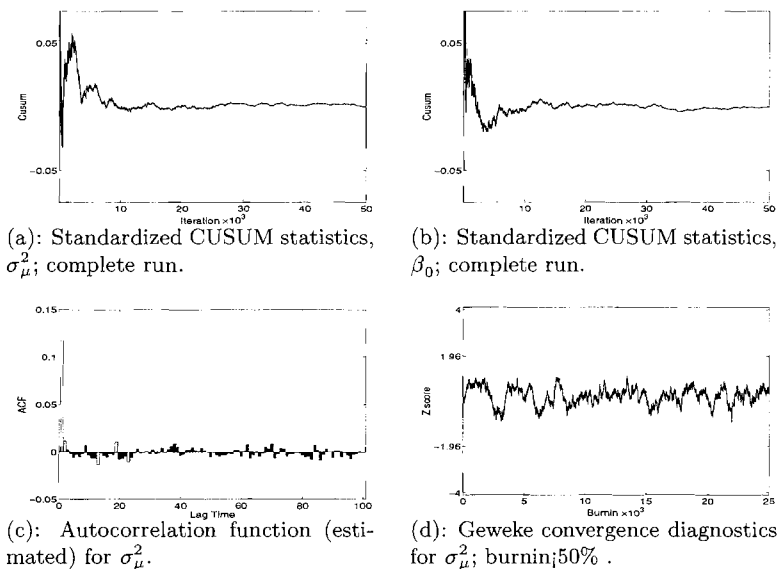


Table 3.4 Posterior simulation results for the Grunfeld panel; direct sampling using Metropolis-Hastings with truncated gamma proposals.

| | Mean | S.d | Mode ^b | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|----------|----------|-------------------|----------|-----------------------|------|--------------------|------|
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -60.695 | 28.810 | -58.277 | 0.01 | 0.129 | 1.00 | 0.128 | 0.99 |
| β_1 | 0.109 | 0.010 | 0.109 | 0.00 | 0.046 ^c | 0.98 | 0.046 ^c | 0.97 |
| β_2 | 0.308 | 0.017 | 0.309 | 0.00 | 0.077 ^c | 1.00 | 0.076 ^c | 0.98 |
| σ_ν^2 | 2813.651 | 292.700 | 2725.891 | 0.07 | 1.395 | 1.14 | 1.373 | 1.10 |
| σ_μ^2 | 7306.661 | 3979.154 | 4989.580 | 0.12 | 19.958 | 1.26 | 19.556 | 1.21 |
| θ_μ | 0.023 | 0.010 | 0.018 | 0.16 | 0.053 ^c | 1.35 | 0.052 ^c | 1.30 |

^a Calculation based on estimated autocorrelation time.

^b Estimated from the kernel density estimate.

^c All values $\times 10^3$.

3.4.1 Determinants of corporate investment

Grunfeld (1958) considered a simple model of corporate investment where the real gross investment of firm i at date j is assumed to be a function of the firms' real value measured by the value of outstanding shares, x_1 , and the real value of its capital stock, x_2 . The panel consists of $N = 10$ firms observed over a period of $T = 20$ years. All runs are made using the same or equivalent prior and 50000 samples are generated in each case. The same prior parameters are used for both variance components. In particular, $\nu_1 = \nu_3 = 1$ and $\nu_2 = \nu_4 = 100$, implying a low prior precision for data that is known to be noisy. For the regression coefficients, $g = 2000$ so that, as things are defined, the contribution of the prior precision will be small. To establish a benchmark and to illustrate the problems associated with the direct application of data augmentation, an attempt is made using Algorithm A.1. While the algorithm is very fast, the results are mixed as illustrated in Figure 3.1 and by Table 3.1.

In Figures 3.1.(a) and (b), an attempt to characterize the convergence properties of the chain is made using the Yu and Mykland (1998) CUSUM plots for the random effect variance and the intercept. For the former, the sample path is consistently keeping within the standard 5% tolerance. However, the convergence towards the final estimate is not smooth, being interrupted at intervals by sudden jumps. For the intercept matters are even worse, the sample path making long excursions away from the convergence path. Inspecting the estimated autocorrelation function for the random effect variance in Figure 3.1.(c) it is small but highly persistent. Because of the inefficiency of the algorithm, estimating numerical standard errors with any confidence is difficult. This creates problems that affect the construction of Geweke (1992) diagnostic statistics in Figure 3.1.(d). To err on the conservative side, an estimate based on the estimated autocorrelation time Chib and Carlin (1999) is used, selecting a large truncation lag. Even so, significant diagnostics are observed after so much as a 40% burnin. Discarding the initial 2500 iterations as burnin, the observations above are complemented and reinforced by the results in Table 3.1. The performance being what it is, the choice of cut-off point is arbitrary. Excess autocorrelation is a problem across the board and the performance, as measured by the relative numerical efficiency, is poor with the idiosyncratic error variance as the only exception.

Contrasting, the improvement offered by Algorithm A.2 is remarkable. In Figure 3.2, the convergence is clean, smooth, and fast, in particular for the intercept. What little autocorrelation is present dies out quickly. The estimate of the numerical standard error is robust, all methods yielding similar results for the various truncation and batch sizes tried. The Geweke statistics for the

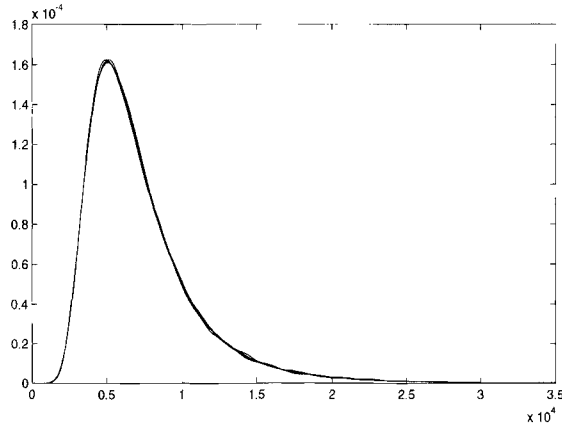
random effect variance are stable and insignificant. Discarding 2500 iterations again, the results in Table 3.2 show how the fortunes are completely reversed, sampling being efficient across the board. For the variance components improvement is concentrated to the random effect variance for obvious reasons. In terms of computational speed, the improvement is bought at a cost of an approximate 78% increase in execution time.

The results being so good for the improved 2-block algorithm leaves little, if any, room for improvement. Experimenting with various grid construction techniques and updating principles, the results for a simple equidistant grid on a truncated support using rejection sampling are illustrated in Figure 3.3 and Table 3.3. Using 51 nodes and refreshing the approximation in every iteration, the results are at least equally good as those using Algorithm A.2, depending primarily on what quantity is monitored (θ_μ or σ_μ^2). The average rejection rate is just below 0.3, meaning that on average 1.4 draws are needed to generate an accepted sample. Convergence seems to be immediate and there is very little autocorrelation in the chain as expected. Because the regression parameters are sampled the same way, that is after marginalization of the random effects, the performance in that part of the model is the same. The numerical standard error is not sensitive to choice of method, though the spectral estimate is fragile when the truncation lag increases beyond a certain point. The Geweke diagnostics are insignificant so there are no obvious signs of trouble. Turning to the table, the slight improvement in numerical efficiency for the random effect variance is not due to more efficient sampling of the effect parameter. Being computed as a function of the effect parameter and the idiosyncratic error variance the source of the apparent improvement is the more efficient sampling of the latter. On the downside, the execution time increased further by 8%.

The last experiment in this segment uses direct sampling truncated gamma proposals and Metropolis-Hastings updates. Cutting the execution time with more than 50% compared to Algorithm A.2, this sampler clearly illustrated the cost of maintaining and adapting a grid in every iteration. As illustrated in Figure 3.4 and Table 3.4, the improvement is achieved with only the slightest loss of efficiency.

Studying all the tables, it is clear how the choice of method does not affect the estimates of the marginal posterior distributions. This is illustrated in Figure 3.4.2 which contains the kernel density estimates based on the output of the various samplers tested. The slight differences present are mainly located in the right tail.

Figure 3.5 Kernel density estimates of random effect variance marginal posterior based on the output from various samplers.



3.4.2 Firm size effect

The firm size effect is usually discussed in the context of a wage premium paid to workers, where the premium depends on the firm size. Typically, individuals are taken as the unit of observation and in the standard framework the role played by the job and the firm is rather limited. Exploring a fresh angle, the job is brought into focus at the expense of individual workers, and the firm is made the unit of observation.

The data used consists of $N = 1571$ small to medium sized firms in the private sector observed over a period of $T = 10$ years. For all firms and each year the available information includes wages and job classification for all employees. Calculating percent deviations of paid wages from an economy wide job specific average wage for each employee, the response variable is the average wage premium paid by the firm to its employees. A simple firm size effect is expected to be reflected in firm specific average premium. Preliminary analysis offers weak evidence but a stronger relation between the intra-firm premium variability and the firm size emerges. It seems like size matters, if not for the mean outcome at least for the variability of outcome. A simple interpretation is that firms with high premium variation are riskier and need to pay a premium. A more speculative interpretation is that the variability is the effect of more correct pricing of heterogeneous labour and that there are increasing returns to efficiency so that there are rents to share. Both the

risk and rent interpretation can be used to explain the observed pattern of small firms exhibiting higher variability. However, an even simpler partial explanation may be that the observed pattern is the result of, in particular, natural turnover. The effect of a senior worker retiring and being replaced could be expected to be larger in a small firm.

No controls except the job classification are used when constructing the wage premium. To capture some of the intra-firm labor stock heterogeneity, the average age, proportion of women, and the proportion of young employees are entered into the set of explanatory variables. The model fitted includes a constant, firm size and turnover, the age, gender, and youth variables, dispersion, as well as the interaction of size and dispersion.

The sampler is initially set up to generate 50000 samples using first Algorithm A.2 and then direct sampling using the traditional griddy Gibbs sampler with a dynamic grid. The results using Algorithm A.2 are summarized in Figure 3.6 and Table 3.5. Studying the evolution of the mean estimate, the CUSUM statistic for the random effect variance exhibits model behavior with a quick and smooth convergence. The Geweke statistics are insignificant and there is only a very slight positive first order autocorrelation detected. Discarding 5000 samples as burnin, which should be more than enough, the kernel density estimate based on the retained samples is nicely and symmetrically placed over estimated mean. Turning to the table, all parameters are sampled efficiently; in particular and as expected this is true for the regression coefficients. Comparing with the results for direct sampling using the griddy Gibbs sampler in Figure 3.7 and Table 3.6 we note how the convergence of the random effect variance seems quicker, albeit more erratic. There is practically no autocorrelation in its sample path, however in the table we note how the auxiliary parameter actually sampled is done so slightly more inefficiently. Comparing posterior means and standard deviations, the choice of method seems irrelevant. Further, comparing execution times the slight advantage of data augmentation in the previous example is no more.

3.5 Model Selection

3.5.1 Extending the prior

Model selection is simply a matter of appropriate extension of the prior. One advantage of direct modeling is that the implementation is simplified when the covariance structure is controlled by just a few parameters rather than a possibly large number of latent variables.

Figure 3.6 Inference in the Labor panel using data augmentation and the improved 2-block algorithm. Select Markov chain properties for the random effect variance σ_μ^2 .

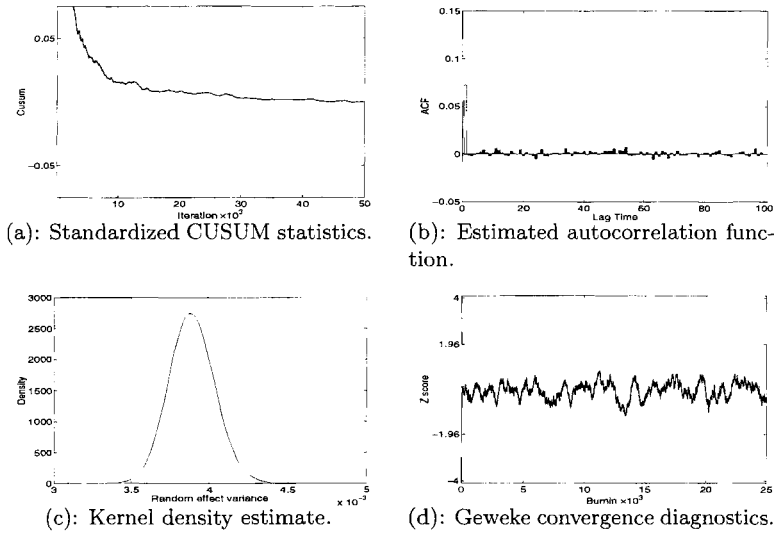


Table 3.5 Posterior simulation results for the Labor panel; data augmentation using the improved 2-block algorithm.

| | Mean ^b | S.d ^b | Mode ^{a,b} | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|-------------------|------------------|---------------------|----------|-----------------------|------|------------------|------|
| | | | | | NSE ^c | RNE | NSE ^c | RNE |
| β_0 | -6.765 | 1.063 | -6.757 | -0.00 | 4.752 | 1.00 | 4.715 | 0.98 |
| β_1 | -0.006 | 0.002 | -0.005 | -0.00 | 0.011 | 1.00 | 0.011 | 0.98 |
| β_2 | -0.004 | 0.003 | -0.004 | -0.00 | 0.015 | 1.00 | 0.015 | 1.00 |
| β_3 | 0.147 | 0.018 | 0.144 | -0.00 | 0.080 | 1.00 | 0.081 | 1.02 |
| β_4 | -2.790 | 0.511 | -2.805 | 0.01 | 2.313 | 1.02 | 2.312 | 1.02 |
| β_5 | -13.262 | 0.659 | -13.293 | 0.00 | 2.948 | 1.00 | 2.968 | 1.01 |
| β_6 | 56.438 | 20.726 | 56.740 | 0.00 | 91.856 | 0.98 | 91.242 | 0.97 |
| β_7 | 49.950 | 1.176 | 49.917 | 0.00 | 5.307 | 1.02 | 5.220 | 0.98 |
| σ_ν^2 | 0.110 | 0.023 | 0.112 | 0.09 | 0.113 | 1.20 | 0.113 | 1.20 |
| σ_μ^2 | 0.389 | 0.025 | 0.388 | 0.09 | 0.124 | 1.18 | 0.124 | 1.18 |

^a See notes in Table 3.1.

^b All values $\times 10^2$.

^c All values $\times 10^5$.

Figure 3.7 Inference in the Labor panel using direct sampling with griddy Gibbs. Select Markov chain properties for the random effect variance σ_μ^2 .

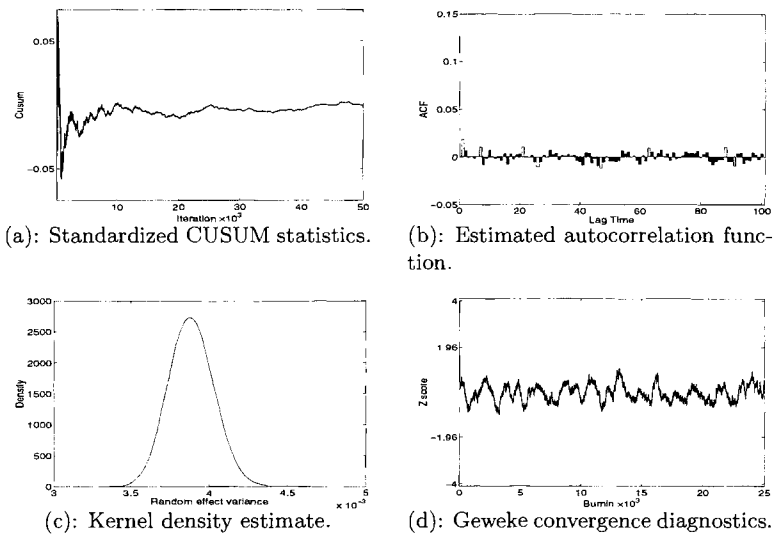


Table 3.6 Posterior simulation results for the Labor panel; direct sampling using the griddy Gibbs sampler.

| | Mean ^b | S.d. ^b | Mode, ^{ab} | ρ_1 | SPECTRAL ^a | | BATCH | |
|----------------|-------------------|-------------------|---------------------|----------|-----------------------|------|--------|------|
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -6.790 | 1.062 | -6.794 | -0.00 | 4.698 | 0.98 | 4.671 | 0.97 |
| β_1 | -0.006 | 0.002 | -0.006 | 0.00 | 0.011 | 0.98 | 0.011 | 1.00 |
| β_2 | -0.003 | 0.003 | -0.003 | -0.00 | 0.015 | 1.00 | 0.015 | 1.00 |
| β_3 | 0.147 | 0.018 | 0.145 | 0.00 | 0.079 | 0.98 | 0.080 | 0.99 |
| β_4 | -2.772 | 0.510 | -2.841 | -0.00 | 2.281 | 1.00 | 2.233 | 0.96 |
| β_5 | -13.245 | 0.657 | -13.311 | -0.00 | 2.937 | 1.00 | 2.921 | 0.99 |
| β_6 | 56.537 | 20.738 | 54.814 | 0.00 | 92.274 | 1.00 | 92.822 | 1.00 |
| β_7 | 49.880 | 1.157 | 49.810 | 0.01 | 5.174 | 1.00 | 5.144 | 0.99 |
| σ_ν^2 | 0.110 | 0.002 | 0.110 | 0.11 | 0.010 | 1.25 | 0.010 | 1.24 |
| σ_μ^2 | 0.389 | 0.014 | 0.388 | 0.03 | 0.066 | 1.05 | 0.066 | 1.04 |
| θ_μ | 2.763 | 0.116 2 | 768.0 | 12.5 | 582.1 | 27.0 | 576.1 | 24. |

^a See notes in Table 3.1.

^b All values $\times 100$.

To illustrate, consider the slightly more elaborate two-way random effects model,

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\beta + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T \\ \varepsilon_{it} &= \mu_i + \lambda_t + \nu_{it} \end{aligned}$$

where the error term ε_{it} now consists of three parts: a unit effect μ_i which is constant across time for a given unit i , a time effect λ_t which is common across units at a given date, and an idiosyncratic error ν_{it} . Assume that each of the random effects is independently and identically normally distributed,

$$\mu_i \sim \mathcal{N}(0, \sigma_\mu^2), \lambda_t \sim \mathcal{N}(0, \sigma_\lambda^2), \nu_{it} \sim \mathcal{N}(0, \sigma_\nu^2),$$

and also independent of each other as well as of \mathbf{x}_{it} . Here, the unknown parameters of the model are $\{\beta, \sigma_\nu^2, \sigma_\mu^2, \sigma_\lambda^2\}$ and, given the assumptions, the covariance matrix may be written as

$$\Sigma_\epsilon = \sigma_\nu^2 (I_N \otimes I_T) + \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\lambda^2 (J_N \otimes I_T).$$

Using a similar procedure Section 3.2.1, defining

$$\begin{aligned} \theta_\mu &\equiv (T\sigma_\mu^2\sigma_\nu^{-2} + 1)^{-1}, \theta_\lambda \equiv (N\sigma_\lambda^2\sigma_\nu^{-2} + 1)^{-1}, \\ 0 &< \theta_\mu, \theta_\lambda < 1 \end{aligned}$$

we may write the covariance matrix as

$$\begin{aligned} \Sigma_\epsilon &= \sigma_\nu^2 [M_1 + \theta_\mu^{-1}M_2 + \theta_\lambda^{-1}M_3 + (\theta_\mu^{-1} + \theta_\lambda^{-1} - 1) M_4] = \sigma_\nu^2 \Sigma_\theta, \\ M_1 &= E_N \otimes E_T, M_2 = E_N \otimes \bar{J}_T, M_3 = \bar{J}_N \otimes E_T, M_4 = \bar{J}_n \otimes \bar{J}_t. \end{aligned}$$

The coefficients are the characteristic roots, with multiplicities $(N-1)(T-1)$, $N-1$, $T-1$ and 1, and the M_i are idempotent and sum to the identity matrix. As a result, the inverse may be written in the desired form given by (3.5), where in terms of the quantities defined in Section 3.2.2

$$\begin{aligned} m &= 4, \theta = \{\theta_\mu, \theta_\lambda\} \\ g_1 &= 1, g_2 = \theta_\mu, g_3 = \theta_\lambda, g_4 = (\theta_\mu^{-1} + \theta_\lambda^{-1} - 1)^{-1}, Q_i = M_i. \end{aligned}$$

Further, a computationally convenient expression for the determinant is given by

$$|\Sigma_\epsilon| = \sigma_\nu^{2T} |\Sigma_\theta| = \sigma_\nu^{2T} \left[\theta_\mu^{-N} \theta_\lambda^{-T} (\theta_\mu + \theta_\lambda - \theta_\mu \theta_\lambda) \right],$$

so that, defining the appropriate coefficients $\{c_1, \dots, c_4\}$, evaluating the full conditional posterior which is on the form of (3.6) is simple.

For inference on the variance of the error components, ignoring again the regression coefficients β , we note how for $\theta_\lambda = 1$ the model reduces to that of the simple individual random effects model detailed in Section 3.2.1. Similarly, $\theta_\mu = 1$ implies no random individual effects. Allowing the possible exclusion of an effect, consider a mixed component priors on the form

$$\pi(\theta_\mu | \sigma_\nu^2) \sim w_\mu \cdot \pi(\theta_\mu | 0, 1) + (1 - w_\mu) \cdot I_{\theta_\mu=1},$$

where $\pi(\theta_\mu | 0, 1)$ a continuous distribution defined over the admissible region, w_μ a weight and $I_{\theta_\mu=1}$ the standard indicator function. Selecting $w_\mu = 1$ imposes the presence of the effect, however, any positive weight $w_\mu < 1$ allocates prior mass on the right boundary of the feasibility region and implies the possible absence of the effect. The priors may be convoluted resulting in an implied prior for the variance components on the form

$$\begin{aligned} \pi(\sigma_\mu^2, \sigma_\lambda^2 | \sigma_\nu^2) = & w_0 I_{\theta_\mu=1} I_{\theta_\lambda=1} + w_\mu \cdot \pi(\sigma_\mu^2 | \sigma_\nu^2) I_{\theta_\lambda=1} + w_\lambda \cdot \pi(\sigma_\lambda^2 | \sigma_\nu^2) I_{\theta_\mu=1} + \\ & + w_{\mu\lambda} \cdot \pi(\sigma_\mu^2 | \sigma_\nu^2) \pi(\sigma_\lambda^2 | \sigma_\nu^2). \end{aligned}$$

Here $\{w_0, w_\mu, w_\lambda, w_{\mu\lambda}\}$ is a collection of prior weights corresponding to models with no random effects, only an random unit effect, only a random time effect, and both random effects present. The mixed prior implies a mixed posterior, where the mixture will be over the models of interest, achieving the desired effect. Inference may then be conducted conditional or averaged over the type of model. Completing the prior specification any continuous components must be selected together with the prior model weights. As before, the choice of continuous prior components is not essential as long as they are proper. Lacking any prior information the weights are typically set equal, often interpreted as a reference uninformative choice.

3.5.2 Adjusting the posterior simulation

With the content of the parameter vector varying between the models corresponding to the components of the mixed prior, the reversible jump Metropolis-Hastings algorithm of Green (1995) is the appropriate tool. The implementation requires proper priors, a set of steps to update parameters conditional on the model, and a set of moves that allows the chain to go from one model to another. Assuming a conjugated normal inverse gamma prior structure for β and σ_ν^2 , and any prior on a structural parameter θ , a generic effect selection algorithm is easy to state.

A generic effects selection algorithm.

1. Conditional on $\{\sigma_\nu^2, \theta\}$, sample and accept a proposal for β from its full conditional posterior Step 1 of Algorithm 1.
2. Conditional on $\{\beta, \theta\}$, sample and accept a proposal for σ_ν^2 from its full conditional posterior Step 2 of Algorithm 1.
3. Conditional on σ_ν^2, β update θ . Querying the current status of the effect, attempt to change its status.
 - (a) If the selected effect is active attempt to delete it. Failing this, update it.
 - (b) If the selected effect is not active attempt to insert it.
4. Repeat steps 1-3 using the most recent values of the conditioning variables.

In the algorithm, Step (3) administrates both the exploration of a given model as well as the transition between model spaces. In (3.a) the updating of an effect surviving an attempted deletion is added to ensure that all effects present are updated in each iteration. While this adds to the overhead of the algorithm it also improves the mixing properties of the chain. Both effect updating and model space transitions may rely on the same approximation of the full conditional posterior. The former to generate updated values for a present effect conditional on the model and the latter to either propose candidate values when attempting to add an effect or balance existing effects when attempting to delete them.

For the moves associated with model space transitions it is necessary to compute an acceptance probability and, depending on it, sample a new state. If successful, the model index is then changed and the effect is added or deleted. The acceptance probability can always be stated in the standard form

$$\alpha = \min \{1, LR \times PR \times QR\},$$

where LR , PR , and QR a likelihood, prior, and proposal ratio respectively. Indexing the current state with c and the proposed state with p , the product of prior and proposal ratios for the addition of an effect will always be on the form

$$PR \times QR = \frac{w_p}{w_c} \pi(\theta_p | \sigma_\nu^2) \frac{j_{pc}}{j_{cp} \cdot q(\theta_p)},$$

where $q(\theta_p)$ the constructed approximation of the full conditional posterior and j_{st} the probability of proposing a move that attempts to move the chain from a model s to a model t . What does separate the cases is the extent to which the likelihood ratio simplifies. For example, when adding a unit random effect in the absence of any time effect the relevant ratio simplifies to

$$LR = \theta_\mu^{N/2} \exp \{ (c_1 + c_3) (\theta_\mu - 1) \},$$

but when the time effect is present the correct ratio is

$$LR = \left(\frac{\theta_\mu^N}{\theta_\mu + \theta_\lambda - \theta_\mu \theta_\lambda} \right)^{1/2} \exp \left\{ \left[c_1 + c_3 \frac{\theta_\lambda^2}{(\theta_\lambda + \theta_\mu - \theta_\mu \theta_\lambda)} \right] (\theta_\mu - 1) \right\}.$$

Similarly, when adding a time effect in the absence of a unit effect

$$LR = \theta_\lambda^{T/2} \exp \{ (c_2 + c_3) (\theta_\lambda - 1) \},$$

but when the unit effect is present

$$LR = \left(\frac{\theta_\lambda^T}{\theta_\mu + \theta_\lambda - \theta_\mu \theta_\lambda} \right)^{1/2} \exp \left\{ \left[c_2 + c_3 \frac{\theta_\mu^2}{(\theta_\lambda + \theta_\mu - \theta_\mu \theta_\lambda)} \right] (\theta_\lambda - 1) \right\}.$$

These results rely on the Q_i being symmetric, idempotent, pairwise orthogonal and, in particular, summing to the identity matrix.

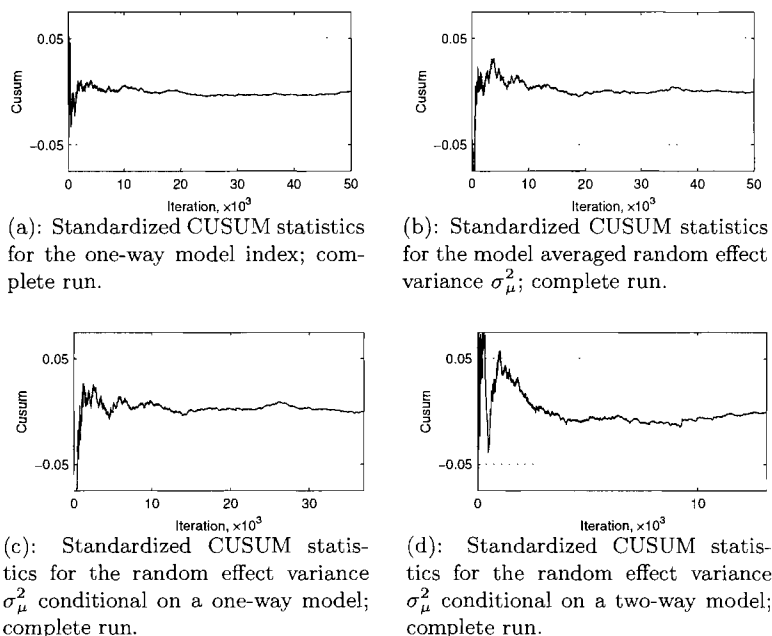
For the inverse moves that attempts to delete the effect, the acceptance probability is just the inverse where the current value of the relevant effect is treated as if it is sampled from the constructed proposal distribution.

3.5.3 A simple illustration

The Grunfeld panel from Section 3.4.1 is used to illustrate effect selection. Setting all weights equal, the sampler is run 50000 iterations and updating is performed using the griddy Gibbs sampler. To be able to compare any results for the simple random individual effect model with those from Section 3.4.1 the same prior parameters are used. The time effect parameters are set equal to the other effect parameters.

One problem that occurs with model selection is that it is not longer clear what and how to monitor convergence. Conditioning on the model index, which is what we typically want to do, quantities such as estimated autocorrelations may be misleading as a gauge of performance. This as a sampler mixing over the model indexes will act as an indirect thinning process. A natural parameter to monitor is the model index which can be hard to pin down

Figure 3.8 Standardized CUSUM plots for the posterior probability of a simple one-way model and various sections of the random effects variance chain.



with precision. In Figure 3.8, CUSUM plots for the model index, the random effect variance conditional on model index and the random effect model averaged are illustrated. The idea is that if the chains conditional on the index as well as the model index probabilities behave well then so should also the model averaged chain. This seems to be the case, possibly with the exception of the output conditional on a two-way specification. However, considering that it is based on the fewest observation in the lot, this is perhaps to be expected. That the model index converges smoothly and reasonably fast is particularly gratifying.

Studying the transition probabilities, the sampler seems to be moving between the one- and two-way specifications at a healthy rate. While the survival probability of a two-way model is low this is countered by a sizeable transition probability from the one- to the two-way model. With a posterior probability of 74%, the evidence for the one-way specification is strong without being overwhelming. In Table 3.7, the results conditional on the one-way model are

Table 3.7 Posterior simulation results for the Grunefeld panel; direct sampling using the griddy Gibbs sampler and featuring model selection.

| $P(\theta_\mu < 1 \wedge \theta_\lambda = 1) = 0.74$ | | | | | | | | |
|--|----------|----------|-------------------|----------|-----------------------|------|---------------------|------|
| | Mean | S.d | Mode ^a | ρ_1 | SPECTRAL ^a | | BATCH | |
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -60.668 | 28.861 | -59.421 | 0.01 | 0.131 | 1.03 | 0.151 | 1.02 |
| β_1 | 0.109 | 0.010 | 0.109 | -0.01 | 0.0 ⁴ 45 | 0.97 | 0.0 ⁴ 53 | 0.97 |
| β_2 | 0.308 | 0.017 | 0.309 | 0.00 | 0.0 ⁴ 78 | 1.02 | 0.0 ⁴ 91 | 1.03 |
| σ_ν^2 | 2806.512 | 293.001 | 2783.182 | 0.04 | 1.360 | 1.08 | 1.553 | 1.04 |
| σ_μ^2 | 7308.590 | 4072.257 | 5149.746 | 0.06 | 19.272 | 1.12 | 22.534 | 1.13 |
| $P(\theta_\mu < 1 \wedge \theta_\lambda < 1) = 0.26$ | | | | | | | | |
| | Mean | S.d | Mode ^a | ρ_1 | SPECTRAL ^a | | BATCH ^a | |
| | | | | | NSE | RNE | NSE | RNE |
| β_0 | -63.216 | 29.279 | -60.189 | -0.00 | 0.131 | 1.00 | 0.257 | 1.01 |
| β_1 | 0.110 | 0.011 | 0.109 | -0.01 | 0.0 ⁴ 49 | 1.05 | 0.0 ⁴ 93 | 0.98 |
| β_2 | 0.314 | 0.018 | 0.312 | -0.00 | 0.0 ⁴ 84 | 1.04 | 0.0 ³ 16 | 1.00 |
| σ_ν^2 | 2749.408 | 292.844 | 2703.463 | 0.02 | 1.334 | 1.04 | 2.596 | 1.03 |
| σ_μ^2 | 7431.542 | 4555.323 | 4986.550 | 0.00 | 20.372 | 1.00 | 39.497 | 0.98 |
| σ_λ^2 | 88.978 | 51.582 | 48.069 | 0.04 | 0.245 | 1.12 | 0.468 | 1.07 |

^a See notes in Table 3.1.

consistent with the results obtained in Section 3.4.1. The transition between specifications hardly affects the parameters in the mean model and results only in slight shifts in the variance components. As calculated, the relative numerical efficiencies cannot be interpreted the usual way. Arguably, their near perfect values reflect that the chain moves between models in an unpredictable way. In general, the results are in line with some of the classical estimates reported in Baltagi (1995). Model averaged the results are similar to those obtained using iterated maximum likelihood to estimate a two-way model.

3.6 Final Remarks

This paper presents a direct sampling method for inference in panel regression models with parametric covariance structures. Developing the tools necessary for posterior simulation, the illustrated examples of direct sampling with and without effect selection offered many valuable insights. When practical, direct sampling is competitive, works well, is reasonably fast and can be relied on

to produce the required inference. It offers the opportunity to re-examine models in search of exciting alternatives outside the mainstream. A good example of that is the Metropolis-Hastings with truncated gamma proposals tested for the one-way model which did something so rare as to combine the best of two worlds; the speed of Algorithm A.1 with the efficiency of Algorithm A.2. Avoiding data augmentation, direct sampling can be quite economical, in particular when the panel size grows. It also offers a greater sense of control, making for instance model selection simpler.

There are several technical issues relating to the posterior simulation that influence the performance, the main issue being how to sample the variance components. In particular, how to construct reliable grids in an efficient manner are important questions. There is ample room for improvement on both counts. Experimentation showed, as expected, that adaptive grids enhanced the performance in terms of numerical efficiency at the cost of fewer iterations per time unit. However, placing nodes in a clever way is time consuming and after some point any clever procedure may end up defeating itself. Finding ways to combine quick and cheap strategies with slow, more elaborate but better methods would lead to quicker samplers. Also, finding means to generate and administrate multidimensional grids efficiently offers an interesting challenge with potentially high rewards. Other technical issues include the efficient parametrization of the models. How it is done should depend on what kind of quantities we are willing to make statements about a priori. Still, the methods outlined and the tools developed do not depend on the prior, at least not as long as the necessary structure is preserved.

While the theory for implementing model selection is standard, the application of the method revealed some practical problems. In particular, the priors on the variance components play an important role and being vague, but proper, is difficult. The mindless application of the standard more or less automatic choices, can lead to strange inference. While falling outside the scope of this paper, finding reasonable default priors should be a priority. Another interesting but more general question is when, if ever, posterior simulation should perhaps be abandoned in favor of classical numerical integration techniques. This because, in the end, doing things only because we can is usually a bad idea.

Bibliography

- Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons, Chichester.
- Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, John Wiley & sons, inc, Chichester.
- Chib, S. and Carlin, B. P. (1999), ‘On MCMC sampling in hierarchical longitudinal models’, *Statistics and Computing* **9**, 17–26.
- Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995), ‘Efficient parametrizations for generalized linear mixed models’, *Biometrika* **82**, 479–488.
- Gelfand, A. E. and Smith, A. F. M. (1990), ‘Sampling-based approaches to calculating marginal densities’, *Journal of the American Statistical Association* **85**, 398–409.
- Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, ‘Bayesian Statistics 4’, Oxford University Press, Oxford, pp. 169–93.
- Gilks, W. R. and Roberts, G. O. (1996), Strategies for improving MCMC, in W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, ‘Markov Chain Monte Carlo in Practice’, Chapman Hall, London, pp. 89–114.
- Green, P. J. (1995), ‘Reversible jump markov chain monte carlo computation and bayesian model determination’, *Biometrika* **82**, 711–732.
- Grunfeld, Y. (1958), The Determinants of Corporate Investment, PhD thesis, University of Chicago.
- Hill, B. M. (1965), ‘Inference about variance components in the one-way model’, *Journal of the American Statistical Association* **60**, 806–825.

- Ritter, C. and Tanner, M. A. (1992), 'Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler', *Journal of the American Statistical Association* **87**, 861–868.
- Tanner, M. A. and Wong, W. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* pp. 528–550.
- Tiao, G. C. and Tan, W. Y. (1965), 'Bayesian analysis of random-effects models in the analysis of variance. I: Posterior distribution of variance components', *Biometrika* **52**, 37–53.
- Vines, S. K., Gilks, W. R. and Wild, P. (1996), 'Fitting bayesian multiple random effects models', *Statistics and Computing* **6**, 337–346.
- Yu, B. and Mykland, P. (1998), 'Looking at Markov samplers through cusum path plots: A simple diagnostic idea', *Statistics and Computing* **8**(3), 275–286.

Appendix A

Two algorithms

A.1 The simple multiple block algorithm

Labeled A1 in Chib and Carlin (1999).

1. Sample β from

$$p(\beta | y, \{\mu_i\}, \sigma_\varepsilon^2, \sigma_\mu^2) \sim \mathcal{N}(\hat{\beta}, \Omega^{-1}), \quad (\text{A.1})$$

where

$$\Omega = \left(\Omega_0 + \sigma_\varepsilon^{-2} \sum_{i=1}^n X_i' X_i \right), \hat{\beta} = \Omega^{-1} \left(\sigma_\varepsilon^{-2} \sum_{i=1}^n X_i' \tilde{\mathbf{y}}_i \right), \tilde{\mathbf{y}}_i = \mathbf{y}_i - \mu_i.$$

2. Sample $\mu = \{\mu_i\}$ independently from

$$p(\mu | y, \beta, \sigma_\varepsilon^2, \sigma_\mu^2) \sim \mathcal{N}(\hat{\mu}_i, \omega^{-1}), \quad (\text{A.2})$$

where

$$\omega = (\omega_0 + \sigma_\mu^{-2} n), \hat{\beta} = \omega^{-1} (\sigma_\mu^{-2} X_i' \tilde{\mathbf{y}}_i), \tilde{\mathbf{y}}_i = \mathbf{y}_i - X_i \beta.$$

3. Sample σ_ε^2 from

$$p(\sigma_\varepsilon^{-2} | y, \beta, \mu, \sigma_\mu^2) \sim \mathcal{G}(a_\varepsilon + nt/2, b_\varepsilon + S/2) \quad (\text{A.3})$$

where

$$S = \sum_{i=1}^n (\mathbf{y}_i - X_i \beta - \mu_i)' (\mathbf{y}_i - X_i \beta - \mu_i).$$

4. Sample σ_μ^2 from

$$p(\sigma_\mu^{-2} | y, \beta, \mu, \sigma_\varepsilon^2) \sim \mathcal{G}(a_\mu + n/2, b_\mu + \mu' \mu) / 2. \quad (\text{A.4})$$

5. Repeat steps 1-4 using the most recent values of the conditioning variables.

A.2 An improved multiple block algorithm

Labeled A2 Chib and Carlin (1999).

1. Sample β and μ from $p(\beta, \{\mu_i\} | y, \sigma_\varepsilon^2, \sigma_\mu^2)$ by sampling

(a) β from $p(\beta | y, \sigma_\varepsilon^2, \sigma_\mu^2)$ (3.7).

(b) μ from $p(\mu | y, \beta, \sigma_\varepsilon^2, \sigma_\mu^2)$ (A.2).

2. Sample σ_ε^2 from $p(\sigma_\varepsilon^2 | y, \beta, \mu, \sigma_\mu^2)$ (A.3).

3. Sample σ_μ^2 from $p(\sigma_\mu^2 | y, \beta, \sigma_\varepsilon^2, \mu)$ (A.4).

4. Repeat steps 1-3 using the most recent values of the conditioning variables.

Paper 4

Serial Correlation and Random Effects in Bayesian Panel Regression

4.1 Introduction

Following the influential work of Lillard and Willis (1978) there has been a continued interest in random effects models which allow for dynamics in the form of a serially correlated error term. Similar to Lillard and Willis, Anderson and Hsiao (1982), MaCurdy (1982) and Baltagi and Li (1991, 1994) consider a one-way error component model with individual specific effects and serially correlated idiosyncratic errors.

This paper proposes novel Bayesian procedures for analyzing panel data models with serial correlation. Models of this type would typically be treated using data augmentation as a facilitating component in a Markov chain Monte Carlo approach. Using the approach of detailed in Paper 3 it is shown how data augmentation in general can be partly avoided and in some important special cases completely eliminated. This allows us to avoid conditioning on starting values when modelling the serial correlation and the full likelihood can be used without simulating latent initial values. In addition, it is well known that data augmentation can lead to slow convergence of the Markov chain in this type of models and a direct approach to sampling the parameters of the variance matrix offer the possibility of greater computational efficiency.

The remainder of the paper proceeds as follows. Section 4.2, treats the pure serial correlation model and offers two simple algorithms for the approximation of the posterior using Markov chain Monte Carlo methods. Section 4.3 introduces random effects in the context of a more general mixed effects model and discusses the special case of individual random effects where the analysis is relatively simple without performing data augmentation. One of the important features of parametric covariance models is the simplicity with which model uncertainty can be accounted for. This is exemplified in Section 4.4 for the simpler pure serial correlation model. Finally, Section 4.5 discusses our experiences to date and offer some ideas for further research.

4.2 Panel regression with serial correlation

Let the response y_{it} of individual i at time t be the sum of a mean $\mathbf{x}_{it}\beta$ and an autoregressive error term ϵ_{it}

$$\begin{aligned}y_{it} &= \mathbf{x}_{it}\beta + \epsilon_{it}, \\ \epsilon_{it} &= \rho\nu_{it-1} + \nu_{it}.\end{aligned}$$

The mean depends on a vector of p explanatory variables \mathbf{x}_{it} with associated fixed effects β , common across all units. The autoregressive process is

restricted to be strictly stationary, so that $|\rho| < 1$, and the errors ν_{it} are assumed to be mutually independent normal, $\nu_{it} \sim \mathcal{N}(0, \sigma_\nu^2)$. Assuming T dates are sampled and $\lambda = \sigma_\nu^{-2}$ the precision, the covariance matrix for the collection of measurements on any unit may be written as

$$\Sigma = [\lambda (1 - \rho^2)]^{-1} \mathbf{M} = \lambda^{-1} \mathbf{M}_\rho,$$

where \mathbf{M} is a symmetric $T \times T$ Toeplitz matrix defined by the $T \times 1$ vector $(\rho^0, \rho^1, \dots, \rho^{T-1})$ with elements $m_{it} = \rho^{|i-t|}$ for $i, t = 1, \dots, T$. Assuming the number of individuals sampled is N , the likelihood is proportional to

$$L(\mathbf{y}, \mathbf{X}, \beta, \lambda, \rho) \propto |\lambda^{-1} \mathbf{M}_\rho|^{-N/2} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \beta)' \mathbf{M}_\rho^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta) \right\} \quad (4.1)$$

where for unit i the vector \mathbf{y}_i contains the measurements of the response variable and the matrix \mathbf{X}_i the stacked explanatory variables.

The unknown parameters of the model are collected in $\theta = \{\beta, \lambda, \rho\}$ and adopting the prior structure

$$\pi(\theta) = \pi(\beta | \lambda) \pi(\rho) \pi(\lambda),$$

the model lends itself to full Bayesian analysis by Markov chain Monte Carlo methods. In particular, ignoring proportionality factors we have for the fixed effects β conditional on λ and ρ and the precision λ conditional on ρ and β ,

$$p(\beta | \mathbf{y}, \mathbf{X}, \lambda, \rho) \propto \exp \left\{ -\frac{\lambda}{2} (\beta - b)' B (\beta - b) \right\} \pi(\beta | \lambda),$$

$$p(\lambda | \mathbf{y}, \mathbf{X}, \beta, \rho) \propto \lambda^{NT} \exp \{-s_e \lambda\} \pi(\lambda),$$

where

$$B = \sum_{i=1}^N \mathbf{X}_i \mathbf{M}_\rho^{-1} \mathbf{X}_i, \quad b = B^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{M}_\rho^{-1} \mathbf{Y}_i,$$

$$s_e = \frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \beta)' \mathbf{M}_\rho^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta),$$

so that the usual conjugate prior choices lead to normal and gamma full conditional posteriors respectively. For the autoregressive parameter ρ conditional on β and λ things are only slightly more complicated. Use that the determinant of the intra-unit covariance matrix is simply given by

$$|\lambda^{-1} \mathbf{M}_\rho| = \lambda^{-T} (1 - \rho^2)^{-1}, \quad (4.2)$$

and that some matrix algebra yields the inverse as

$$\mathbf{M}_\rho^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} = \sum_{i=0}^2 \rho^i \mathbf{M}_i, \quad (4.3)$$

for the obvious definitions of the $T \times T$ matrices \mathbf{M}_i . Then, again ignoring a proportionality factor, the full conditional posterior is just

$$p(\rho | \mathbf{y}, \mathbf{X}, \beta, \lambda) \propto (1 - \rho^2)^{N/2} \exp \left\{ -\frac{1}{2} \sum_{i=0}^2 c_i \rho^i \right\} \pi(\rho), \quad (4.4)$$

$$c_i = \lambda \sum_{j=1}^N \mathbf{e}_j' \mathbf{M}_i \mathbf{e}_j, \quad \mathbf{e}_j = \mathbf{y}_j - \mathbf{X}_j \beta.$$

In a univariate context, that is $N = 1$, Fomby and Guilkey (1978) referring to Zellner and Tiao (1964) and Zellner (1971) use a Jeffreys prior which is of the beta form with parameters (0.5, 0.5) over the range $|\rho| < 1$, resulting in a normal full conditional posterior. Adapting this to the panel setting we obtain

$$\pi(\rho) \propto (1 - \rho^2)^{-N/2}, \quad (4.5)$$

the factor in front of the exponential cancels and the full conditional posterior simplifies to a truncated normal,

$$p(\rho | \mathbf{y}, \mathbf{X}, \beta, \lambda) \propto \exp \left\{ -\frac{1}{2s_r^2} (\rho - r)^2 \right\}, \quad (4.6)$$

for

$$r = -\frac{c_1}{2c_2}, \quad s_r = \sqrt{\frac{1}{c_2}}.$$

Completing the specification with the independent conditional conjugate priors

$$\pi(\beta) \sim \mathcal{N}_p(\beta_0, B_0), \quad \pi(\lambda) \sim \mathcal{G}(\delta_1, \delta_2), \quad (4.7)$$

then leads to a particularly simple updating scheme.

No effects exponential errors model

1. Conditional on λ and ρ , sample and accept a proposal for β from

$$\beta | \lambda, \rho \sim \mathcal{N}_p(b, S),$$

where

$$S = [B_0^{-1} + \lambda B]^{-1}, \quad b = S \left(B_0^{-1} \beta_0 + \lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_\rho^{-1} \mathbf{y}_i \right).$$

2. Conditional on β and ρ , sample and accept a proposal λ from its full conditional posterior. The conjugate structure results in

$$\lambda | \beta, \rho \sim \mathcal{G}(d_1, d_2),$$

where

$$d_1 = \delta_1 + NT/2, \quad d_2 = \delta_2 + s_e.$$

3. Conditional on λ and β sample ρ using the full conditional posterior in (4.6).
4. Iterate steps 1-3 until convergence or a stopping criterion is met.

Leading as it does to a simple full conditional posterior, from a pragmatic point of view the prior in (4.5) is very attractive. The fact that the use of a Jeffreys prior on ρ leads to a convenient full conditional posterior has not been appreciated in the literature and we note that this leads to very efficient simulations in the standard univariate regression model with AR(1) errors.

Unfortunately, its shape and behavior when N increases makes it implausible and as such hard to justify. However, even if this prior is rejected, the results above are still very useful. The crucial point is the means to effortlessly construct the inverse and also compute the determinant leading the likelihood. As in the previous paper, with a simple and computationally cheap general expression for the full conditional posterior it is possible to construct a first order approximation of the conditional posterior. This approximation may then be used either to retain the overall Gibbs structure, employing the griddy Gibbs sampler of Ritter and Tanner (1992), or to generate good candidate values to be used in Metropolis type updates. Regardless of method, the results in (4.6) can be used to help place the grid where it matters, thus improving the quality

of the approximation. The posterior simulation is then carried out much in the same way as in Algorithm 4.2. Thus, assuming for instance

$$\pi(\rho) \sim U(-1, 1),$$

and using the Metropolis-Hastings algorithm, the following modified updating scheme may be used.

No effects exponential errors model, modified.

1. Conditional on λ and ρ , sample β as in Algorithm 4.2.
2. Conditional on β and ρ , sample λ as in Algorithm 4.2.
3. Conditional on λ and β sample ρ using the full conditional posterior in (4.4) construct a first order approximation $q_\rho(\rho)$. Indexing the current and candidate values with c and p respectively, the acceptance probability simplifies to

$$\begin{aligned} \alpha &= \min \left\{ 1, \left(\frac{1 - \rho_p^2}{1 - \rho_c^2} \right)^{N/2} \frac{\exp \left\{ -\frac{1}{2} \sum_{i=0}^2 c_i \rho_p^i \right\} q_\rho(\rho_c)}{\exp \left\{ -\frac{1}{2} \sum_{i=0}^2 c_i \rho_c^i \right\} q_\rho(\rho_p)} \right\} \\ &= \min \left\{ 1, \kappa \exp \left\{ -\frac{1}{2} (\rho_p - \rho_c) [c_1 + c_2 (\rho_p + \rho_c)] \right\} \right\} \\ \kappa &= \left(\frac{1 - \rho_p^2}{1 - \rho_c^2} \right)^{N/2} \frac{q_\rho(\rho_c)}{q_\rho(\rho_p)} \end{aligned}$$

Depending on the quality of the approximation, α_ρ is more or less close to one.

4. Iterate steps 1-3 until convergence or a stopping criterion is met.

4.3 The mixed effects model with serial correlation

The set up in the previous section can easily accommodate the popular fixed individual effects model either via the inclusion of individual dummy variables or the within transformation. Turning to random effects models we consider this in the more general set up of a mixed effects model. In the Bayesian paradigm this amounts to adding a level with unit effects to the model, where the associated prior includes parameters to estimate. The model represents

the response y_{it} of unit i at time t as the sum of a mean $\mathbf{x}_{it}\beta$, a unit effect $\mathbf{z}_{it}\gamma_i$ and an autoregressive error term ν_{it}

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\beta + \mathbf{z}_{it}\gamma_i + \epsilon_{it}, \\ \epsilon_{ij} &= \rho\nu_{it-1} + \nu_{it}. \end{aligned}$$

The unit effects $\gamma_i \sim \mathcal{N}_q(0, \mathbf{\Omega})$ operate on q explanatory variables \mathbf{z}_{ij} . The design matrices \mathbf{Z} and \mathbf{X} may share elements, and typically q is smaller than p . At one extreme, a random coefficients model is obtained with $\mathbf{Z} = \mathbf{X}$ and at the other we have a random individual effects model with \mathbf{Z} a matrix of individual dummy variables.

As before, there are p explanatory variables with associated effects β , common across all units, the autoregressive process is restricted to be strictly stationary, and the errors ν_{it} are assumed to be mutually independent normal, $\nu_{it} \sim \mathcal{N}(0, \lambda^{-1})$. The likelihood in (4.1) is adjusted in the obvious way,

$$L(\mathbf{y}, \mathbf{X}, \beta, \lambda, \rho, \mathbf{\Omega}) \propto \prod_{i=1}^N |\mathbf{\Sigma}_i|^{-1/2} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i\beta)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta) \right\}$$

for $\mathbf{\Sigma}_i = \lambda^{-1}\mathbf{V}_i = \lambda^{-1}\mathbf{M}_\rho + \mathbf{Z}_i\mathbf{\Omega}\mathbf{Z}_i'$.

Adding a prior for $\mathbf{\Omega}$ to the prior structure yields

$$p\left(\beta, \{\gamma_i\}_{i=1}^N, \lambda, \rho, \mathbf{\Omega}\right) = p(\beta) p\left(\{\gamma_i\}_{i=1}^N\right) p(\lambda) p(\rho) p(\mathbf{\Omega}),$$

where

$$\begin{aligned} \pi(\mathbf{\Omega}^{-1}) &\sim \mathcal{W}(v_0^{-1}\mathbf{W}_0, v_0), \pi(\rho) \sim U(-1, 1), \\ \pi(\beta) &\sim \mathcal{N}_p(\beta_0, B_0), \pi(\lambda) \sim \mathcal{G}(\delta_1, \delta_2). \end{aligned}$$

Attempting inference without recourse to data augmentation turns out to be quite complicated and costly due to problems sampling ρ . On the upside, sampling, augmenting and conditioning on the unit effects allows simulation of ρ using the procedure already developed in Section 4.2. Indiscriminate implementation of data augmentation, as detailed and discussed in Tanner and Wong (1987), may result in a Markov chain with severe convergence problems. This is generally true for random effects specifications and substantial effort has been exerted to refine the algorithms used. Chib and Carlin (1999) offer an overview and also present two of the best procedures available to date.

The suggested procedure is based on and extends their simpler Algorithm 2 to include serial correlation.

The approach relies on the conditional Gaussian structure yielding the conditional distribution of \mathbf{y}_i

$$\mathbf{y}_i | \beta, \{\gamma_i\}_{i=1}^N, \lambda, \rho, \mathbf{\Omega} \sim \mathcal{N}_t(\mathbf{X}_i\beta + \mathbf{Z}_i\gamma_i, \lambda^{-1}\mathbf{M}_\rho), \quad (4.8)$$

to sample the unit random effects and the same marginalized over the random effects

$$\mathbf{y}_i | \beta, \lambda, \rho, \mathbf{\Omega} \sim \mathcal{N}_t(\mathbf{X}_i\beta, \lambda^{-1}\mathbf{V}_i), \quad (4.9)$$

to sample the regression coefficients β . This marginalization has proven essential for the quality of the Markov chain, leading to substantial gains in relative numerical efficiency, see Chib and Carlin (1999). Approximating the posterior, the sampler iterates four simple steps.

General random effects exponential errors model

1. Conditional on λ , ρ , and $\mathbf{\Omega}$, sample and accept proposals for β and $\{\gamma_i\}_{i=1}^N$ from the appropriate full conditional posteriors.

- (a) Combining the marginalized conditional likelihood (4.9) for all units with the prior gives

$$\beta | \lambda, \rho, \mathbf{\Omega} \sim \mathcal{N}_p(b, B),$$

where

$$B = \left[B_0 + \lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right]^{-1}, \quad b = B \left(B_0 \beta_0 + \lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right),$$

- (b) Combining the conditional likelihood in (4.8) with respective prior component, cycle through units and sample proposals from

$$\gamma_i | \beta, \lambda, \rho, \mathbf{\Omega} \sim \mathcal{N}_q(\mathbf{g}_i, \mathbf{G}_i^{-1}),$$

where

$$\mathbf{G}_i = \mathbf{\Omega}^{-1} + \lambda \mathbf{Z}_i' \mathbf{M}_\rho^{-1} \mathbf{Z}_i, \quad \mathbf{g}_i = \mathbf{G}_i^{-1} (\mathbf{Z}_i' \mathbf{M}_\rho^{-1} (\mathbf{y}_i - \mathbf{X}_i' \beta)).$$

2. As specified, all the information on Ω contained in the sample is passed down through the collection of unit random effects $\{\gamma_i\}_{i=1}^N$. It follows that the relevant conditioning is $\{\gamma_i\}_{i=1}^N$ and Ω is sampled from,

$$\Omega^{-1} | \lambda, \{\gamma_i\}_{i=1}^N \sim \mathcal{W}(\mathbf{W}_1, v_1),$$

where

$$\mathbf{W}_1^{-1} = v_0^{-1} \mathbf{W}_0^{-1} + \sum_{i=1}^N \gamma_i \gamma_i', \quad v_1 = v_0 + N.$$

3. Conditional on β , $\{\gamma_i\}_{i=1}^N$, ρ sample λ from

$$\lambda | \beta, \{\gamma_i\}_{i=1}^N, \rho \sim \mathcal{G}(d_1, d_2),$$

where

$$d_1 = \delta_1 + NT/2, \quad d_2 = \delta_2 + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \gamma_i)' \mathbf{M}_\rho^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \gamma_i) / 2.$$

4. Conditional on β , $\{\gamma_i\}_{i=1}^N$, and λ sample ρ as in Section 4.2 with

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{Z}_j \gamma_j.$$

5. Iterate steps 1-4 until convergence or a stopping criterion is met.

While we typically need to use data augmentation to make progress there is one interesting case when it can be avoided without having to recourse to brute force inversion of the covariance matrix. In the random individual effects model \mathbf{Z}_i is just a dummy variable for individual i and Ω reduces to a single element. The model is conventionally written as

$$y_{it} = \mathbf{x}_{it} \beta + \mu_i + \epsilon_{it}$$

with $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$ and ϵ_{it} as before. The covariance matrices for the individuals are identical and simplifies to

$$\Sigma = \lambda^{-1} \mathbf{V} = \sigma_\mu^2 \mathbf{J}_T + \lambda^{-1} \mathbf{M}_\rho$$

for $\mathbf{J}_T = \mathbf{j}_T \mathbf{j}_T'$ a matrix of ones. A convenient expression for the inverse is given by Baltagi and Li (1991). Consider the Prais-Winsten transformation matrix \mathbf{C}

$$\mathbf{C}^{-\top} \mathbf{C}^{-1} = \mathbf{M}_\rho.$$

Applying it to \mathbf{V} we obtain

$$\mathbf{V} = \lambda \sigma_\mu^2 \mathbf{J}_t + \mathbf{M}_\rho = \mathbf{C}^{-\top} (\lambda \sigma_\mu^2 \mathbf{C}' \mathbf{J}_t \mathbf{C} + \mathbf{I}) \mathbf{C}^{-1} = \mathbf{C}^{-\top} (\theta^{-1} \bar{\mathbf{J}}^\alpha + \bar{\mathbf{E}}^\alpha) \mathbf{C}^{-1},$$

for

$$\bar{\mathbf{J}}_T^\alpha = \mathbf{j}_T^{\alpha'} \mathbf{j}_T^{\alpha'} / d^2, \mathbf{j}_T^{\alpha'} = \mathbf{j}_T' \mathbf{C} = (\alpha, \mathbf{j}_{T-1}')^{\alpha'}$$

and

$$\bar{\mathbf{E}}_T^\alpha = \mathbf{I}_T - \bar{\mathbf{J}}_T^\alpha$$

with

$$d^2 = \mathbf{j}_T^{\alpha'} \mathbf{j}_T^\alpha = \alpha^2 + T - 1, \alpha = \sqrt{(1 + \rho_v) / (1 - \rho_v)}$$

and

$$\theta^{-1} = d^2 (1 - \rho)^2 \lambda \sigma_\mu^2 + 1.$$

The inverse is then given by

$$\mathbf{V}^{-1} = \mathbf{C} (\theta \bar{\mathbf{J}}^\alpha + \bar{\mathbf{E}}^\alpha) \mathbf{C}' = \mathbf{C} [(\theta - 1) \bar{\mathbf{J}}^\alpha + \mathbf{I}] \mathbf{C}' = \frac{\theta - 1}{d^2} \mathbf{M}_\rho^{-1} \mathbf{J} \mathbf{M}_\rho^{-1} + \mathbf{M}_\rho^{-1}. \quad (4.10)$$

Let

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \beta \quad \mathbf{e}_i^* = (e_{i2}, \dots, e_{iT-1}),$$

the exponent in the likelihood can then be written as

$$\begin{aligned} & -\frac{\lambda}{2} \sum_{i=1}^N \mathbf{e}_i' \left(\frac{\theta - 1}{d^2} \mathbf{f} \mathbf{f}' + \mathbf{M}_\rho^{-1} \right) \mathbf{e}_i \\ & = -\frac{\lambda(\theta - 1)}{2d^2} \left[\begin{aligned} & (1 - \rho)^2 \sum_{i=1}^N (e_{i1}^2 + e_{iT}^2)^2 \\ & + 2(1 - \rho)^3 \sum_{i=1}^N (e_{i1}^2 + e_{iT}^2) (\mathbf{e}_i^* \mathbf{j}_{T-2}) \\ & + (1 - \rho)^4 \sum_{i=1}^N (\mathbf{e}_i^* \mathbf{j}_{T-2})^2 \end{aligned} \right] + \sum_{j=0}^2 \rho^j c_j, \end{aligned} \quad (4.11)$$

where

$$\mathbf{f} = \mathbf{M}_\rho^{-1} \mathbf{j}_T = (1 - \rho, (1 - \rho)^2, \dots, (1 - \rho)^2, 1 - \rho)'$$

In addition, the determinant of \mathbf{V} is given by $|\mathbf{V}| = (1 - \rho^2)^{-1} \theta^{-1}$ since θ^{-1} and 1 are the characteristic roots with multiplicity 1 and $T - 1$ of $\theta^{-1} \bar{\mathbf{J}}^\alpha + \bar{\mathbf{E}}^\alpha$.

The determinant and (4.11) are simple functions of ρ and σ_μ^2 and can be used to tabulate both the full conditional posteriors of both ρ and σ_μ^2 in order to construct a first order approximation of the posterior. Extending Algorithm 4.3 to allow for random individual effects is thus a simple matter of replacing step 3 with two steps that sample ρ and σ_μ^2 using either the griddy Gibbs sampler or Metropolis type updates with the approximation as the proposal density.

4.4 Model Selection

The models discussed presents various model selection possibilities. Certainly the design matrix is one dimension where an aspect of model selection may be included. Reinterpreting the fixed effects variables as possible explanatory variables, the status of each may be labeled as active or not. The question of random effects or not offers particularly rich model selection possibilities. Similar to the fixed effects the issue of what variables to include in the design matrix is amenable to model selection. A simpler, arguably more natural question for random effects, asks if there should be any at all. Treating variables one by one or as an ensemble does not matter much for the implementation where the reversible jump (Green (1995)) sampler provides a unifying framework.

A simple question, particular to the models considered here is that of serial correlation or not. When $\rho = 0$ the model collapses into that of a ordinary mixed effects model. Adding or nesting a reversible jump move (Green (1995)) model selection or model averaging is straightforward. The practical implementation involves introducing a prior on the form

$$\rho \sim w_0 I_{\rho=0} + (1 - w_0) U(-1, 1),$$

where w_0 is a weight allocating prior mass to the model with no serial correlation present and $I_{\rho=0}$ the standard indicator function. The uniform density in this mixed component prior may be exchanged as long as it restricts the parameter to the stationary region. There being two major states, with serial correlation present or absent, only a single move and its implied inverse needs to be specified.

The move that attempts to add serial correlation to an ordinary random effects model is executed conditional on all else and relies on the same approximation used for updating ρ in the algorithms presented in Sections 4.2 and 4.3. Specifically, calculating the coefficients of the polynomial in (4.4), an approximation of the full conditional posterior is constructed and a proposed value r is sampled. Assuming Algorithm 4.2 or 4.3 is used, the move to this new state is accepted with probability

$$\alpha_{0 \rightarrow \rho} = \min \left\{ 1, (1 - r^2)^{N/2} \exp \{ (r) [c_1 + c_2(r)] \} \frac{(1 - w_0)/2}{w_0} \frac{j_{\rho \rightarrow 0}}{j_{0 \rightarrow \rho} q_{\rho}(r)} \right\}.$$

The generic expression for the acceptance probability is a simple product of a likelihood, prior, proposal ratio, and Jacobian which in this case is equal to one. The proposal ratio includes move probabilities from state s to state t ,

$j_s \rightarrow t$. The acceptance probability for the tandem move is just the inverse of the expression, where in the balancing move the current value ρ is treated as if it was sampled from $q_\rho(\rho)$.

4.5 Discussion

We apply the template for parametric covariance modelling approach of Paper 3 to the case of serially correlated data. In the simplest case of pure serial correlation we show how it is possible, assuming that a convenient Jeffrey's prior is used, to sample the posterior with a sequence of simple Gibbs steps. A strength of the parametric approach is the simplicity with which model uncertainty may be introduced and administrated.

Extending the model to include also random effects we show how the template is generally practical. Then, using data augmentation it is possible to split up the problem into two distinct parts, each being trivial to execute. For the special case of a random individual effects model it is possible to proceed without using data augmentation, albeit at increased computational costs that are probably worth considering only when the cross-section dimension is large.

As always there are many practical and technical issues that offer ample room for improvement. This is especially true for the complicated model with serial correlation and random effects. As is, the setup described will typically sample regression coefficients and random effects as efficiently as the extent of serial correlation allows. When it comes to variance components, the idiosyncratic error variance usually presents no serious problems. However, slow mixing over the unique components of Ω may be an issue. Chib and Carlin (1999) suggest a procedure based on a marginalization similar to that used in updating β above, this time over both β and $\{\gamma_i\}_{i=1}^n$. While elegant and based on simple computations, the extent of calculations necessary in each iteration quickly renders the sampler inefficient in terms of iterations per time unit, in particular as the cross-section size increases. An untried possibility is to go a small step in that direction by using a similar marginalization when updating the random effects. How well this works remains to be investigated.

Extending the model to more general dynamics is certainly possible though for the intended applications perhaps not a fruitful exercise. More interesting then is the possibility of extending the pure serial correlation model to be able to handle unequally spaced data. Using the alternative specification of the covariance this should be a simple but rewarding task.

Bibliography

- Anderson, T. W. and Hsiao, C. (1982), 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics* **18**, 47–82.
- Baltagi, B. H. and Li, Q. (1991), 'A transformation that will circumvent the problem of autocorrelation in an error component model', *Journal of Econometrics* **48**, 385–393.
- Baltagi, B. H. and Li, Q. (1994), 'Estimating error component models with general MA(q) disturbances', *Econometric Theory* **10**, 396–408.
- Chib, S. and Carlin, B. P. (1999), 'On MCMC sampling in hierarchical longitudinal models', *Statistics and Computing* **9**, 17–26.
- Fomby, T. B. and Guilkey, D. K. (1978), 'On choosing the optimal level of significance for the durbin-warson test and the bayesian alternative', *Journal of Econometrics* **8**, 203–213.
- Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* pp. 711–732.
- Lillard, L. A. and Willis, R. J. (1978), 'Dynamic aspects of earnings mobility', *Econometrica* **46**, 985–1012.
- MaCurdy, T. E. (1982), 'The use of time series processes to model the error structure of earnings in longitudinal data analysis', *Journal of Econometrics* **18**, 83–114.
- Ritter, C. and Tanner, M. A. (1992), 'Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler', *Journal of the American Statistical Association* **87**, 861–868.
- Tanner, M. A. and Wong, W. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* pp. 528–550.

- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.
- Zellner, A. and Tiao, G. C. (1964), 'Bayesian analysis of the regression model with autocorrelated errors', *Journal of the American Statistical Association* **59**, 763–778.

EFI

The Economic Research Institute

Reports since 1998

A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2003

- Andersson, H.**, Valuation and Hedging of Long-Term Asset-Linked Contracts.
- Bergman, M.**, Essays on Human Capital and Wage Formation.
- Damsgaard, N.**, Deregulation and Regulation of Electricity Markets.
- Eklund, B.**, Four Contributions to Statistical Inference in Econometrics.
- Hakkala, K.**, Essays on Restructuring and Production Decisions in Multi-Plant Firms.
- Holgersson, C.**, Rekrytering av företagsledare. En studie i hoimosocialitet.
- Ivaschenko, I.**, Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.
- Lange, F.**, Brand Choice in Goal-derived Categories - What are the Determinants?
- Le Coq, C.**, Quantity Choices and Market Power in Electricity Market.
- Magnusson, P.R.**, Customer-Oriented Product Development - Experiments Involving Users in Service Innovation.
- Meisiek, S.**, Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.
- Mårtensson, A.**, Managing Mission-Critical IT in the Financial Industry.
- Nilsson, G.**, Processorientering och styrning – Regler, mål eller värderingar?
- Sandberg, R.**, Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.
- Sturluson, J.T.**, Topics in the Industrial Organization of Electricity Markets.
- Tillberg, U.**, Ledarskap och samarbete – En jämförande fallstudie i tre skolor.
- Waldenström, D.**, Essays in Historical Finance.
- Wallén, U.**, Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.
- Ögren, A.**, Empirical Studies in Money, Credit and Banking - The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834 – 1913.

2002

- Barinaga, E.**, Levelling Vagueness – A study of cultural diversity in an international project group.
- Berglund, J.**, De otillräckliga - En studie av personalspecialisternas kamp för erkännande och status.
- Bolander, P.**, Anställningsbilder och rekryteringsbeslut.
- Damjanovic, T.**, Essays in Public Finance.
- Ekman, M.**, Studies in Health Economics – Modelling and Data Analysis of Costs and Survival.
- Företagerskan – Om kvinnor och entreprenörskap.** Holmquist, C. och Sundin, E (red)
- Heyman, F.**, Empirical Studies on Wages, Firm Performance and Job Turnover.
- Kallifatides, M.**, Modern företagsledning och omoderna företagsledare.
- Kaplan, M.**, Acquisition of Electronic Commerce Capability - The Cases of Compaq and Dell in Sweden.

Mähring, M., IT Project Governance.
Nilsson, M., Essays in Empirical Corporate Finance and Governance.
Rekratering av koncernstyrelsen – Nomineringsförfaranden och styrelsesammansättning med focus på kvinnors ställning och möjligheter.
 Sjöstrand, S-E. och Petrelius, P.,(red)
Scener ur ett företag – Organiseringsteori för kunskapssamhället. Löwstedt, J. Stymne, B.,(red).
Schenkel, A., Communities of Practice or Communities of Discipline - Managing Deviations at the Øresund Bridge.
Schuster, W., Företagets Valutarisk – En studie av horisontella och vertikala styrprocesser.
Skogsvik, S., Redovisningsmått, värder relevans och informationseffektivitet.
Sundén, D., The Dynamics of Pension Reform.
Ternström, I., The Management of Common-Pool Resources - Theoretical Essays and Empirical Evidence.
Tullberg, J., Reciprocitet – Etiska normer och praktiskt samarbete.
Westling, G., Balancing Innovation and Control – The Role of Face-to-face Meetings in Complex Product Development Projects.
Viklund, M., Risk Policy – Trust, Risk Perception, and Attitudes.
Vlachos, J., Risk Matters - Studies in Finance, Trade and Politics.

2001

Adolfson, M., Monetary Policy and Exchange Rates – Breakthrough of Pass-Through.
Andersson, P., Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.
Björklund, C., Work Motivation - Studies of its Determinants and Outcomes.
Center for Management and Organization 50 (1951-2001).
Charpentier, C., Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämndsreformen.
Dahlén, M., Marketing on the Web - Empirical Studies of Advertising and Promotion Effectiveness.
Eckerlund, I., Essays on the Economics of Medical Practice Variations.
Ekelund, M., Competition and Innovation in the Swedish Pharmaceutical Market.
Engström, S., Success Factors in Asset Management.
Ericsson, D., Kreativitetets mysteriet – Ledtrådar till arbetslivets kreativering och skrivandets metafysik.
Eriksson, R., Price Responses to Changes in Costs and Demand.
Frisell, L., Information and Politics.
Giordani, P., Essays in Monetary Economics and Applied Econometrics.
Gustavsson, P., Essays on Trade, Growth and Applied Econometrics.
Hedlund, A., Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.
Hill, M., Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.
Hvenmark, J., Varför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.
Hägglund, P.B., Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.
Höök, P., Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.
Johansson, C., Styrning för samordning.

Josephson, J., Evolution and Learning in Games.

Kjellberg, H., Organising Distribution - Hakonbolaget and the efforts to rationalise food distribution, 1940-1960.

Lange, F. och Wahlund, R., Category Management – När konsumenten är manager.

Liljenberg, A., Customer-gearred competition – A socio-Austrian explanation of Tertius Gaudens.

Lindkvist, B., Kunskapsöverföring mellan produktutvecklingsprojekt.

Ljunggren, U., Nyckeltal i grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen.

Läkemedel – Kostnad eller resurs för sjukvården? Jönsson, B.,(red).

Löf, M., On Seasonality and Cointegration.

Martensen, K., Essays on Entry Externalities and Market Segmentation.

Matros, A., Stochastic Stability and Equilibrium Selection in Games.

Mårtensson, P., Management Processes – An Information Perspective on Managerial Work.

Nilsson, A., Market Transparency.

Norberg, P., Finansmarknadens amoralitet och det kalvinska kyrkorummet – En studie i ekonomisk mentalitet och etik.

Persson, B., Essays on Altruism and Health Care Markets.

Rech, G., Modelling and Forecasting Economic Time Series with Single Hidden-layer Feedforward Autoregressive Artificial Neural Networks.

Skoglund, J., Essays on Random Effects Models and GARCH.

Strand, N., Empirical Studies of Pricing.

Thorén, B., Stadsdelsnämndsreformen och det ekonomiska styrsystemet - Om budgetavvikelser.

2000

Berg-Suurwee, U., Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.

Bergkvist, L., Advertising Effectiveness Measurement: Intermediate Constructs and Measures.

Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Koncernchefen och ägarna.

Bornefalk, A., Essays on Social Conflict and Reform.

Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform.

Edman, J., Information Use and Decision Making in Groups.

Emling, E., Svenskt familjeföretagande.

Ericson, M., Strategi, kalkyl, känsla.

Gunnarsson, J., Wahlund, R., Flink, H., Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.

Hellman, N., Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.

Hyll, M., Essays on the Term Structure of Interest Rates.

Håkansson, P., Beyond Private Label – The Strategic View on Distributor Own Brands.

I huvudet på kunden. Söderlund, M (red). *EFI och Liber Förlag*.

Karlsson Stider, A., Familjen och firman.

Ljunggren, U., Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.

Ludvigsen, J., The International Networking between European Logistical Operators.

Nittmar, H., Produktutveckling i samarbete – Strukturförändring vid införande av nya Informationssystem.

Robertsson, G., International Portfolio Choice and Trading Behavior.
Schwarz, B., Weinberg, S., Serviceproduktion och kostnader – att söka orsaker till kommunala skillnader.
Stenström, E., Konstiga företag.
Styrning av team och processer – Teoretiska perspektiv och fallstudier. Bengtsson, L., Lind, J., Samuelson, L.A., (red).
Sweet, S., Industrial Change Towards Environmental Sustainability – The Case of Replacing Chlorofluorocarbons.
Tamm Hallström, K., Kampen för auktoritet – standardiseringsorganisationer i arbete.

1999

Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Expert Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Flink, H., Gunnarsson, J., Wahlund, R., Svenska hushållens sparande och skuldsättning– ett konsumentbeteende perspektiv.
Gunnarsson, J., Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
Hamrefors, S., Spontaneous Environmental Scanning.
Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Japanese Production Management in Sunrise or Sunset. Karlsson, C., (red).
Jönsson, B., Jönsson, L., Kobelt, G., Modelling Disease Progression and the Effect of Treatment in Secondary Progressive MS. Research Report.
Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., Indikatorer i grundskolan i Stockholms stad före stadsdelsnämndsreformen – en kartläggning.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Mägi, A., Store Loyalty? An Empirical Study of Grocery Shopping.
Mölleryd, B.G., Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
Nilsson, K., Ledtider för ledningsinformation.
Osynlig Företagsledning. Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red).
Rognes, J., Telecommuting – Organisational Impact of Home Based – Telecommuting.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Spagnolo, G., Essays on Managerial Incentives and Product-Market Competition.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Svedberg Nilsson, K., Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
Söderström, U., Monetary Policy under Uncertainty.
Werr, A., The Language of Change The Roles of Methods in the Work of Management Consultants.
Wijkström, F., Svenskt organisationsliv – Framväxten av en ideell sektor.

1998

Andersson, M., On Testing and Forecasting in Fractionally Integrated Time Series Models.

Berg-Suurwee, U., Styrning av kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.

Berg-Suurwee, U., Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdelsnämndsreformen.

Bergström, F., Essays on the Political Economy of Industrial Policy.

Bild, M., Valuation of Takeovers.

Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.

Eriksson-Skoog, G., The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.

Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.

Ioannidis, D., I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.

Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.

Levin, J., Essays in Company Valuation.

Ljunggren, U., Styrning av grundskolan i Stockholms stad innan stadsdelsnämndsreformen.

Mattsson, S., Från stat till marknad – effekter på nätverksrelationer vid en bolagiseringsreform.

Nyberg, A., Innovation in Distribution Channels – An Evolutionary Approach.

Olsson, P., Studies in Company Valuation.

Reneby, J., Pricing Corporate Securities.

Roszbach, K., Essays on Banking Credit and Interest Rates.

Runsten, M., The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.

Segendorff, B., Essays on Bargaining and Delegation.

Sjöberg, L., Bagozzi, R., Ingvar, D.H., Will and Economic Behavior.

Sjögren, A., Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.

Studier i kostnadsintäktsanalys. Jennergren, P., (red)

Söderholm, J., Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.

Thorburn, K., Cash Auction Bankruptcy and Corporate Restructuring

Wijkström, F., Different Faces of Civil Society.

Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.