JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS

70

Jani Luoto

Bayesian Applications in Dynamic Econometric Models

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS 70

Jani Luoto

Bayesian Applications in Dynamic Econometric Models

Esitetään Jyväskylän yliopiston taloustieteiden tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Villa Ranan Blomstedtin salissa
tammikuun 10. päivänä 2009 kello 12.

Academic dissertation to be publicly discussed, by permission of
the School of Business and Economics of the University of Jyväskylä,
in the Building Villa Rana, Blomstedt Hall, on January 10, 2009 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2009

Bayesian Applications in Dynamic Econometric Models

JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS 70

Jani Luoto

Bayesian Applications in Dynamic Econometric Models

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2009

Editors
Tuomo Takala
School of Business and Economics, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-3568-9
ISBN 978-951-39-3568-9 (PDF)

ISBN 978-951-39-3434-7 (nid.)
ISSN 1457-1986

Jyväskylä University Printing House, Jyväskylä 2009

ABSTRACT

Jani Luoto
Bayesian Applications in Dynamic Econometric Models
Jyväskylä: University of Jyväskylä, 2009, 148 p.
(Jyväskylä Studies in Business and Economics,
ISSN 1457-1986; 70)
ISBN 978-951-39-3568-9 (PDF), 978-951-39-3434-7 (nid.)
Diss.

The purpose of this thesis is to provide a few new ideas to the field of Bayesian econometrics. In the first essay, we provide an easily implementable method for the Bayesian analysis of a simple hybrid DSGE model of Clarida et al. (1999). The forecasting properties of the model are tested against commonly used forecasting tools, such as Bayesian VARs and naïve forecasts based on univariate random walks. Our posterior evidence implies that the model captures the predictable behavior of the three U.S. key macroeconomic variables, inflation, short-term nominal interest rate and a measure of output gap, very well. In the second essay, using GARCH-in-Mean models based on the following Intertemporal Capital Asset Pricing Model $E_{t-1}(r_t) = \mu_0 + \mu_1 Var_{t-1}(r_t)$ for the expected excess return r_t, we study the robustness of the risk-return relationship in U.S. stock market returns. The issue is important, since unnecessarily including μ_0 in the previous expression is known to distort conclusions, while restricting μ_0 to zero forces the expected excess return to equal the risk-free interest rate under the hypothesis $\mu_1 = 0$. Our evidence indicates that the existence of a risk-return relationship is fairly robust in that it does not *strongly* depend on the prior beliefs concerning the intercept, especially when the true value of μ_0 is sufficiently close to zero. In the third essay, we expand Kleibergen and Zivot's (2003) Bayesian Two Stage (B2S) model by allowing for unequal variances. To the best of our knowledge there is no single Bayesian study of instrumental variable (IV) models with unequal variances, although from the Bayesian point of view modelling heteroscedasticity should improve the precision of estimates and the quality of predictive inference. As an application we present a cross-country Cobb-Douglas production function estimation. In the fourth essay, we provide a simple epidemiology model where households, when forming their inflation expectations, adopt the past release of inflation with certain probability rather than the forward-looking newspaper forecast as suggested in previous literature. The posterior model probabilities based on the Michigan survey data strongly support the proposed model. Our results show that this simple model is also able to capture the heterogeneity in households' expectations very well. In the fifth essay, the Bayesian structural vector autoregressive model and the Finnish aggregate infrastructure capital series from 1860 to 2003 are used to explore how government infrastructure policy affects long-run output growth. We base our conclusions on posterior analysis, since it allows us to draw exact inference on parameters with near non-stationary data. We find strong and robust support in the Finnish data to indicate that permanent changes in government infrastructure policy can have permanent effects on the growth rate of output.

Keywords: Bayesian inference, Monte Carlo methods, Prior elicitation

Author's address Jani Luoto
School of Business and Economics
P.O. Box 35, FIN-40014 University of Jyväskylä,
Finland

Supervisor Markku Lanne
Department of Economics
University of Helsinki

Reviewers Sune Karlsson
Swedish Business School
Örebro University

Matti Virén
Department of Economics
University of Turku

Opponent Antti Kanto
Department of law
University of Tampere

ACKNOWLEDGEMENTS

It is time to kindly thank the many persons from whom I have received guidance and support during this process. I am most grateful to my brilliant supervisors Ph.D Arto Luoma and Professor Markku Lanne for the opportunity to work with them. Especially I would like to thank Arto and Markku for their patience in always being available to answer questions, comment on and give advice. Arto is also my old good friend who has played especially vital role in this process. I also wish to express my sincere gratitude to my co-author Mr. Jan-Erik Antipin who has prepared his doctoral thesis at the same time as me and has been my friend from my undergraduate years in Tampere.

During the research process I have been part of some research-related communities. It's has been a pleasure to work at the Department of Economics in University of Jyväskylä. I would like to thank all the staff and especially Mr. Juhani Raatikainen, Professor Kari Heimonen, Ph.D Esa Mangeloja and Mrs. Lea Nevalainen. Special thanks to Mr. Erkki Siivonen at VATT for many rewarding moments. I also like to thank Ph.D Vesa Vihriälä and Ph.D Raija Volk at PTT who have had a particular influence in my academic career.

In the early phases of the project I was privileged to receive financial support from the Yrjö Jahnsson Foundation and University of Tampere. I wish to thank them for their valuable support. I would also like to thank Professors Sune Karlsson and Matti Virén providing valuable comments to all of my papers.

Finally, I want to thank my family, Sanni and Emma-Sofia, all my brothers and sisters in Christ (including you Arto), and my parents and sister; it would not have been possible to complete this project without your help!

Tampere, November 2008

Jani Luoto

CONTENTS

1 INTRODUCTION

1.1 A Brief Overview on Bayesian Econometrics

The purpose of this thesis is to provide a few new ideas to the field of Bayesian econometrics. For example, it proposes a new econometric model, shows how to use priors to check the empirical validity of theoretical restrictions and provides an efficient way to improve the forecasting performance of popular New Keynesian models. Bayesian methods are used to solve problems of statistical inference which cannot be easily solved in the non-Bayesian framework. In particular, the focus of the thesis is on analyzing dynamic econometric models. Most models include nonlinear components and we often deal with small samples or near unit root data.

Since prior distributions have a very important role in Bayesian analysis, considerable effort is spent on prior elicitation and discussion on the used priors. Throughout the thesis economic theory ('collectively' subjective beliefs) and information derived from data sets not included in the estimation sample are used as a basis of prior knowledge. In some models standard uninformative priors are used, though.

The estimated models have nonstandard posterior distributions requiring the use of Monte Carlo methods in their evaluation. The correct implementation of these methods is of course crucial for the validity of results. Therefore, detailed descriptions of the used algorithms and simulation routines are given. The convergence of the Markov Chain Monte Carlo (MCMC) samplers are also assessed using formal diagnostics.

A major finding of this thesis is that when classical methods fail to produce reliable inference, Bayesian methods can still, in many cases, be successfully applied. Our empirical evidence implies that Bayesian methods are particularly useful when the competitive models are non-nested, the likelihood function is multimodal or peaks in an economically non-sensible region, economic theory provides important or necessary information that is not contained in the estimation sample, or the economic variables have some 'silent' information which can not be modeled using the likelihood only. These results are of course well known in Bayesian econometrics. Our contribution is to give some new ideas and new empirical evidence to the literature.

Finally, the reader will note that we are not fully consistent in using Bayesian analysis everywhere. In some cases we use classical tests in preliminary data analysis, since these tests are readily available and do not demand extra programming efforts. In addition, their use increases comparability with some earlier studies.

In the following subsections we will give, among some discussion and brief literature review, an illustrative example of the usefulness of Bayesian inference in modern economic research. This is done by taking a brief look at the Bayesian method, which has been widely used among economists and policy makers over the past 15 years. In section 1.2 we will give the summaries of the essays.

1.1.1 Some Discussion

We begin our 'tour' by a brief discussion on the main distinction between classical and Bayesian methods. We then continue by taking a short look at the prior issue. After looking at the priors we will give an incomplete overview on the history of Bayesian econometrics. To give an example of the practicability of Bayesian methods, for a reader who is unfamiliar with the Bayesian approach, the Bayesian estimation of popular dynamic stochastic general equilibrium (DSGE) models is also discussed. In particular, an overview of the literature and a short discussion on why Bayesian inference works well in this environment are given.

Given the observations y (the data) and the model M, the starting point of Bayesian analysis is to determine the prior density function $p(\theta\,|\,M)$ of unobservable parameters θ, which together with the density of observations $L(y\,|\,\theta,M)$ (likelihood function), yields the conditional density of θ given y and M,

$$p(\theta|y,M)=\frac{p(\theta,y|M)}{p(y,|M)}=\frac{p(\theta\mid M)L(y\mid\theta,M)}{\int p(\theta\mid M)L(y\mid\theta,M)d\theta}. \qquad (1)$$

Expression (1) is called the posterior density of θ and the corresponding distribution is called posterior distribution. The marginal density $\int p(\theta|M)L(y;\theta,M)d\theta$, which is called marginal likelihood, is relevant in Bayesian decision making, since Bayesians often base model inference on them (see e.g. Hoeting et al., 1999, for a complete introduction to Bayesian model averaging where marginal likelihoods are utilized in an efficient and practical way).

The major difference between the Bayesian and classical schools of thought is in conditioning (see e.g. DeGroot, 1970, and Poirier, 1988) for a discussion on the distinction between Bayesian and non-Bayesian methods. The classical school conditions on an unknown θ, and compares the likelihood function with the data. Bayesians condition on actually available information y and base their inference on the full density $p(\theta,y|M)$. Thus, classical statistics has the problem of conditioning on what is unknown, but it does not need any statement about prior density. Bayesian statistics instead provides a coherent tool for decision making, but requires the specification of prior density $p(\theta|M)$ (see Geweke, 2005) for discussion.

The prior density reflects the researcher's subjective beliefs concerning plausible parameter values (or models etc.). When the researcher does not have a clear idea about the value of a parameter or does not want his prior knowledge to affect estimation results, he can assign to it a noninformative prior distribution. Noninformative distributions have typically a large variance or are improper, that is, non-integrable. Bayesian econometrics has a long tradition of noninformative priors (especially improper priors) and there is a variety of standard textbook models which follow this tradition (see e.g. Jeffreys, 1961, Zellner, 1971, Bauwens et al., 1999, Koop, 2003, and Lancaster, 2004).

A commonly known problem with improper priors is that they cause marginal likelihoods, which are typically used in model comparison, to be indeterminate, since the normalizing constants of these priors are not defined. On the other hand, there are methods for model comparison which allow for the use of improper priors. For example, Gelman et al. (2004) introduce the average discrepancy, defined as the posterior mean of deviance, as a criterion for model comparison. They prefer using discrepancy between data and model to using marginal likelihoods in model comparisons. They consider marginal likelihoods to be in most cases irrelevant, since they are used to compute the relative probabilities of the models conditional on one of them being true. The usual argument against deviances is that they have no proper scale. However, a similar statement can be made of posterior model probabilities when the list of competitive models excludes the true model. Furthermore, priors have a prominent role in determining the values of marginal likelihoods (see e.g. Adolfson et al., 2007a, for discussion), whereas the average discrepancy is not sensitive to priors.

A typical informative prior, on the other hand, reflects the researcher's subjective beliefs (when she/he has them), summarizes information from the data not included in the estimation sample, or is based on both of them. Often the underlying economic theory provides a natural starting point for prior elicitation in econometrics.

The use of informative priors is usually subject to strong criticism, sometimes rightly, by reason of their tendency to increase the degree of subjectivity in statistical inference. However, informative priors can – when well designed – facilitate numerical maximization and provide a practical tool to handle potential model misspecification and lack of parameter identification (see e.g. Fernández-Villaverde and Rubio-Ramírez, 2004). Furthermore, if the researcher is unsure of whether his prior influences on the posterior in an undesirable way, there are methods to control for the robustness of posterior results with respect to changes in the prior distribution. The recent Bayesian analysis of very popular dynamic stochastic general equilibrium (DSGE) models provides an excellent example of an efficient prior use and we will return to this issue in the next subsection (see e.g. Sungbae and Schorfheide, 2007, and Del Negro and Schorfheide, 2008, for discussion).

1.1.2 Related Literature

There is a rabidly growing literature using Bayesian methods to solve problems of statistical inference. For example, Poirier (1989, 2004) find strong empirical evidence on a rapid growth of Bayesian publications in statistics over the past 35th years. An upward trend of Bayesian publications in economics (mainly in theoretical economics but also in empirical economics) can also be found, but surprisingly not in econometrics (see Poirier, 1989, 2004, for discussion).

The increasing interest of Bayesian methods in statistics is largely a result of two elements. Firstly, the computational revolution of the 1990s strongly increased the practicability of computationally intensive Bayesian methods. Secondly, Bayesian methods can be, in many cases, successfully implemented even when the mainstream statistics fails to produce reliable inference. This is particularly common when the statistical inference is based on a system of equations (cf. the recent Bayesian analysis of DSGE models). Despite of this the Bayesian approach has been dismissed by several mainstream *econometricians* mainly because of its subjectivism (see Qin, 1993, and Zellner, 2008, for discussion). However, the true reason for the unpopularity of Bayesianism in econometrics is likely the complexity of Bayesian methods, not subjectivism, since the methods used in mainstream econometrics are often much simpler than the corresponding Bayesian methods.

Anyhow, all econometricians use prior information more or less subjectively. Model selection provides a fairly illustrative example of this proposition. In model selection Bayesians typically combine model priors and marginal likelihoods to utilize posterior model probabilities associated with alternative models to averaging over these models, whereas frequentists typically select a model from some class of models (typically chosen using their past experience i.e. subjective prior information) and then proceed as if the selected model had generated the data. According to Zellner (2008), Bayesian econometricians learn using an explicit model, Bayes' theorem that allows prior information to be employed in a formal and reproducible manner whereas classical econometricians learn in an informal, subjective manner.

An incomplete list of important works which have been particularly influential in Bayesian econometrics, to the best of our knowledge, is given below (see e.g. Berry et al., 1996, and Zellner, 2006, 2008, for a complete analysis of the topic).

The central Bayesian book of Jeffreys (1939), The Theory of Probability, provides important methods for deriving diffuse priors and constructing Bayesian (statistical) tests, among other things (see Zellner, 1980 and 2008, for discussion). Jeffreys' work has influenced all areas of Bayesian econometrics, and a large amount of econometricians have based their statistical inference on his work.

Friedman and Savage (1948, 1952) provide an early Bayesian work which combined economic utility theory and statistical theory to the decision theoretic Bayesian approach. The illustrative book of Raiffa and Schlaifer (1961) provides methods and applications of the Bayesian decision theory. The work of Fisher (1962), in which the author thoroughly analyzes the Bayesian solution of the decision problem, is also worth of mention.

Very important and influential works on the Bayesian analysis of simultaneous equation models are given by Drèze (1962, 1976), Rothenberg (1963) and Drèze and Richard (1983), whereas earlier Bayesian works on evaluating and estimating structural dynamic econometric models can be found in Zellner and Palm (1974, 1975) and Zellner (1965) (see also, for example, Kleibergen and Zivot, 2003, for a recent treatment of Bayesian simultaneous equation models).

A Bayesian solution for regression models with autocorrelated errors was initiated by Zellner and Tiao (1964) (see Bauwens et al., 1999, for a recent treatment of regression models with autocorrelated errors), while the paper of Tiao and Zellner (1964) provides an original work of multivariate regression models (see also Geisser, 1965). The influential book of Zellner (1971), An Introduction to Bayesian Inference in Econometrics, has also had a very prominent role in Bayesian econometrics, and is therefore important to mention in this context.

A pioneering work of Bayesian model comparison in econometrics is given by Geisel (1975). In particular, his study was the first, to the best of our knowledge, which used posterior odds to compare competitive macroeconomic models, an approach which has become very popular in current empirical macro-economics (see e.g. Smets and Wouters, 2003, 2005 and 2007, Sala-i-Martin et al., 2004, and Sungbae and Schorfheide, 2007, and references therein).

Stein (1956, 1962) developed Bayesian shrinkage estimators that outperform a variety of other estimators relative to quadratic loss in several models. Over the past 30th years methods based on his 'shrinkage idea' have become popular, especially in forecasting, among economists working in private, government and academic sectors (see e.g. Bawa et al., 1979, Quintana et al., 1995, Schirm and DiCarlo, 1998 and Cogley et al., 2005). The work of the so called 'Minnesota crew' (see Litterman 1980, Doan et al. 1984, and Litterman 1986) has had a prominent role in this development (see also Kadiyala and Karlsson, 1997, who provide systematic treatment and simulation methods for Bayesian vector autoregressive (VAR) models with Minnesota-type priors).

Finally, important works concerning posterior simulation, which has a crucial role in Bayesian econometrics, include Kloek and van Dijk (1978), Drèze and Richard (1983) and Gelfand and Smith (1990) (see also Geweke, 1999a and 2005, and references therein).

To convince the non-Bayesian reader (econometrician) of the practicability of Bayesian methods in econometrics, we will discuss below on Bayesian methods developed to estimate and evaluate DSGE models in recent years. This rapidly growing empirical literature provides a useful starting point for such a mission, since there is nearly consensus among macroeconomists on the usefulness of the Bayesian approach in this context.

The maximum likelihood (ML) estimation of DSGE models has, in general, turned out to be a challenging task. Often DSGE models have multimodal likelihoods that may peak in economically unreasonable areas. This may easily lead to absurd parameter estimates. For instance, the ML estimates of structural parameters, such as the constant discount factor or the elasticity of labor supply, are often observed to be at odds with evidence obtained from micro-level data, due to the stylized nature of DSGE models.

In situations like this, informative priors can be used to incorporate additional information into estimation, at the cost of increasing subjectivism in empirical analysis. However, if we are ready to accept this cost, marginal priors can be used, for instance, to facilitate numerical maximization, down-weigh the regions of the parameter space which are at odds with observations not included in the estimation sample or add curvature to a likelihood function which is flat in some dimensions of the parameter space (see e.g. Sungbae and Schorfheide, 2007). Thus, priors, when they are well designed, can provide a fairly practical tool to handle many potential problems in empirical analysis, such as model misspecification and lack of parameter identification (see e.g. Fernández-Villaverde and Rubio-Ramírez, 2004).

Not surprisingly, Bayesian methods have become very popular in the analysis of DSGE models. Sungbae and Schorfheide (2007) provide an excellent review of the Bayesian methods which have been developed in recent years to estimate and evaluate models of this class. Table 1 lists the major contributors of this literature. As we can see, Bayesian approaches to calibration are proposed by Canova (1994), DeJong, Ingram, and Whiteman (1996) and Geweke (1999b). These early works provide empirical applications that assess business cycle and asset pricing implications of stochastic growth models.

The likelihood-based Bayesian estimation of DSGE models was initiated by the works of Landon-Lane (1998), DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), and Otrok (2001). Among these authors DeJong, Ingram, and Whiteman (2000) examined the forecasting performance of stochastic growth models. Schorfheide (2000) considered cash-in-advance monetary DSGE models, while Otrok (2001) explored the welfare costs of business cycles using a real business cycle model with habit formation.

TABLE 1 List of Contributors

Contributors	Content
Canova (1994), DeJong et al. (1996) and Geweke (1999b)	These authors used Bayesian calibration in their analyses and explored the business cycle and asset pricing implications of stochastic growth models.
Landon-Lane (1998), DeJong et al. (2000), Schorfheide (2000) and Otrok (2001)	Started the likelihood-based Bayesian estimation of DSGE models. Examined a variety of issues, such as the forecasting performance of stochastic growth models, cash-in-advance economy and the welfare costs of business cycles.
DeJong and Ingram (2001), Chang and Schorfheide (2003), Galí and Rabanal (2004) and Fernández-Villaverde and Rubio-Ramírez (2004a)	Expanded the field: by studying the cyclical behavior of skill accumulation, estimating a home-production model to study the importance of labor supply shocks, estimating a cattle-cycle model and studying the effect of technology shocks on hours worked.
Rabanal Rubio-Ramírez (2005)	Compared variants of the small-scale New Keynesian DSGE model using marginal likelihoods.
Schorfheide (2005)	Allowed for a regime-switching of the target inflation level in the monetary policy rule.
Lubik and Schorfheide (2007) and Del Negro (2003)	Used marginal likelihoods and small-scale open economy models to answer the question whether the central banks respond to exchange rates.
Lubik and Schorfheide (2005), Rabanal and Tuesta (2005), and deWalque and Wouters (2004)	Estimated multi-country DSGE models.
Adolfson et al. (2007), Adolfson et al. (2007a) and Adolfson et al. (2008)	Used large-scale open economy DSGE models to explore the out-of-sample forecasting performance of these models and captured the volatility and persistence of real exchange rate without assuming unreasonable degrees of price rigidity.
Smets and Wouters (2003, 2005 and 2007), Christiano et al. (2005), Laforte (2004), Onatski and Williams (2004), and Levin et al. (2005)	Analyzed large-scale closed economy DSGE models, which include capital accumulation, additional shocks and real (and nominal) frictions and can be used for policy analysis and forecasting.

The cyclical behavior of skill accumulation was studied by the work of DeJong and Ingram (2001). Chang and Schorfheide (2003) estimated a home-production model to study the importance of labor supply shocks, whereas Fernández-Villaverde and Rubio-Ramírez (2004a) estimated a cattle-cycle model.

Galí and Rabanal (2004) used the DSGE model to study the effect of technology shocks on hours worked. Variants of the small-scale New Keynesian DSGE model are compared by Rabanal and Rubio-Ramírez (2005) using marginal likelihoods and U.S. data. Schorfheide (2005) allowed for regime-switching of the target inflation level in the monetary policy rule, whereas

Canova (2006) explored the stability of the structural parameters in the U.S. economy by estimating a small-scale New Keynesian model recursively.

In the open economy literature Lubik and Schorfheide (2007) computed marginal likelihoods using the small-scale open economy DSGE model to answer the question whether the central banks of Australia, Canada, England, and New Zealand respond to exchange rates. A similar model is applied by Del Negro (2003) to Mexican data. Justiniano and Preston (2004) studied imperfect exchange rate passthrough, whereas Lubik and Schorfheide (2005), Rabanal and Tuesta (2005), and De Walque and Wouters (2004) have estimated multi-country DSGE models. In addition, the research staff of Sveriges Riksbank (the central bank of Sweden) has also put a large amount of effort to the evaluation and (Bayesian) estimation of open-economy DSGE models (see e.g. Adolfson et al., 2005, Adolfson et al., 2007a and b, Adolfson et al., 2007, and Adolfson et al., 2008).

Finally, large-scale models that include capital accumulation and additional shocks and real and nominal frictions are analyzed by Smets and Wouters (2003, 2005) both for the U.S. and the Euro Area. The interested reader would also like to read the very intuitive paper of Smets and Wouters (2007). These authors show that a large-scale DSGE model with capital accumulation and various nominal and real frictions can have a model fit comparable to that of reduced-form Bayesian vector autoregressive (VAR) models (estimated with well-designed shrinkage methods). Models similar to that of Smets and Wouters (2003), which can be used for policy analysis and forecasting, have been estimated by Christiano et al. (2005), Laforte (2004), Onatski and Williams (2004), and Levin et al. (2005).

Table 2 summarizes the major issues involved in the likelihood based inference of DSGE models. As Table 2 indicates, in the DSGE framework, model misspecification may occur because of non-singularity in the forecast error covariance matrix. For example, Sims (2002b) suggests dealing with this issue using procedures that can be applied despite the singularity, whereas Sargent (1989), Altug (1989), Ireland (2004), Leeper and Sims (1994) and Smets and Wouters (2003) propose removing the singularity by adding measurement errors or additional structural shocks to the model mechanism. The major problem in the latter solution is related to model uncertainty. That is, how to decide which shocks (or measurement errors) should be added to the DSGE model mechanism? A feasible Bayesian solution to this problem is model averaging.

TABLE 2 Major Issues Involved in the Likelihood Based Inference of DSGE Models

Issue	Problem	Solution	Problem in Solution
Model misspecification	● Non-singularity in forecast error covariance matrix	● Procedures that can be applied despite the singularity ● Remove the singularity by adding so-called measurement errors or additional structural shocks to the model	● Model uncertainty: which shocks or measurement errors should be introduced to the model? ● A feasible solution: Bayesian model averaging
	● Invalid cross-coefficient restrictions which cause poor out-of -sample fit	● Larger DSGE models equipped with capital accumulation and various nominal and real frictions	● Complexity of models, which reduces practic-ability. ● Model uncertainty: which variables should be introduced to the model? ● A feasible solution: Bayesian model averaging
	● Absurd parameter estimates due to the stylized nature of DSGE models	● Use of informative priors to down-weigh the regions of the parameter space which are at odds with micro-level evidence	● Subjectivism: what features of the posterior are generated by the prior rather than the likelihood? ● A direct comparison of priors and posteriors may produce one solution to this problem.
Identification	● Lack of identification of the structural parameters	● Use of informative priors to add curvature to a likelihood function which is flat in some dimensions of the parameter space	● Subjectivism: the use of informative priors may require a considerable amount of effort to confirm that the important features of posterior results are not driven by these priors ● A direct comparison of priors and posteriors.
Model comparison	● Posterior odds, typically used in model comparison, are proportional to prior odds; model comparison is thus sensitive to the prior. ● Typically, DSGE priors are in part based on data-peeking, which weakens the credibility of model comparison.	● Use of a training-sample to cancel the influence of the prior on the marginal likelihood ● Use of alternative model comparison methods, such as the average discrepancy, defined as the posterior mean of deviance.	● How to determine the size of the training sample? ● Training sample priors for small training samples can easily be eccentric. ● The usual argument against deviances is that they have no proper scale.

The second source of model misspecification comes from potentially invalid cross-coefficient restrictions, which often cause poor out-of-sample fit. The literature has responded to this issue by developing large-scale DSGE models equipped with capital accumulation and various nominal and real frictions (see Smets and Wouters, 2003, 2005 and 2007, and Christiano et al., 2005). Again Bayesian model averaging is needed, since introducing additional variables into the model mechanism increase model uncertainty. Smets and Wouters (2007) take steps to this direction by testing which frictions are empirically important by comparing the marginal likelihoods of the alternative models. Another problem of large-scale models is their complexity, which decreases their practicability.

Absurd parameter estimates are, due to the stylized nature of DSGE models, also common in empirical analysis. In situations like this, the use of informative priors to down-weigh the regions of the parameter space which are at odds with micro-level studies may be practical. This, however, easily provokes questions about the reliability of the analysis. That is, which features of the posterior are generated by the prior rather than the likelihood? A direct comparison of priors and posteriors may provide one solution to this problem.
The lack of identification of the structural parameters is studied by, e.g., Beyer and Farmer (2004), Canova and Sala (2005) and Sungbae and Schorfheide (2007). A Bayesian solution to this problem is to use priors to add curvature to a likelihood function which is flat in some dimensions of the parameter space. However, the use of informative priors may require a considerable amount of effort to confirm that the important features of posterior results are not driven by these priors.

Finally, posterior odds are typically used in model comparison. However, they are proportional to prior odds, which makes model comparison sensitive to the prior. Furthermore, the DSGE priors are typically based on data-peeking in part, which weakens the credibility of model comparison. We can of course use a training sample to cancel the influence of the prior on the marginal likelihood. However, training sample priors, especially with small training samples, can easily be eccentric. More importantly, determining the size of the training sample is somewhat difficult, and far from unique. As an alternative strategy, we suggest using other model comparison methods, such as the average discrepancy, defined as the posterior mean of deviance. However, the usual argument against deviances is that they have no proper scale.

In sum, Bayesian methods can provide straightforward solutions to many fundamental problems related to the analysis of DSGE models. However, all approaches have their pros and cons, as Table 2 clearly points out.

Finally, it is clear that Bayesian methods can provide useful tools for all areas of economic research, not just for the analysis of DSGE models. However, since this class of models has become very popular in economics over the past 15 years, it provides fairly illustrative examples of the practicability of the Bayesian approach in econometrics. In the next subsection, when giving the summaries of the essays, we will suggest several other Bayesian tools having important roles in reliable econometric analysis.

1.2 Essays

1.2.1 The Forecasting Performance of the Small-Scale Hybrid New Keynesian Model

In essay 1, we provide a method for the Bayesian analysis of a simple hybrid DSGE model (or the New Keynesian (NK) model) of Clarida et al. (1999). This method is easily implementable and leads to savings in the CPU time required in posterior simulation, compared to the commonly used Kalman filter approach. Specifically, the Bayesian full information framework is adopted.

We choose to use Bayesian methods in estimation since the full information maximum likelihood (FIML) estimates turned out to be very sensitive to the starting values of maximization due to a multimodal likelihood. According to Lindé (2005), the maximum likelihood estimation of the hybrid model can with different starting values converge to local equilibria with more or less plausible parameter values. Surprisingly, this problem remained even if the parameter space was restricted to an economically feasible region.

The Bayesian full information analysis instead turned out to produce reliable parameter estimates. This is mainly since our joint prior is well designed in allowing the parameters to be estimated fairly freely, while being informative enough to keep the posterior distribution away from economically unreasonable values. To be more concrete, the marginal priors of the taste and policy parameters of the hybrid model come from micro-level studies, while the priors of the autoregressive parameters are based on a simple parameter transformation, which forces the posteriors of these parameters to be located in the interval [–1, 1]. Furthermore, the standard deviations of structural shocks are assumed to follow an inverse-gamma distribution with the shape and scale parameters yielding fairly loose priors.

The main objective of this essay is to compare the forecasting properties of the hybrid DSGE model against commonly used forecasting tools, such as Bayesian VARs and naïve forecasts based on univariate random walks. To do this properly we use quarterly ex post and real-time U.S. data from 1953:2 to 2004:4. In particular, the predictability of three key macroeconomic-variables, inflation, short-term nominal interest rate and a measure of output gap are studied. The source of the ex post data is the FREDII databank of the Federal Reserve Bank of St. Louis, while the source of the real-time data is the Federal Reserve Bank of Philadelphia.

Our out-of-sample forecast evidence implies that in the entire forecast sample (1976:4-2004:4) the forecasts of the hybrid DSGE model outperform those of the Bayesian VARs, while in the low inflation subsample (1990:1-2004:4) all the multivariate forecasting methods seem to produce equally accurate forecasts. In addition, both the DSGE and VAR models turned out to produce inflation forecasts which outperformed the naïve forecasts up to six quarters in all samples. This result is particularly important, since Atkeson and

Ohanian (2001) found that the one-year-ahead Federal Reserve's Greenbook inflation forecast has not been better on average than the naïve forecast since 1984.

We may conclude that the simple hybrid DSGE model captures the predictable behavior of the three U.S. key macroeconomic variables very well. The result is very interesting, since several recent papers have suggested different ways to improve the forecast performance of DSGE models at the cost of increasing the complexity of model mechanisms, thus decreasing the practicability of these approaches.

1.2.2 Robustness of the Risk-Return Relationship in the U.S. Stock Market

In essay 2, we study the robustness of the risk-return relationship in monthly U.S. stock market returns with respect to the specification of the conditional mean equation. The starting point of our analysis is the following linear equation for the expected excess return r_t, implied by Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM),

$$E_{t-1}(r_t) = \mu_0 + \mu_1 Var_{t-1}(r_t),$$

where the slope coefficient μ_1 is expected to be positive and the intercept μ_0 should be zero. Most of the previous empirical work on the ICAPM is based on the GARCH-in-Mean (GARCH-M) model, and it is the model used in our empirical analysis.

The empirical evidence of ICAPM is mixed, but typically μ_1 has been found insignificant (see Ghysels et al., 2005, and references therein). However, Lanne and Saikkonen (2006) recently showed that adding the theoretical restriction $\mu_0 = 0$ to the previous expression produces reasonable empirical results. Unfortunately, this restriction may be criticized, since it forces the expected excess return of the stock market to equal the risk-free interest rate under the null hypothesis $\mu_1 = 0$. This is in conflict with the stock premium puzzle literature put forth by Mehra and Prescott (1985), according to which the average stock return, in the U.S. particularly, has been excessively high relative to the risk.

Therefore, the results of Lanne and Saikkonen (2006) concerning the risk-return relationship should be checked for robustness to confirm that they are not driven by the zero restriction on the intercept. To this end, we make use of Bayesian methods and estimate a model assuming a number of different prior beliefs in order to observe the sensitivity of the estimate of μ_1.

In particular, we assume a zero mean normal prior distribution for μ_0 and consider several alternative plausible values for the prior variance of μ_0 to see how the tightness of the prior assumption $\mu_0 = 0$ affects the estimation results. Specifically, we let the prior standard deviation (σ_μ) of μ_0 vary from $6^{-1}\times0.003$ to 10×0.003, where 0.003 is the standard error of the maximum likelihood estimate of μ_0 reported by Lanne and Saikkonen (2006). In practice, $\sigma_\mu = 10\times0.003$ yields a noninformative prior distribution for μ_0, while $\sigma_\mu = 6^{-1}\times 0.003$ is small enough to force the posterior mean of μ_0 to be close to zero. Although the equity premium puzzle indicates a positive μ_0, in this exercise we rely on the underlying economic theory (ICAPM), which suggests zero mean on μ_0, as a basis of our prior knowledge. Values of the prior variance of μ_0 close to zero are, thus, more in accordance with the ICAPM, while higher values of the prior variance of μ_0 allow for the possibility of the equity premium.

The data are downloaded from Kenneth French's homepage. The data set consists of the monthly excess return on the value-weighted CRSP index (in excess of the 3-month Treasury bill rate) from 1928:1 to 2004:12. In addition to the entire sample, the models are estimated for the 1928:1 – 1966:6 and 1966:7 – 2004:12 subsample periods of equal length. This serves as a convenient check for robustness and parameter constancy.

Our empirical evidence lends support to the robustness of the positive risk-return relationship in the U.S. stock market data, albeit the evidence is weaker in the first subsample period. In other words, this conclusion is shown not to be affected by (falsely) restricting the intercept term in the equation for the expected excess return equal to zero, when its true value is close to zero. The estimates of μ_1, however, were shown to be affected by imposing this restriction if untrue, as was the case in the full sample and the first subsample periods.

One possible drawback of our analysis comes from the role which priors have in determining the values of marginal likelihoods. That is, when the prior variance of μ_0 is set to be small enough to force the posterior mean of μ_0 to be close to zero, the prior may become informative enough to have non-negligible influence on the marginal likelihood. We control this issue by two ways. Firstly, we estimate the marginal likelihood with the full restriction of zero intercept ($\mu_0 = 0$). The results of this exercise were very close to that where the posterior mean of μ_0 was forced to be close to zero using small prior variance of μ_0. Secondly, in addition to marginal likelihoods, estimated average discrepancies, which are not sensitive to priors, are used as well (see Gelman et al., 2004). Fortunately, both methods of model comparison turned out to produce similar conclusions.

1.2.3 Bayesian Two-Stage Regression with Parametric Heteroscedasticity

In essay 3, we expand Kleibergen and Zivot's (2003) Bayesian Two Stage (B2S) model by allowing for unequal variances. In classical analysis, modelling heteroscedasticity improves the efficiency of estimation and enables the variance estimates to be consistent. Thus, not surprisingly, modelling heteroscedasticity has become standard in classical IV literature. However, there is no single Bayesian study (to our knowledge) of IV models with unequal variances, although from the Bayesian point of view, modelling heteroscedasticity should improve the precision of estimates and the quality of predictive inference. The latter follows from the fact that modelling heteroscedasticity allows predictive inferences to be more precise for some units and less precise for other.

Our choice for modelling heteroscedasticity is a fully Bayesian parametric approach. As a prior distribution we will use a modification of the Jeffreys prior distribution (see e.g. Zellner, 1971). Although the derivation of our prior distribution is somewhat arbitrary (as will be seen) it yields a prior density which has a very practical property. The B2S model with heteroscedasticity correction is not informative regarding the vector of structural parameters β when the parameter matrix Π of the regression coefficients of the first-stage regression has reduced rank. In the presence of our prior, the marginal posterior of Π does not have a non-integrable asymptote in this situation. The parameter matrix Π is close to zero or close to having reduced rank in the case of weak instruments, that is, when the instruments are only weakly correlated with the endogenous regressors.

To make an empirical illustration of the properties of the heteroscedastic B2S model, we follow Benhabib and Spiegel (1994) and Papageorgiou (2003) and construct a simple exercise of aggregate production function estimation using a cross-country sample of 85 non-oil-producing countries. In this essay, we use the Penn World Table 6.1 dataset. The usefulness of this dataset is that the data is adjusted for differences in purchasing power across countries, which allows for strict comparability of the levels of GDP. The source of the years-of-schooling series is Cohen and Soto (2001). In the formation of their dataset, the authors put effort to raise the quality of the years of schooling data by minimising extrapolations and keeping the data as close as possible to those directly available from national censuses.

Our motivation to cross-country analysis comes from the fact that the problems of endogenoity and heteroscedasticity are well documented in cross-country growth literature (see e.g. Benhabib and Spiegel, 1994, and the surveys of Temple, 1999, and Durlauf et al., 2005).

Our parameter estimates of the input shares of human capital and physical capital are close to their theoretical values 2/3 and 1/3, respectively, when the endogeneity of inputs is controlled. Our results also indicate that the variation of standard deviations between the countries is large, which causes estimation inefficiency if not modelled. Finally, on the basis of the estimated heteroskedasticity parameter, it seems that the less-developed countries in 1970 had more variability in their growth possibilities over 1970-2000.

1.2.4 A Naïve Sticky Information Model of Households' Inflation Expectations

In essay 4, we provide a new model for the households' inflation expectations formation process. The model is based on Carroll's (2003) expectation formation model, where the general public adopt professionals' forecast with certain probability, rather than form their own rational forecasts. The structure of his model was inspired by simple models of disease spread from the epidemiology literature, and it provides promising microfoundations for sticky information models. To the best of our knowledge, it is also unique in relaxing the assumption that an ordinary person either knows the true probability distribution of the economy or can estimate some sophisticated econometric model when forming expectations. This relaxation is, however, important, since although trained economists might have this kind of knowledge, it would probably be an overwhelming task for an ordinary person (see Shiller, 1997). According to Carroll (2003), it might require, for example, obtaining a Ph.D. degree in economics first.

Our model differs from that of Carroll (2003) in that it no longer assumes agents to be 'infected' by rare newspaper forecasts. Rather, the source of 'infection' is the past release of annualized monthly inflation, the most commonly reported figure in the news coverage of inflation. This model is referred as the naïve sticky information model, and is motivated by the finding of Atkeson and Ohanian (2001) that since 1984 the one-year-ahead inflation forecast of professionals has not been better than the "naïve" forecast given by the inflation rate over the previous year. Furthermore, recent work has cast doubt on the reliability of traditional approaches to inflation forecasting (see e.g. Atkeson and Ohanian, 2001, Fisher et al., 2002, Sims, 2002a, Stock and Watson, 2002 and 2007, and Brave and Fisher, 2004). These findings give us a reason to question households' rationale to search for relatively rare newspaper forecasts or to form their own rational forecasts.

The model is estimated with both population- and household-level survey data from 1981/3 to 2001/4, constructed by the Survey Research Center (SRC) at the University of Michigan. In the population level, the naïve sticky information model is tested empirically against Carroll's sticky information approach. Specifically, we compare the posterior probabilities of the alternative models in which households update their expectations either to the forward-looking newspaper forecast or to the most recently reported past inflation statistic. As will be seen, Michigan data strongly support the latter.

The posterior model probabilities are computed, since the models are non-nested and conventional tests cannot be used to the test them against each other. There are of course classical alternatives for model comparison of non-nested models (such as the Akaike Information Criterion and Davidson-McKinnon J-test), but these 'tests' have no proper scale. The tested models are standard linear regression models corresponding to their theoretical counterparts. We assume uniform independent prior distributions on the interval [0,1] for parameters which economic theory suggests to be households' updating probabilities. Furthermore, for simplicity, uniform independent prior distributions on given intervals are also assumed for other parameters.

In the household-level analysis, we extend the agent-based epidemiology model, proposed by Carroll (2006), by deriving a relatively simple adaptation of that model, suitable for estimation. The model is derived, since there is likely to be heterogeneity in households' expectations that cannot be captured by the population level model. The marginal priors of model parameters are based on the underlying economic theory and on information derived from data sets not included in the estimation sample.

Our posterior evidence indicates that the agent-based epidemiology model captures the heterogeneity between agents' expectations fairly well, in the sense that the variance of unexplained heterogeneity (σ^2), i.e. heterogeneity in agents' expectations which the underlying model cannot explain, is quite small (approximately 1.1) relative to the high degree of heterogeneity observed in the actual micro level data. For example, in Branch's (2007) Rationally Heterogeneous Expectations (RHE) sticky information model, σ^2 was 36. Although our result is not fully comparable to that of Branch (2007), we note that, opposite to our model, in the RHE model most variation in agents' expectations is attributed to unexplained heterogeneity.

1.2.5 Aggregate Infrastructure Capital Stock and Long-Run Growth: Evidence from Finnish Data

In essay 5, we explore the relationship between infrastructure capital and long-run output growth in the time-series context. Exogenous and endogenous growth theories give substantially different predictions on this relationship. The conventional wisdom of the exogenous neoclassical model has been that government actions can have an effect on the income level but only a temporal effect on the growth rate, while the endogenous growth theory predicts that permanent changes in government policy, e.g. investments in infrastructure capital or human capital, can also have a permanent effect on the growth rate of output.

Earlier growth studies on productive public expenditure use different kinds of measures of the public capital stock (or subcomponents of the aggregate infrastructure capital stock) in their growth regressions. Instead, we use the Finnish land and water construction investment series from 1860 to 2003, and the perpetual-inventory method, to obtain a constant-euro value of the aggregate infrastructure capital stock for each year. This estimated capital series covers all important factors of infrastructure investment in Finland during that period.

Our estimations were based on the non-residential sector of the Finnish economy. The source of the investment series and the gross domestic product (GDP) series is Statistics Finland (Historical Series of Finnish National Accounts). All the endogenous variables are I(1) processes. We, however, find only very weak evidence in the data for a long-run relationship between the infrastructure capital, non-infrastructure capital and GDP series. This increases the complexity of reliable classical inference, decreasing its practicability. Fortunately, Bayesian methods allow us to easily draw an exact inference on the parameters even with near non-stationary data. At least Sims and Uhlig (1991), Sims and Zha (1999), and Bauwens, Lubrano and Richard (1999: 136) have discussed the usefulness of posterior analysis of near non-stationary data.

Boarnet (1995) reports a number of shortcomings in traditional infrastructure (public capital) studies. First, a common trend in both the output and infrastructure capital time series can lead to spurious correlation. Second, some of the productivity effect of infrastructure might spill over across regions, and thus regional panel studies do not measure the full effect of infrastructure. Third, infrastructure capital investment both causes and is caused by output growth, which means that the series are endogenous.

We avoid the reported shortcomings by using the Bayesian vector autoregressive (VAR) approach. Specifically, we use a structural-form Bayesian VAR to get a consistent parameter estimate for the long-run multiplier of output w.r.t. aggregate infrastructure capital. To acquire such a parameter we identify our Bayesian VAR using a Barro (1990) type growth model. Furthermore, in this essay, a standard non-informative prior is used, since our VAR system is very parsimonious and, hence, does not suffer from the over-parameterization problem.

Our analysis is close to that of Kocherlakota and Yi (1997); however, when we analyse a country where public infrastructure investments are typically less than 5% of total public expenditures (as is typical in OECD countries), we should not include a tax variable (like Kocherlakota and Yi, 1997), but instead include a private capital variable in the regression. If we estimate the model with infrastructure capital and the tax variable on the right-hand side, as is typical in a distributed-lag approach, only the revenue part of the government's budget is controlled. This biases the result towards exogenous growth. However, including a private capital variable allows us to control the effect of the agents' decisions on the allocation of investments between the infrastructure and other forms of capital.

The posterior evidence gives strong support for endogenous growth. However, failing to include the private capital variable in the regression biases the result towards exogenous growth. In general, our estimation results are in line with previous studies on the dynamic effects of infrastructure (public) capital. However, as we will show, when the dynamic effect of aggregate infrastructure capital (not public capital) is studied, one should put forth an effort to modelling in order to obtain consistent parameter estimates for the long-run coefficient of output w.r.t. infrastructure capital.

2 ESSAY 1: THE FORECASTING PERFORMANCE OF THE SMALL-SCALE HYBRID NEW KEYNESIAN MODEL[1]

ABSTRACT

This paper uses quarterly ex post and real-time U.S. data to show that the very simple hybrid New Keynesian model of Clarida, Galí and Gertler [1999. The Science of Monetary Policy: A New Keynesian Perspective. Journal of Economic Literature 37, 1661-1707] can provide forecasts comparable to those based on Bayesian reduced-form vector autoregressive models. The issue is important, since several recent papers have suggested different ways to improve the forecast performance of New Keynesian models at the cost of increasing the complexity of model mechanisms, thus reducing the practicability of these approaches.

Keywords: New Keynesian model; Forecasting; Real-time data; Bayesian inference; Vector autoregressive models.

---

1 This paper has been published in the Helsinki School of Economics Working Papers W-451. Jani Luoto is responsible for derivation of equations, most of the estimation, interpretation of the results, and most of the writing. Jan-Erik Antipin assisted in estimation, in interpreting the results and in some of the writing.

2.1 Introduction

There is an increasing volume of literature focusing on developing New Keynesian (NK) models suitable for forecasting and quantitative policy analysis (see Sungbae and Schorfheide, 2007, and references therein). Within this literature, Smets and Wouters (2003, 2005 and 2007), Christiano et al. (2005), Adolfson et al. (2007a), and Adolfson et al. (2005, 2008), construct large-scale NK models aiming to find a structural macroeconomic model which has a fit comparable to that of reduced-form Bayesian vector autoregressive (VAR) models. In these studies, additional shocks, frictions and measurement errors are introduced to the NK model mechanisms until the desired fit is achieved. This approach ignores model uncertainty, leading to inferences which are over-confident and decisions which are riskier than the policy-maker believes them to be. A promising alternative strategy is provided by Del Negro and Schorfheide (2004). In their approach an NK model is used to generate a prior distribution for the parameters of the VAR to improve the forecast and policy analysis performance of these models. Although this approach is promising, it is nonetheless complicated and the numerical methods required in estimation are time-consuming. The practicability of this approach can therefore be questioned (see also Del Negro et al., 2007). In the light of recent NK literature, it would be thus interesting to see whether a simple NK model, including only few shocks and the standard price rigidity, can have a fit comparable to other forecasting methods such as the Bayesian VARs commonly used as a benchmark.

This paper has two objectives. First, it provides a method for the Bayesian analysis of a simple hybrid NK model of Clarida et al. (1999). The method is very easy to implement and leads to savings in the CPU time required in posterior simulation, compared to the commonly used Kalman filter approach. Lindé (2005) estimates a version of the hybrid NK model with the full information maximum likelihood (FIML) method using U.S. data. We instead adopt a Bayesian full information framework, since the FIML estimates turned out to be very sensitive to starting values and since Bayesian methods allow incorporation of prior information which facilitates numerical maximization.

Our second objective is to compare the forecasting properties of the hybrid NK model against commonly used forecasting tools such as Bayesian VARs and naïve forecasts based on univariate random walks. Using quarterly U.S. data we show that the hybrid model can provide forecasts of key macroeconomic variables, inflation and short-term nominal interest rate, and a measure of the output gap comparable to forecasts based on reduced-form Bayesian VARs. Our results also indicate that the hybrid model predicts more accurately than naïve forecasts based on univariate random walks. In particular, these results hold for both ex post data and real-time data, which are available to policy-makers when forecasts are being made. Our results also confirm the finding of Smets and Wouters (2007) that the cross-equation restrictions implied by NK models work especially well in forecasting at medium-term horizons (from four to twelve quarters). For policy-makers, comparisons of forecasts at longer than one quarter horizon are of interest, since policy actions typically depend on expected future developments in the economy.

Finally, we find two major reasons for the good forecasting performance of the hybrid model. Firstly, the model allows both for the endogenous persistence in inflation and output and for the persistence of exogenous shock processes. This approach is commonly used in large-scale NK models, which forecast well. Secondly, our joint prior is well designed in allowing the parameters to be estimated fairly freely, while being sufficiently informative to keep the posterior distribution away from economically non-meaningful values.

The remainder of the paper is organized as follows. In section 2, we discuss the model, the prior and the data. We continue the analysis by reporting the posterior distributions of the parameters. In section 3, we explain the forecasting comparison methods, and present and discuss the results of a forecasting exercise. Section 4 concludes the paper.

2.2 Likelihood, Prior, Data, and Posterior

In this section we introduce a hybrid NK model. Its likelihood and the joint prior density function of the structural parameters are specified. We then describe the data and continue the analysis by reporting the posterior distributions of the parameters.

2.2.1 Model Likelihood

Let us consider the following hybrid NK model for period t inflation[2], π_t, and a measure of the output gap, x_t, respectively,

$$\pi_t = \alpha E_t \pi_{t+1} + (1-\alpha)\pi_{t-1} + \gamma x_t + \varepsilon_{\pi,t}\,, \tag{1}$$

$$x_t = \beta E_t x_{t+1} + (1-\beta)x_{t-1} - \beta_r (R_t - E_t \pi_{t+1}) + \varepsilon_{x,t}\,, \tag{2}$$

where parameters a and β satisfy the conditions $0 \leq a \leq 1$ and $0 \leq \beta \leq 1$. Equation (1) is the hybrid New-Keynesian Phillips curve (NKPC), similar to that analyzed in Rudd and Whelan (2006), while Equation (2) is the aggregate demand equation. The model is very close to that carefully studied in Clarida et al. (1999).

The disturbance terms $\varepsilon_{\pi,t}$ and $\varepsilon_{x,t}$ in Equations (1) and (2) are assumed to follow univariate AR(1) processes:

---

2 Price inflation is defined as the percent change in the price level from t – 1 to t.

$$\varepsilon_{\pi,t} = \rho_\pi \varepsilon_{\pi,t-1} + u_{\pi,t}, \tag{3}$$

$$\varepsilon_{x,t} = \rho_x \varepsilon_{x,t-1} + u_{x,t}, \tag{4}$$

where $\rho_\pi, \rho_x \in [-1,1]$, and $u_{\pi,t}$ and $u_{x,t}$ are independently and identically distributed (i.i.d.) random variables with zero means and variances σ_π^2 and σ_x^2, respectively.

We close the model with the following Taylor rule for the nominal interest rate R_t,

$$R_t = (1-\rho)(\gamma_\pi \pi_t + \gamma_x x_t) + \rho R_{t-1} + \varepsilon_{R,t}, \tag{5}$$

where the parameter $\rho \in [0, 1]$ measures the degree of interest rate smoothing, the disturbance term $\varepsilon_{R,t}$ obeys $\varepsilon_{R,t} = \rho_R \varepsilon_{R,t-1} + u_{R,t}$, $\rho_R \in [-1,1]$, and $u_{R,t}$ is an i.i.d. random variable with zero mean and variance σ_R^2.

The model in Equations (1)-(5) can be solved analytically by using standard first-order log-linear methods. In particular, this paper follows Lindé (2005) in applying the solution algorithm of Söderlind (1999). The solution gives the equilibrium law of motion for the relevant state variables. Specifically, the state equation is given by $z_t = Cz_{t-1} + v_t$, where $z_t = (\varepsilon_{\pi,t}, \varepsilon_{x,t}, \varepsilon_{R,t}, \pi_{t-1}, x_{t-1}, R_{t-1})'$, $v_t = (u_{\pi,t}, u_{x,t}, u_{R,t}, 0, 0, 0)'$ and C is a nonlinear function of structural parameters. Given that the shocks are normally distributed and that the vector of observables $y_t = (\pi_t, x_t, R_t)'$ is a linear combination of the state variables, the common approach is to specify a recursive likelihood function for the model using the Kalman filter. The estimates of the model can then be obtained using standard non-linear optimization methods.

Alternatively, the analytical solution of the model can be written as a full information system of the vector of observables (see Lindé, 2005). Specifically,

$$y_t = C_y y_{t-1} + C_\varepsilon \varepsilon_t, \tag{6}$$

where $\varepsilon_t = (\varepsilon_{\pi,t}, \varepsilon_{x,t}, \varepsilon_{R,t})'$ and C_y and C_ε are partitions of the solution matrix C conformably with y_t and ε_t, respectively.

Then denote $\varepsilon_t = \mathrm{P}\varepsilon_{t-1}$, where P is a diagonal matrix whose diagonal entries are given by ρ_π, ρ_x and ρ_R. The likelihood function for a sample of T observations can be written as

$$L(Y;\theta) \propto |C_\varepsilon|^{-T} |\Lambda|^{-T/2} \exp\{-0.5 \times tr(U'U\Lambda^{-1})\}, \tag{7}$$

where $\theta = (\beta, a, \beta_r, \gamma, \gamma_\pi, \gamma_x, \rho, \rho_\pi, \rho_x, \rho_R, \sigma_\pi, \sigma_x, \sigma_R)'$ is a vector comprising all model parameters and Λ a diagonal covariance matrix with diagonal entries

σ_π^2, σ_x^2 and σ_R^2. Furthermore, the *t*th rows of ($T\times m$) matrices Y and U are given by y_t' and u_t', respectively, where m is the number of observables and

$$u_t = C_\varepsilon^{-1}\left(y_t - \left(C_y + C_\varepsilon \mathrm{P} C_\varepsilon^{-1}\right)y_{t-1} + C_\varepsilon \mathrm{P} C_\varepsilon^{-1} C_y y_{t-2}\right). \quad (8)$$

In what follows, we adopt the full information approach of Equation (7), since the optimization algorithm based on it proved faster than the algorithm based on the recursive Kalman filter. Specifically, the Kalman filter approach requires roughly 4.5 times as much CPU time for posterior simulation as our approach (with a sample of 200 observations). Furthermore, both estimation methods were also found to produce similar results.

The model described in Equations (1)-(5) contains 13 parameters, collected in θ. It is fairly easy to see that all parameters are identifiable from the data. However, the maximum likelihood (ML) estimation of the model turned out to be a challenging task. In particular, the ML estimates of the parameters were very sensitive to the starting values of maximization due to a multimodal likelihood. This problem remained even when the parameter space was restricted to an economically feasible region. To illustrate this problem we give an example from the previous literature. Lindé (2005) estimates a version of the model in Equation (1)-(5) on U.S. data with full information maximum likelihood (FIML).[3] He finds positive and highly significant parameter estimates for the slope coefficients γ (≈ 0.05) and β_r (≈ 0.09). However, there exists a local equilibrium in which the likelihood is higher than that in Lindé's solution. At this equilibrium the slope coefficients γ and β_r are still positive, but rather close to zero. According to Lindé (2005), the estimation can with different starting values converge to local equilibria with more or less plausible parameter values. To facilitate numerical maximization, we suggest using Bayesian methods, which allow incorporation of prior beliefs on parameters. While restricting, for example, the slope coefficients γ and β_r to be equal to some theoretical values gives an example of a very strong prior belief, other kinds of beliefs cannot easily be considered in the classical framework.

As seen in current literature, Bayesian methods have become a standard workhorse in analysing the NK models. Sungbae and Schorfheide (2007) provide an excellent review of the Bayesian methods developed in recent years to estimate and evaluate this class of models (see also Adolfson et al., 2007b). Rather than elaborating the details of Bayesian methods in analysing the NK models, which is already done in Sungbae and Schorfeide (2007), we discuss our choices of marginal prior distributions in the next subsection.

---

3 Lindé (2005) adds additional lags in the aggregate demand equation (2) and the monetary policy rule (3) to make disturbance terms $\varepsilon_{x,t}$ and $\varepsilon_{R,t}$ white noise.

2.2.2 Marginal Priors

The starting-point in the Bayesian analysis is to determine the prior density function of the parameters, $p(\theta)$, which together with the likelihood function (7) yields the posterior density

$$q(\theta|Y)=\frac{p(\theta)L(Y;\theta)}{\int p(\theta)L(Y;\theta)d\theta}. \tag{9}$$

A typical informative prior reflects the researcher's subjective beliefs, summarizes information from the data not included in the estimation sample, or is based on both of them. Often the underlying economic theory provides a natural starting-point for the prior elicitation. We will use a very simple structural model as the basis of our prior knowledge. The model can be obtained by log-linearizing the aggregation of individual firms' pricing decisions and the consumption Euler equation without using ad hoc assumptions such as backward inflation indexation or habit formation in consumption. Specifically, the prior means of the parameters in θ are based on the following model,

$$\pi_t = bE_t\pi_{t+1} + \frac{(1-\kappa)(1-b\kappa)(1+\zeta)}{\kappa}x_t, \tag{10}$$

$$x_t = E_t x_{t+1} - (R_t - E_t\pi_{t+1}), \tag{11}$$

where b is the subjective discount factor, κ the frequency of price adjustment and ζ the elasticity of labor supply. Note that, for simplicity, a standard assumption on prior independence is used (see e.g. Zellner, 1971). Del Negro and Schorfheide (2008) criticize this assumption as having the drawback that the resulting joint prior distribution may assign a non-negligible amount of probability mass to regions of the parameter space where the model is unreasonable. It is fairly easy to see that the undesirable property suggested by Del Negro and Schorfheide (2008) is not present in our joint prior.

Table 1 lists the marginal prior distributions of the parameters. The beta prior distributions of the parameters a and β are concentrated towards unity, but are nonetheless only weakly informative (see Equations 10 and 11 for motivation). The prior mean of the slope coefficient β_r is set at unity, while the prior mean of γ (1.00) can be obtained by setting the subjective discount factor, the elasticity of labor supply and the frequency of price adjustment at their standard calibrated values, e.g. 0.99, 2 and 0.57, respectively, in Equation (10). The prior variances of these parameters (γ, β_r) are set to be small enough to keep the posterior distribution away from economically non-meaningful values. The prior means of the policy parameters γ_π (1.50) and γ_x (0.50) are obtained from

Taylor (1993).[4] However, some interest rate smoothing is also allowed a priori. That is, the prior mean of ρ is set at 0.50. With the given prior variances, the marginal prior distributions of these parameters (γ_π, γ_x, ρ) turned out to be practically noninformative.

The standard deviations σ_π, σ_x, and σ_R are assumed to follow inverse-gamma distributions with shape and scale parameters yielding fairly loose priors. Finally, the normal prior distribution with zero mean and $(3/4)^2$ variance is used for the transformed parameters

$$\phi_\pi = \frac{1}{2}\log\frac{1+\rho_\pi}{1-\rho_\pi}, \ \phi_x = \frac{1}{2}\log\frac{1+\rho_x}{1-\rho_x} \text{ and } \phi_R = \frac{1}{2}\log\frac{1+\rho_R}{1-\rho_R}. \qquad (12)$$

These marginal priors force the posterior distributions of the autoregressive parameters ρ_π, ρ_x and ρ_R to be located in the interval [–1, 1]. The marginal priors are also very loose, but nevertheless turned out to improve simulation efficiency.

2.2.3 Data and Results

Throughout this study the quarterly U.S. data from 1953:2 to 2004:4 are used. In addition to the entire sample, the models are estimated for the subsample periods 1953:2-1982:2 and 1982:3-2004:4, capturing the "Great Inflation" and "Great Moderation" periods, respectively. This serves as a convenient check for robustness and parameter constancy. We are aware that the nominal interest rate, as the instrument of monetary policy, provides a reasonable description of the Federal Reserve's operating procedures only after 1964 (see Clarida et al., 1999). However, the first ten years of data are required to have a sufficiently long out-of-sample forecasting period. We form out-of-sample forecasts from 1976:4 to 2004:4 to have forecast series which covers a diverse spectrum of inflation volatility.

The output gap is measured as a logarithmic difference between the actual and the potential output level. Two measures of actual output are used: real gross domestic product (GDP) and non-farm business (NFB) sector output. The logarithm of the potential output is proxied by the one-sided Hodrick-Prescott (HP) trend estimate in the model

$$g_t = \tau_t + \eta_{1t}, \qquad (13)$$

$$(1 - L)^2\tau_t = \eta_{2t}, \qquad (14)$$

---

4 In Taylor (1993), the interest rate and the inflation rate are expressed on a yearly basis. Since we express them on a quarterly basis, the prior mean of γ_x should be set at 0.125 (0.5 divided by 4). However, the standard deviation of the measure of the output gap used in Taylor (1993) is markedly higher than the standard deviation of the measure of the output gap used in this paper. Thus, the prior mean of 0.5 can be seen to be justified in our case.

where g_t is the logarithm of the measure of actual output, L is a lag operator and η_{1t} and η_{2t} are mutually uncorrelated white noise sequences with the relative variance $q = \text{var}(\eta_{1t})/\text{var}(\eta_{2t})$. The value of $q = 0.67\times10^{-3}$ is taken from Stock and Watson (1999). We use the previous approximation of potential output, since our focus is on forecasting and since it does not use the future values of the detrended variable, as the optimal two-sided trend extraction HP-filter for Equations (13) and (14) does.[5] Furthermore, Stock and Watson (1999) find, after experimenting with several methods suitable for forecasting, that this procedure produces plausible gap estimates which work fairly well in inflation forecasting.

The price inflation is measured as the log difference of the Implicit Price Deflator of GDP (NFB). All the series are seasonally adjusted. The source of the final vintage data is the FREDII databank of the Federal Reserve Bank of St. Louis, while that of the real-time data is the Federal Reserve Bank of Philadelphia. The Federal Funds rate (FFR) is used as the instrument of monetary policy. The nominal interest rate and inflation rate series are measured as quarterly changes corresponding to their appearance in the structural model. Finally, the data are demeaned prior to estimation.

The estimation results are presented in Table 1, in the topmost panel (A) for the entire sample and in the lower panels (B and C) for the two subsample periods.[6] The data appear to be particularly informative in all these samples.

---

5 We also tested for detrending linear and quadratic trend methods which are suitable for forecasting, and found that the results presented are not sensitive to using these measures of potential output. Furthermore, we ran several regressions with the dataset used in Lindé (2005). The results of the regressions with our and his datasets were quite similar.

6 To generate a Monte Carlo sample from the posterior of θ we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MMCMC). The algorithm uses a multivariate normal distribution for the jump distribution on changes in θ. Our simulation procedure was as follows: we first simulated 10,000 draws using a diagonal covariance matrix with diagonal entries 0.00001 in the jump distribution. We then used these draws to estimate the posterior covariance matrix of θ and scale it by the factor $2.4^2/13$, to obtain an optimal covariance matrix for the jump distribution; see e.g. Gelman et al. (2004). We continue by a simulating 10,000 draws and calculated a more accurate covariance matrix for θ. We repeated this roughly 2 times. We then added noise to the posterior median to obtain overdispersed starting values and simulated three chains of length 30,000. We excluded the first 5000 simulations as a burn-in period in each chain and picked out every 25th draw from the Markov chain, yielding a sample of 3000 draws, which economizes on storage space and reduces autocorrelation across draws. The convergence of the chains was checked using Gelman and Rubin's convergence diagnostic R (also called 'potential scale reduction factor') (see Gelman and Rubin, 1992). The diagnostic values close to 1 indicate approximate convergence and values smaller than 1.1 are acceptable in most cases. In our case the diagnostic was estimated to be between 1.01 and 1.03 for all parameters and all models. The multivariate version of Gelman and Rubin's diagnostic proposed by Brooks and Gelman (1998) was between 1.01 and 1.02 for each model; the convergence was thus fairly good. The frequencies of accepted jumps were roughly 0.21. Finally, the previous adaptive Metropolis algorithm is used because the covariance matrix estimate based on the local behaviour of the posterior at its highest peak turned out to give too optimistic a view of precision, and thus failed to yield an efficient covariance matrix for the normal jump distribution.

That is, the variances of the posterior distributions are found to be systematically smaller than the prior variances. The posteriors are also relatively stable between the data sets and the subsamples with two exceptions. The variances of the stochastic error processes seem to have fallen in the second subsample period. Sims and Zha (2006) and Smets and Wouters (2007) find similar evidence in U.S. data concerning the variance of monetary policy shocks. Our results also indicate that the Federal Reserve seemed to respond to the output gap and inflation more strongly during the second subsample period. The latter result is in accordance with that of Boivin and Giannoni (2006) and Smets and Wouters (2007), while Bernanke and Mihov (1998), Leeper and Zha (2003) and Canova (2006) find a relatively stable interest rate rule for the post WWII sample.

The Taylor principle is fulfilled in all the samples. This contradicts Clarida et al. (2000), who report that the Federal Reserve responded less than one-to-one to inflation during the period 1960-1979 (pre-Volcker period), thus violating the Taylor principle. In line with our result, for example, Smets and Wouters (2007) and Rabanal and Rubio-Ramírez (2005) find the inflation coefficient to be greater than one.

The point estimate of a (0.08) indicates a very insignificant role for the forward-looking behavior in the Phillips curve. This result is in accordance with those of Fuhrer (1997), Lindé (2005) and Rudd and Whelan (2006), but at odds with the results of Smets and Wouters (2003, 2005 and 2007), Adolfson et al. (2005) and Galí et al. (2005). The latter group of authors obtains relatively low parameter estimates for the degree of price indexation. Our estimates were obtained using a statistical measure of the output gap. Galí and Gertler (1999) and Galí et al. (2005) have suggested that the key reason for the lack of success of the forward-looking NKPC is that the detrended output is not a good proxy for real marginal costs. Contrary to their finding, Rudd and Whelan (2006), who used both the output gap and labor's share as a proxy for real marginal cost, found that the evidence for the forward-looking behavior in the NKPC was very weak.

The point estimates of β are high, supporting the traditional forward-looking intertemporal Euler equation. In contrast, previous studies have typically observed a high degree of habit persistence (see e.g. Christiano et al., 2005, and Smets and Wouters, 2007). However, there seems to be a trade-off between the forward-looking behavior of the demand equation and the persistence of autoregressive demand shocks. In our paper, the high autoregressive parameter of the exogenous shock process (ρ_x = 0.79) takes into account the degree of persistence observed in the data. In Smets and Wouters (2007), the habit formation of consumption takes into account the high persistence, while the autoregressive parameter of exogenous shocks is estimated to be relatively small (0.36). Smets and Wouters (2007), however, assume a high habit parameter 0.7 (with 0.01 prior variance), a priori.

TABLE 1 Priors and Posteriors of the hybrid NK model

	Prior Distr.			Posterior Distr. (GDP)			Posterior Dist. (NFB)		
				Panel A: Sample 1954:2 – 2004:4					
Par.	Distr.	Mean	St.Dev.	Median	5%	95%	Median	5%	95%
a	Beta	0.67	0.24	0.08	0.02	0.19	0.08	0.02	0.18
γ	Gamma	1.00	0.32	0.03	0.02	0.05	0.03	0.02	0.05
β	Beta	0.67	0.24	0.75	0.65	0.84	0.74	0.65	0.83
β_r	Gamma	1.00	0.32	0.10	0.05	0.16	0.12	0.07	0.20
γ_π	Gamma	1.5	0.61	1.82	1.50	2.32	1.82	1.46	2.34
γ_x	Gamma	0.5	0.35	0.59	0.40	0.87	0.49	0.33	0.76
ρ	Beta	0.5	0.22	0.87	0.83	0.91	0.89	0.85	0.92
ρ_π	Normal	0	0.54	-0.38	-0.46	-0.28	-0.42	-0.50	-0.33
ρ_x	Normal	0	0.54	0.79	0.67	0.87	0.78	0.66	0.86
ρ_R	Normal	0	0.54	0.12	-0.00	0.24	0.11	-0.00	0.24
σ_π	Invgam.	0.40	3.96	0.29	0.26	0.32	0.34	0.31	0.37
σ_x	Invgam.	0.40	3.96	0.16	0.12	0.20	0.21	0.16	0.28
σ_R	Invgam.	0.40	3.96	0.22	0.20	0.24	0.22	0.20	0.24
				Panel B: Sample 1954:2 – 1982:2					
a	Beta	0.67	0.24	0.08	0.02	0.20	0.08	0.02	0.21
γ	Gamma	1.00	0.32	0.05	0.03	0.07	0.05	0.03	0.07
β	Beta	0.67	0.24	0.79	0.66	0.94	0.77	0.66	0.92
β_r	Gamma	1.00	0.32	0.19	0.11	0.32	0.21	0.12	0.35
γ_π	Gamma	1.5	0.61	1.86	1.46	2.46	1.81	1.41	2.47
γ_x	Gamma	0.5	0.35	0.52	0.29	0.86	0.47	0.25	0.77
ρ	Beta	0.5	0.22	0.84	0.78	0.90	0.87	0.80	0.92
ρ_π	Normal	0	0.54	-0.35	-0.48	-0.21	-0.41	-0.52	-0.28
ρ_x	Normal	0	0.54	0.77	0.60	0.87	0.76	0.60	0.87
ρ_R	Normal	0	0.54	0.11	-0.06	0.29	0.10	-0.06	0.27
σ_π	Invgam.	0.40	3.96	0.34	0.30	0.39	0.41	0.36	0.46
σ_x	Invgam.	0.40	3.96	0.24	0.17	0.34	0.32	0.22	0.43
σ_R	Invgam.	0.40	3.96	0.28	0.25	0.31	0.28	0.25	0.32
				Panel C: Sample 1982:3 – 2004:4					
a	Beta	0.67	0.24	0.08	0.02	0.20	0.08	0.02	0.20
γ	Gamma	1.00	0.32	0.05	0.03	0.08	0.04	0.03	0.06
β	Beta	0.67	0.24	0.83	0.70	0.97	0.86	0.73	0.98
β_r	Gamma	1.00	0.32	0.19	0.11	0.32	0.23	0.13	0.37
γ_π	Gamma	1.5	0.61	2.65	1.93	3.64	2.63	1.88	3.75
γ_x	Gamma	0.5	0.35	0.89	0.57	1.35	0.69	0.42	1.08
ρ	Beta	0.5	0.22	0.90	0.86	0.93	0.91	0.87	0.94
ρ_π	Normal	0	0.54	-0.35	-0.50	-0.17	-0.37	-0.50	-0.22
ρ_x	Normal	0	0.54	0.88	0.79	0.94	0.88	0.80	0.94
ρ_R	Normal	0	0.54	0.29	0.11	0.47	0.34	0.16	0.52
σ_π	Invgam.	0.40	3.96	0.22	0.19	0.25	0.24	0.20	0.28
σ_x	Invgam.	0.40	3.96	0.09	0.06	0.12	0.11	0.08	0.16
σ_R	Invgam.	0.40	3.96	0.12	0.10	0.13	0.11	0.10	0.13

Finally, the persistence of monetary policy shocks (ρ_R) is relatively low and equal to that estimated by Smets and Wouters (2007).

2.3 Forecast Comparison

In this section we first discuss the forecasting methods. We then provide some details for the forecasting comparison methods. Finally, we report the results of a forecasting exercise.

2.3.1 Measuring the Prediction Performance of Competitive Models

It is fairly easy to see that Equation (6) can be treated as a reduced-form VAR with lag-length 2 and normally distributed errors with covariance matrix $\Sigma = C_\varepsilon \Lambda C_\varepsilon'$. Thus, the conditional predictive distribution of Equation (6) for the joint lead time 1 through H, $p(y_{t+1},\ldots,y_{t+H} \mid Y,\theta)$, is multivariate normal (see Lütkepohl, 1993). This facilitates straightforward simulations from $p(y_{t+1},\ldots,y_{t+H} \mid Y,\theta)$, given the posterior p.d.f. of θ. The method for obtaining the posterior p.d.f. of θ was explained in the previous section.[7]

The predictive performance of the hybrid NK model is compared to two Bayesian VARs and to naïve forecasts based on univariate random walks. The VAR systems consist of the same three variables, $y_t = (\pi_t, x_t, R_t)'$, as the hybrid NK model. The data are not however demeaned prior to estimation. Diffuse and Normal-Diffuse priors are used for the parameters of the VAR models (see Kadiyala and Karlsson, 1997) for discussion. Parameterization of the Normal-Diffuse prior is based on the assumption that the variables behave as if they had random walk components (see Litterman, 1980). That is, the prior means are set at zero except for the elements corresponding to the first own lag of each variable. The prior variances of the parameters in the ith equation of a p-lag VAR ($k = 1,\ldots, p$)[8] are given by π_1/k, $\pi_2 s_i^2/s_j^2 k$ ($i \neq j$) and $\pi_3 s_i^2$, for the parameters on own lags, foreign lags and a constant, respectively (see Litterman, 1986, and Kadiyala and Karlsson, 1997, for the motivation of this prior variance specification). A scale factor accounting for the different scales of the variables,

---

7 In a recursive forecast exercise, a total of 113 chains were simulated from each model. The posterior estimates of θ are based on 30,000 draws. The first 6,000 draws were discarded as a burn-in period. To reduce the size of output files, every 12th draw was saved. The predictive likelihoods are thus computed on the basis of 2000 draws from the Markov chain. Geweke (1992) proposed a convergence diagnostic for Markov chains based on a test for the equality of means of the first and last parts of the chain (in this paper the first 10% and the last 50% of observations were used). The test statistic is a standard Z-score; the difference between the two sample means divided by its estimated standard error. The standard error is estimated from the spectral density at zero and so takes into account any autocorrelation. The hypothesis of the equality of means was not rejected for most parameters at the 5 % significance level.

8 In our paper, p is set at 4. The fractional marginal likelihoods (FML) of Villani (2001), which were used in preliminary data analysis, supported this choice in over 99% of the estimated regressions.

s_i^2, is set at the residual standard error of equation *i*. The relative tightness of the prior is set at the commonly used values of hyper-parameters, π_1 = 0.05 and π_2 = 0.005 (see e.g. Kadiyala and Karlsson, 1997, and Litterman, 1986). The tightness of the constant terms is set at π_3 = 0.05, which shrinks the processes towards driftless univariate random walk. This prior specification provides a suitable description for the processes of inflation, nominal interest rate and detrended output. The posterior distributions were simulated using the Gibbs sampling algorithm[9] of Kadiyala and Karlsson (1997) for the Normal-Diffuse prior specification and the matricvariate Student's *t* distribution for the Diffuse prior specification. The predictive likelihoods were computed on the basis of 2,000 draws from the posteriors.

The forecasting performance of the models is examined using the standard recursive forecast procedure, which entails making forecasts using data dated before the forecasting period. The forecasting procedure is as follows: using data up to a given time point *T* all the parameters in the model are estimated and the predictive distribution over $y_{T+1},\ldots,y_{T+H}$ is computed.[10] Moving forward one period, all the parameters are re-estimated and the forecast distribution of $y_{T+2},\ldots,y_{T+H+1}$ is computed. This is continued until no more data are available to compute the one-step-ahead forecast errors. The period over which the dynamic forecast distributions are computed in this manner is 1976:4 through 2004:4. In addition to the entire forecasts sample, the forecasts are also compared for the subsample period 1990:1-2004:4 (the sample period of Smets and Wouters, 2007). This serves as a check of robustness of the results and increases the comparability of our results to those in previous literature; especially in the paper of Smets and Wouters (2007).

Adolfson et al. (2007a) recommend use of several univariate and multivariate measures to determine the accuracy of the point forecasts. The two commonly used univariate measures of accuracy, the root mean squared forecast error (RMSE) and the mean absolute forecast error (MAE) are computed as

$$RMSE_i(h) = \sqrt{N_h^{-1} \sum_{t=T}^{T+N_h-1} e_{i,t}^2(h)}, \tag{15}$$

$$MAE_i(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \left| e_{i,t}(h) \right|, \tag{16}$$

---

9 2,200 draws were simulated and the first 200 draws from the Markov chain were neglected as a burning period.

10 Note also that when the forecasts are evaluated the data is demeaned and the gap estimates are computed using the data up to time *T*. Furthermore, when the analysis is based on demeaned data, the posterior median forecasts are computed and the means are added to the median forecasts.

respectively, where $e_{i,t}(h) = y_{i,t+h} - \hat{y}_{i,t+h|t}$ is the *i*th element of the *h*-step-ahead forecast error, $\hat{y}_{t+h|t}$ the *h*-step-ahead posterior median forecast of y_{t+h} and N_h the number of the *h*-step-ahead forecasts (h = 1,…,H). However, only the RMSEs are reported, since these two measures turned out to produce equal results.

Two multivariate accuracy measures of point forecast, the log determinant statistic and the trace statistic, are also used in addition to the univariate measures. The multivariate statistics are based on the scaled *h*-step-ahead mean squared error (MSE) matrix

$$T_M(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \bar{e}_t(h) \cdot \bar{e}_t{}'(h) \,, \tag{17}$$

where $\bar{e}_t(h) = M^{-1} e_t(h)$ and M is a scaling matrix accounting for the different scales of the variables being forecasted.[11] As discussed in Adolfson et al. (2007a), the forecasting performance of the least predictable dimensions, that is, those corresponding to the highest eigenvalues of the square matrix $T_M(h)$, mainly determine the trace statistic $\mathrm{tr}[T_M(h)] = \lambda_1 + \ldots + \lambda_m$, while the log determinant statistic $\log |T_M(h)| = \log \lambda_1 + \ldots + \log \lambda_m$ also takes into account the forecasting performance of the most predictable dimensions (the lowest eigenvalues). It is also obvious that when the lowest eigenvalue of $T_M(h)$ approaches zero, the most predictable dimension determines the log determinant statistic.

Finally, in view of the increasing interest for forecast uncertainty, we also compare the prediction performance of the competitive models using the log predictive density score (LPDS), which is a measure of the accuracy of multivariate density forecasts (see Adolfson et al., 2007a). To be more concrete, let $\hat{y}_{t+h|t}$ and $\Omega_{t+h|t}$ denote the posterior mean and covariance matrix of the *h*-step-ahead forecast distribution $p_t(y_{t+h})$. Then, under the normality assumption of $p_t(y_{t+h})$, the LPDS of the *h*-step-ahead predictive density at time *t* is defined as

$$\begin{aligned} S_t(y_{t+h}) &= -2 \log p_t(y_{t+h}) \\ &= m \log(2\pi) + \log\left|\Omega_{t+h|t}\right| + \left(y_{t+h} - \hat{y}_{t+h|t}\right)' \Omega_{t+h|t}^{-1} \left(y_{t+h} - \hat{y}_{t+h|t}\right). \end{aligned} \tag{18}$$

We report the averages of the LPDSs over the evaluated *h*-step-ahead forecasts,

$$S(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} S_t(y_{t+h}). \tag{19}$$

This measure takes into account the forecasting performance of the predictive density as a whole.

---

11 We follow Adolfson et al. (2007a) and set M equal to the diagonal of the sample covariance matrix of the y_t from 1976:4 to 2004:4 (1990:1 to 2004:4).

2.3.2 Results

Figures 1-3 summarize the forecasting performance of the competitive models. Specifically, Figure 1 reports the RMSEs in quarterly percentage terms, Figure 2 the log determinant and the trace statistics, and Figure 3 the averages of the LPDS statistic. Figures 4-6 gives the corresponding statistics for the forecasts based on real-time data. The results based on the NFB data were similar to those based on the GDP data and in order to save space we report only the latter. All the statistics are reported at the 1 to 12 quarters horizons. [12] In the figures, a small value favors the model.

A few key findings emerge from the figures. Firstly, although the models are very simple they seem to forecast particularly well. According to the RMSEs, the small-scale models appear to produce more accurate point forecasts, on both inflation and the Federal Funds rate,[13] than the large-scale Bayesian VAR of Smets and Wouters (2007). In addition, the models turned out to produce real-time inflation forecasts which outperformed the naïve forecasts up to six quarters in the 1990:1-2004:4 subsample (see Figure 4). This result gives some perspective to the forecast accuracy of the hybrid model, when we take into account the finding of Atkeson and Ohanian (2001) that the one-year-ahead Federal Reserve's Greenbook inflation forecast has not been better on average than the naïve forecast since 1984.

Secondly, all the forecast comparison methods appear to yield similar conclusions. In the entire sample the forecasts of the hybrid model outperform those of the Bayesian VARs, while in the low inflation subsample (1990:1-2004:4) all the multivariate forecasting methods seem to produce equally accurate forecasts. Thus, the restrictions (stationary and cross-equation) implied by the hybrid model appear to help in forecasting especially well during high inflation periods. According to the univariate and multivariate measures of forecast accuracy, this result is most obvious at medium-term horizons. One exception is the nominal interest rate. The hybrid model forecasts this series very well in all samples and forecasting horizons (see Figure 1 and 4). In particular, all these results hold for both ex post data and real-time data.

Taking a closer look at the figures we see that the hybrid model is superior to the naïve forecasts at all samples and horizons, except for the longer horizon inflation forecasts in the low inflation subsample. In this subsample, the Bayesian VARs also give slightly better inflation and output gap forecasts than the hybrid model, according to the RMSEs. However, the improvement in the predictability of the variables is clearly negligible.

It also seems that the shrinking prior does not improve the forecasting performance of VARs in terms of point forecasting accuracy. This is not surprising, since the VAR systems are particularly parsimonious and, hence, do

---

12 We do not report the marginal likelihood, since it captures only the one-step-ahead predictive performance of the full model and is therefore too restricted for forecasting comparison.

13 The GDP forecasts are not directly comparable to results of Smets Wouters (2007), since they use the log difference of GPD series, while we use the GDP gap series.

not suffer from the over-parameterization problem. However, the LPDS statistics (see Figures 3 and 6) support a slightly better forecasting density for the Normal-Diffuse prior specification in the low inflation subsample. Over the entire sample the LPDSs support Bayesian VARs at the shorter forecasting horizons (1 to 4 quarter); however, the hybrid model again outperforms the VARs at the longer horizons.

In sum, it seems fair to say that the simple hybrid NK model captures the predictable behavior of the three U.S. key macroeconomic variables very well. The reason for its good forecasting performance may be that the model allows both for the endogenous persistence in inflation and output and for the persistence of the exogenous shock processes. This approach is commonly used in large-scale NK models, which forecast well. Our joint prior is also well designed in allowing the parameters to be estimated fairly freely, while being informative enough to keep the posterior distribution away from economically non-meaningful values.

2.4 Conclusion

Several recent papers have suggested different ways to improve the forecast performance of New Keynesian models. Unfortunately, improvement in fit is achieved at the cost of increasing the complexity of model mechanisms, which reduces the practicability of these approaches. This paper, in contrast, has shown that the very simple hybrid New Keynesian model of Clarida et al. (1999) can provide forecasts comparable to those based on commonly used benchmark models such as reduced-form Bayesian VARs and univariate random walks.

Our forecasting evidence indicates that the restrictions implied by the hybrid model work especially well in high inflation regimes. According to several univariate and multivariate measures of forecast accuracy, the forecasts of the hybrid model outperform those of the Bayesian VARs when high inflation periods are forecast. In the low inflation forecast subsample, the methods produce equally accurate forecasts. One exception was the nominal interest rate. The hybrid model seems to forecast this series very well in all samples and horizons. The hybrid model also predicts more accurately than the naïve forecasts based on univariate random walks. Finally, we note that all these findings hold for both ex post and real-time data.

FIGURES

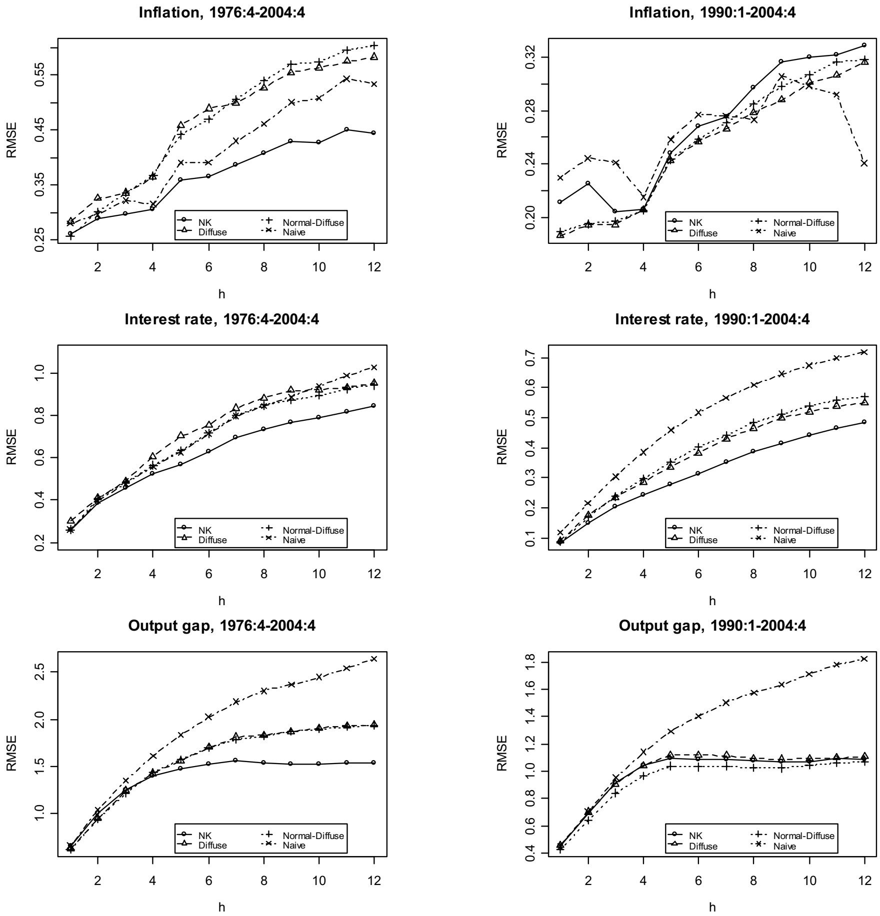

FIGURE 1 Root mean squared forecast errors for the competitive models

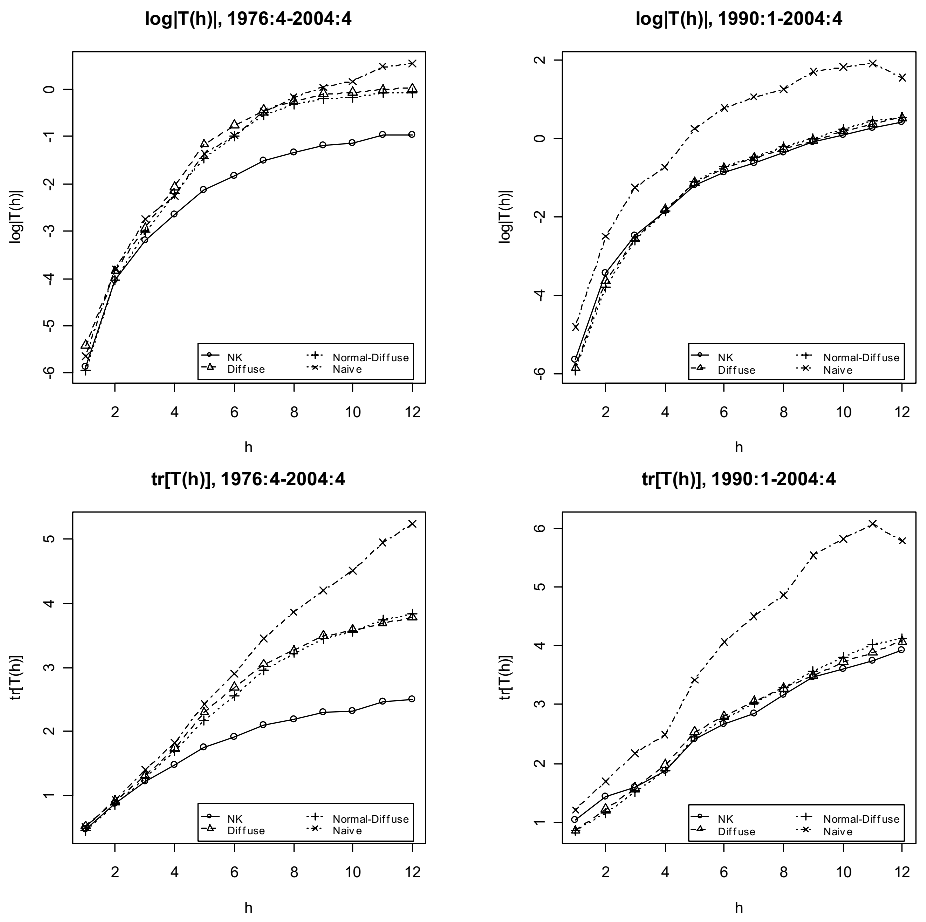

FIGURE 2 Log determinant statistics and trace statistics for the competitive models

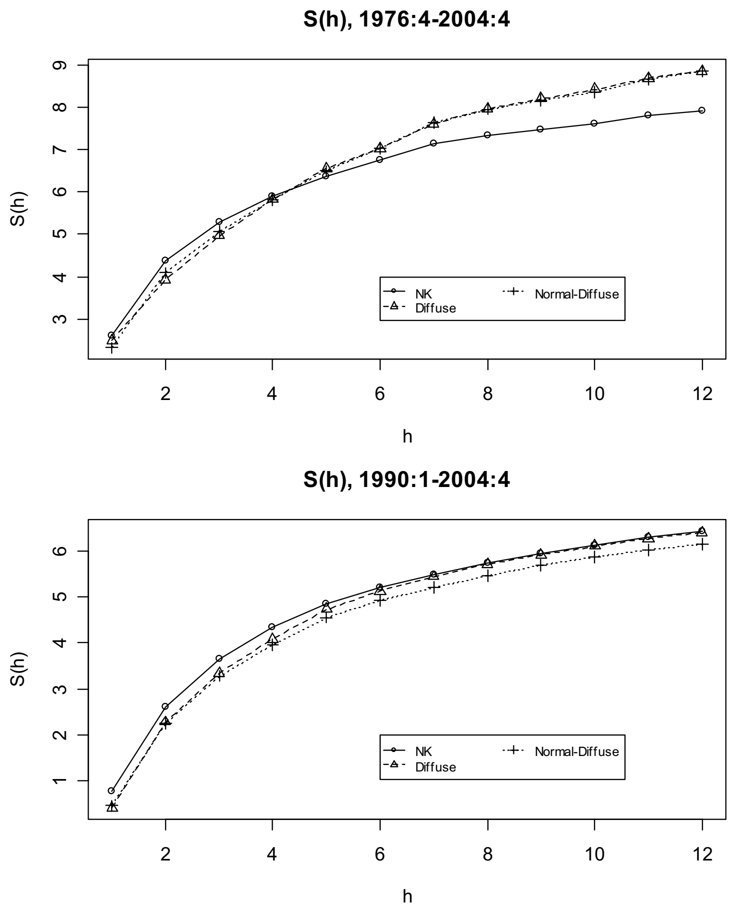

FIGURE 3 The average log predictive density scores for the competitive models

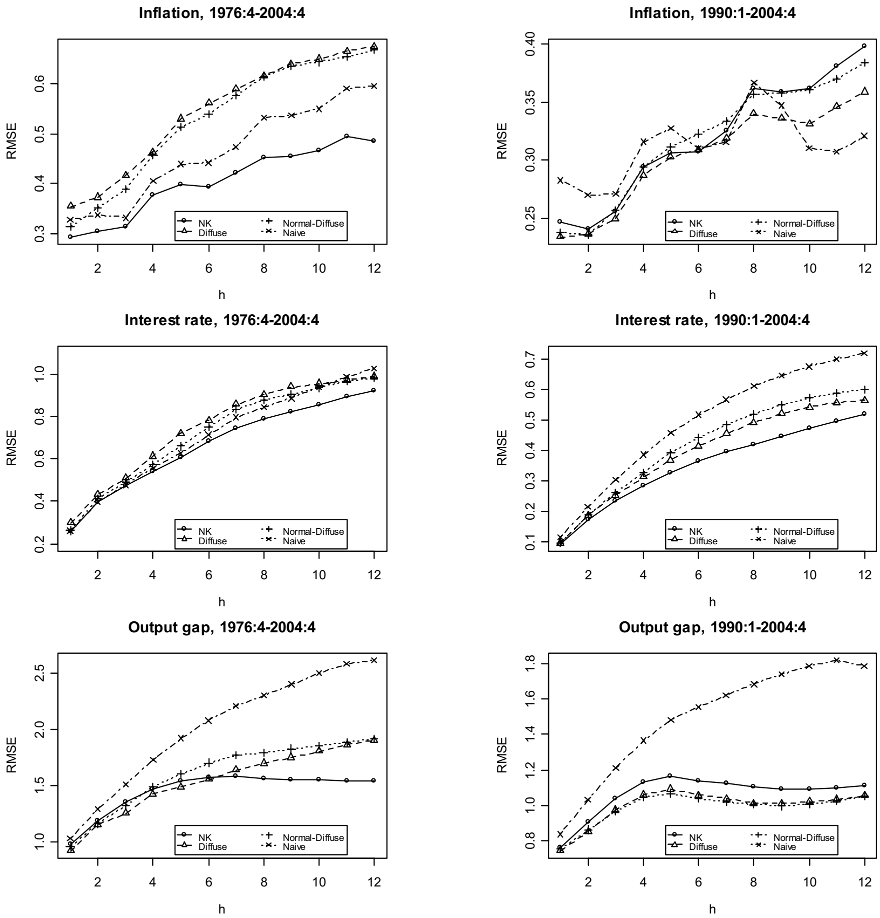

FIGURE 4 Root mean squared forecast errors for the competitive models (real-time data)

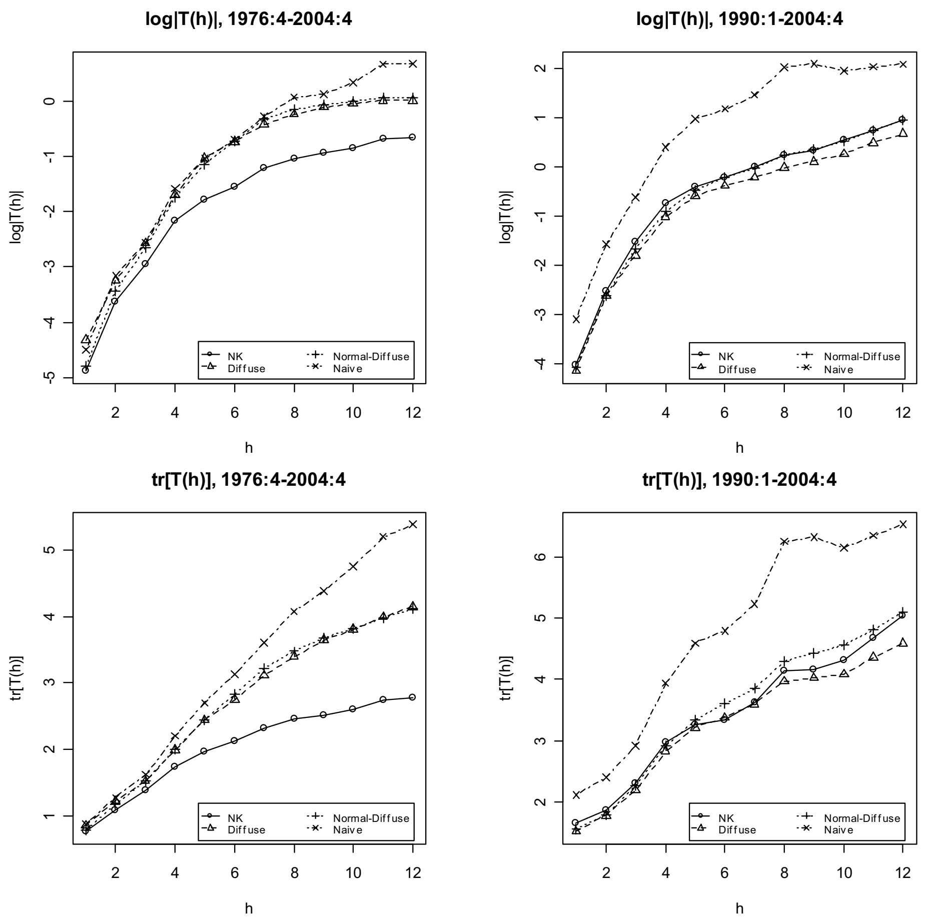

FIGURE 5 Log determinant statistics and trace statistics for the competitive models (real-time data)

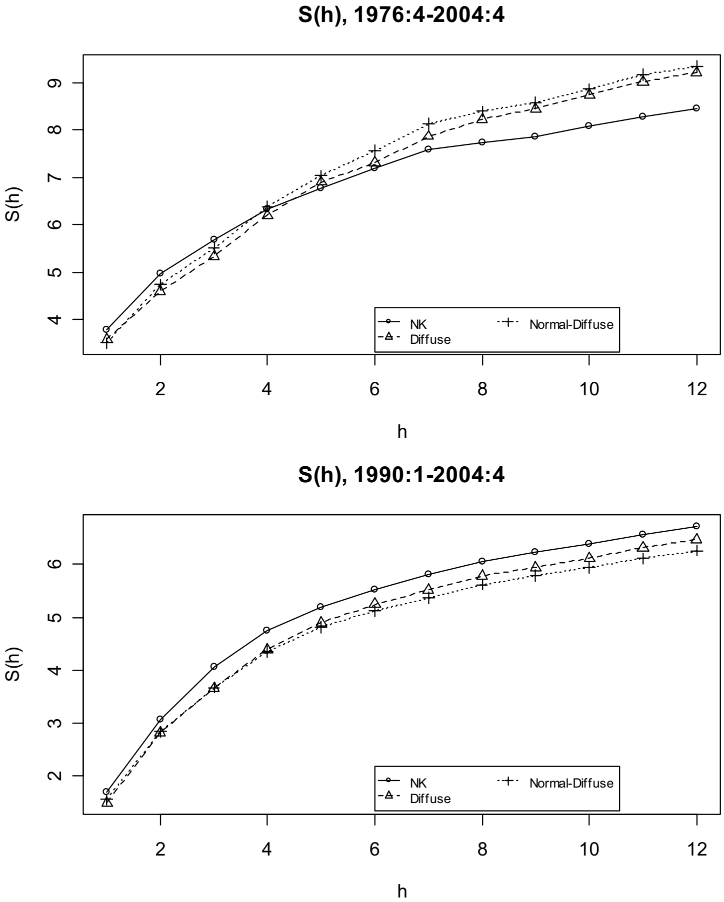

FIGURE 6 The average log predictive density scores for the competitive models (real-time data)

REFERENCES

Adolfson, M., Laséen, S., Lindé, J. and Villani, M., 2008. Evaluating an Estimated New Keynesian Small Open Economy Model. *Journal of Economic Dynamics and Control*. Forthcoming.

Adolfson, M., Lindé, J. and Villani, M., 2007a. Forecasting performance of an open economy DSGE model. *Econometric Reviews* 26, 289–328.

Adolfson, M., Laséen, S., Lindé, J. and Villani, M., 2005. The Role of Sticky Prices in an Open Economy DSGE Model: A Bayesian Investigation. *Journal of the European Economic Association* 3, 444–457.

Adolfson, M., Lindé, J. and Villani, M., 2007b. Bayesian Inference in DSGE Models - Some Comments. *Econometric Reviews* 26, 173–185.

Atkeson, A. and Ohanian, L.E., 2001. Are Phillips curves useful for forecasting inflation?. *Federal Reserve Bank of Minneapolis Quarterly Review* 25, 2–11.

Bernanke, B. and Mihov, I., 1998. Measuring Monetary Policy. *The Quarterly Journal of Economics* 113, 869–902.

Boivin, J. and Giannoni, M.P., 2006. Has Monetary Policy Become More Effective?. *The Review of Economics and Statistics* 88, 445–462.

Brooks, S.P. and Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434–455.

Canova, F., 2006. Monetary Policy and the Evolution of the US Economy. CEPR Discussion Papers No. 5467.

Christiano, L.J., Eichenbaum, M. and Evans, C.L., 2005. Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* 113, 1–46.

Clarida, R., Galí, J. and Gertler, M., 1999. The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature* 37, 1661–1707.

Clarida, R., Galí, J. and Gertler, M., 2000. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115, 147–180.

Del Negro, M., and Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs. *International Economic Review* 45, 643–673.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R., 2007. On the fit of New-Keynesian Models. *Journal of Business & Economic Statistics* 25, 143–162.

Del Negro, M., and Schorfheide, F., 2008. Forming Priors for DSGE Models. Manuscript.

Fuhrer, J.C., 1997. The (un)importance of forward-looking behavior in price specifications. *Journal of Money, Credit and Banking* 29, 338–350.

Galí, J., and Gertler, M., 1999. Inflation Dynamics: A Structural Econometric Analysis. *Journal of Monetary Economics* 44, 195–222.

Galí, J., Gertler, M. and López-Salido, J.D., 2005. Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve. *Journal of Monetary Economics* 52, 1107–1118.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004. *Bayesian Data Analysis 2nd edition*, Boca Raton: Chapman & Hall/CRC.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments., in José M. Bernardo, James O. Berger, A. Philip David and Adrian F.M. Smith (Eds.), *Bayesian Statistics*, Vol. 4 (Oxford: Oxford University Press, UK, pp. 169–193.

Kadiyala, K.R. and Karlsson, S., 1997. Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics* 12, 99–132.

Leeper, E. and Zha, T., 2003. Modest Policy Interventions. *Journal of Monetary Economics* 50, 1673–1700.

Lindé, J., 2005. Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach. *Journal of Monetary Economics* 52, 1135–1149.

Litterman, R.B., 1980. A Bayesian Procedure for Forecasting with Vector Autoregressions. Mimeo, Massachusetts Institute of Technology.

Litterman, R.B., 1986, Forecasting with Bayesian vector autoregressions - Five years of experience. *Journal of Business and Economic Statistics* 4, 25–38.

Lütkepohl, H., 1993. Introduction to Multiple Time Series Analysis, 2nd edition, Springer-Verlag, New York.

Rabanal, P. and Rubio-Ramírez, J.F., 2005. Comparing New Keynesian models of the business cycle: A Bayesian approach. *Journal of Monetary Economics* 52, 1151–1166.

Rudd, J. and Whelan, K., 2006. Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?. *American Economic Review* 96, 303–320.

Sims, C. and Zha, T., 2006. Were There Regime Switches in U.S. Monetary Policy?. *American Economic Review* 96, 54–81.

Smets, F. and Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association* 1, 1123–1175.

Smets, F. and Wouters, R., 2005. Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach. *Journal of Applied Econometrics* 20, 161–183.

Smets, F. and Wouters, R., 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review* 97, 586–606.

Stock, J. and Watson, M., 1999. Forecasting Inflation. *Journal of Monetary Economics* 44, 293–335.

Sungbae, A. and Schorfheide, F., 2007. Bayesian Analysis of DSGE Models. *Econometric Reviews* 26, 113–172.

Söderlind, P., 1999. Solution and Estimation of RE Macromodels with Optimal Policy. *European Economic Review* 43:4, 813–823.

Taylor, J.B. 1993. Discretion Versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy* 39, 195–214.

Villani, M. 2001. Fractional Bayesian Lag Length Inference in Multivariate Autoregressive Processes. *Journal of Time Series Analysis* 22, 67–86.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, Inc, New York.

3 ESSAY 2: ROBUSTNESS OF THE RISK-RETURN RELATIONSHIP IN THE U.S. STOCK MARKET[1]

ABSTRACT

Using GARCH-in-Mean models, we study the robustness of the risk-return relationship in monthly U.S. stock market returns (1928:1 – 2004:12) with respect to the specification of the conditional mean equation. The issue is important because in this commonly used framework, unnecessarily including an intercept is known to distort conclusions. The existence of the relationship is relatively robust, but its strength depends on the prior belief concerning the intercept. The latter applies in particular to the first half of the sample, where also the coefficient of the relative risk aversion is smaller and the equity premium greater than in the latter half.

Keywords: Bayesian analysis ; ICAPM model; GARCH-in-Mean model

---

1 This paper has been published in the Finance Research Letters 5 (2008), pp. 118–127. Jani Luoto is responsible for empirical analysis. The writing and interpretation of results have been done jointly with Markku Lanne.

3.1 Introduction

There is a voluminous literature studying the relationship between stock market returns and their risk. The starting point in most of this research is the following linear equation for the expected excess return, r_t, implied by Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM),

$$E_{t-1}(r_t) = \mu_0 + \mu_1 Var_{t-1}(r_t), \qquad (1)$$

where the slope coefficient μ_1 is expected to be positive and the intercept μ_0 should be zero. The empirical evidence is mixed, but typically μ_1 has been found insignificant (see Ghysels et al., 2005).

Most of the previous empirical work on the ICAPM is based on the GARCH-in-Mean (GARCH-M) model. Within this model, Lanne and Saikkonen (2006) recently showed that omitting the theoretically justified zero restriction on the intercept in the conditional mean equation may lead to a considerable power loss in the standard Wald test. This may explain why the slope coefficient has often been found insignificant. Constraining the intercept to zero, Lanne and Saikkonen (2006) found a positive risk-return relationship in the U.S. stock market.

While the restriction $\mu_0 = 0$ produces reasonable empirical results, it may be criticized because it forces the expected excess return of the stock market to equal the risk-free interest rate under the null hypothesis $\mu_1 = 0$. This is in conflict with the stock premium puzzle literature put forth by Mehra and Prescott (1985), according to which the average stock return, in the U.S. in particular, has been excessively high relative to the risk. Therefore, the results of Lanne and Saikkonen (2006) concerning the risk-return relationship should be checked for robustness to confirm that they are not driven by the zero restriction on the intercept. To this end, we make use of Bayesian methods, and estimate the model assuming a number of different prior beliefs to observe the sensitivity of the estimate of μ_1. In the U.S. data, the finding of a positive risk-return relationship seems to be relatively robust with respect to the prior.

The plan of the paper is as follows. In Section 2, the GARCH-M model and the employed Bayesian methods are discussed. In Section 3, the empirical results are reported. Finally, Section 4 concludes.

3.2 Econometric Methodology

Let us consider the following GARCH-M model for the excess stock return r_t,

$$r_t = \mu_0 + \mu_1 h_t + h_t^{1/2} \varepsilon_t, \qquad (2)$$

where μ_0 and μ_1 are parameters, ε_t is an independently and identically distributed random variable with mean zero and variance unity and h_t is the conditional variance. We assume that h_t follows the GARCH(1,1) process of Bollerslev (1986),

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 h_{t-1}\varepsilon_{t-1}^2, \tag{3}$$

where the parameters β_0, β_1 and β_2 satisfy the conditions $\beta_0 \geq 0$, $\beta_1 \geq 0$, $\beta_2 > 0$ and $\beta_1 + \beta_2 < 1$. The GARCH (1,1) specification has almost invariably been found an adequate description of the dynamics of the conditional variance of stock returns (see e.g. Hansen and Lunde, 2005). The model can be estimated by the method of maximum likelihood (ML). Various distributional assumptions concerning ε_t have been entertained in the previous literature. Because the returns typically exhibit high excess kurtosis, normality has been found inadequate, while the standardized Student's t distribution with a relatively small degrees-of-freedom parameter ν has turned out to be quite satisfactory, and that is the assumption made in our empirical analysis. Assuming that $\nu > 2$, the likelihood function for a sample of T observations $r = (r_1, r_2, \ldots, r_T)'$ can be written as (Bollerslev, 1986)

$$l(\eta; r) = \prod_{t=1}^{T} \frac{c(\nu)}{\sqrt{h_t}} \left(1 + \frac{u_t^2}{(\nu - 2)h_t}\right)^{-\frac{\nu+1}{2}}, \tag{4}$$

where $u_t = h_t^{1/2}\varepsilon_t$, $\eta = (\beta_0, \beta_1, \beta_2, \mu_0, \mu_1, \nu)'$ is the vector consisting of all the parameters, and

$$c(\nu) = \Gamma\left(\frac{\nu+1}{2}\right) \Big/ \left[\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu - 2)}\right]$$

with $\Gamma(\cdot)$ the gamma function.

As discussed by Engle et al. (1987), the classical maximum likelihood estimator of η is consistent and asymptotically normally distributed if the model is correctly specified. This facilitates standard asymptotic inference, but as pointed out by Lanne and Saikkonen (2006), the standard Wald test has very low power if the model is misspecified in that the intercept term μ_0 is, in fact, equal to zero. Therefore, Lanne and Saikkonen (2006) recommended imposing this constraint if it is true. However, even though this restriction is implied by the theoretical model in our application of interest, it may not be consistent with the data, and the robustness of the results with respect to it needs to be checked. To this end, we suggest using Bayesian methods that easily allow for incorporating prior beliefs about the parameters. While restricting the intercept term to be equal to zero represents a very strong prior belief, other kinds of beliefs cannot easily be considered in the classical framework.

Most of the empirical work concerning (G)ARCH models is based on classical inference. However, at least Geweke (1989), Kleibergen and van Dijk (1993), Bauwens and Lubrano (1998, 2002), Nakatsuma (2000), Kaufmann and Fruhwirth-Schnatter (2002), Vrontos et al. (2000, 2003), and Bauwens et al. (2006) have employed the Bayesian framework.

The starting point of the Bayesian analysis of the GARCH-M model is the likelihood function (4) multiplied by the prior density function of the parameters, $p(\eta)$, yielding the posterior density:

$$q(\eta|r) \propto p(\eta) l(\eta; r) \tag{5}$$

The prior density reflects the researcher's prior beliefs concerning plausible parameter values. Our goal is to find out how the prior beliefs on μ_0 affect the estimate of μ_1 in the GARCH-M model, and this can be accomplished by comparing the posterior distributions based on a number of different priors.

In this paper, a standard assumption on prior independence, $p(\eta) = p(\beta_0)\cdot p(\beta_1)\cdot p(\beta_2)\cdot p(\mu_0)\cdot p(\mu_1)\cdot p(\nu)$, is used. The marginal priors are assumed to be proper, since improper priors, i.e. priors that are not well-defined density functions, cause marginal likelihoods (which we use in model comparison) to be indeterminate. This is because the normalizing constants of these priors are not defined. As shown by Bauwens and Lubrano (1998), sufficient prior information is needed on the Student-t degree of freedom parameter ν to force the posterior, in order to be integrable, to tend to zero quickly enough at the tail. We decided to follow Geweke (1993) in using an exponential density for the prior of ν. Furthermore, we decided to use log normal prior densities on β_0, β_1 and β_2, which reflects the nonnegativity constraint on these parameters. We also assume that, a priori, μ_0 and μ_1 are symmetrically distributed, and hence the normal distribution is entitled for them. The prior means of the parameters are based on previous empirical studies, especially that of Lanne and Saikkonen (2006).[2] Specifically, we set $\beta_0 \sim$ logN(-3, 4), $\beta_1 \sim$ logN(-0.2, 4), $\beta_2 \sim$ logN(-2.3, 4), $\mu_0 \sim$ N(0, σ_μ), $\mu_1 \sim$ N(0,10), and $\nu \sim$ Exp(0.1), where N(m, s) denotes the normal distribution with mean m and standard deviation s, logN(m, s) the corresponding log normal distribution and Exp(τ) is an exponential distribution with rate τ. The joint prior density function $p(\eta)$ in Eq. (5) is, hence, a product of univariate normal, log normal and exponential distributions. Finally, with the given values of prior standard deviations, the marginal priors turned out to be practically noninformative[3].

---

2 Note that our ML results (not reported here in order to save space) are similar to those of Lanne and Saikkonen (2006), since we use the same data, with the exception that our sample covers the period from 1928:1 to 2004:12, while their sample ends at 2000:12.

3 When one does not have a clear idea about the value of a parameter or does not want his prior knowledge to affect estimation results, he can assign to it a noninformative prior distribution. Noninformative distributions have typically a large variance or are

As we are interested in the effect of the prior distribution, several models based on different priors are considered. These models are compared by means of Bayes factors and estimated average discrepancies. A Bayes factor is defined to be the ratio of the marginal likelihoods of the two competing models, while the estimated average discrepancy approximates the posterior expectation of the deviance between the data and the model. Since the Bayes factor approach is based on the potentially invalid assumption that one of the models is true, the estimated average discrepancies are used as well (see Gelman et al., 2004). The Bayes factor is based on the marginal likelihood

$$q(r|M_k) = \int L(\eta_k; r, M_k) p(\eta_k|M_k) d\eta_k$$

of each of the models M_k (k = 1, 2,…, K). It is estimated from the simulated independent posterior sample using the reciprocal importance estimator with a multivariate normal importance density[4] (see Gelfand and Dey, 1994). Given these marginal likelihoods, the Bayes factors can be computed using the formula

$$\hat{B}_{ij} = \frac{\hat{q}(r|M_i)}{\hat{q}(r|M_j)}, \tag{6}$$

where the hat indicates an estimator. In the previous equation a small $\hat{B}_{ij}$ favors the model M_j. As Miazhynskaia and Dorffner (2004) show, for GARCH-type models the accuracy of the reciprocal importance estimator is statistically equal to more complex estimators, such as the bridge sampling estimator, Chib's candidate's estimator and the reversible jump MCMC estimator. Finally, the estimated average discrepancy is defined as

$$\hat{D}_{avg}(r) = \frac{1}{N}\sum_{i=1}^{N} D(r, \eta^i), \tag{7}$$

---

improper (nonintegrable). Opposite statement can be made for informative priors.

4 The reciprocal importance estimator is based on the following result

$$\frac{1}{q(r|M_k)} = \int \frac{f(\eta_k)}{L(\eta_k; r, M_k) p(\eta_k|M_k)} q(\eta_k|r, M_k) d\eta_k ,$$

where $f(\eta_k)$ is importance density (in our case multivariate normal). The associated estimator of $q(r\,|\,M_k)$ is

$$\hat{q}(r|M_k) = \left[\frac{1}{N}\sum_{i=1}^{N} \frac{f(\eta_k^i)}{L(\eta_k^i; r, M_k) p(\eta_k^i|M_k)}\right]^{-1}$$

where $\{\eta_k^i\}_{m=1}^{N}$ is a sample from the posterior $q(\eta_k\,|\,r, M_k)$ and N is the sample size.

where $r = (r_1, r_2, \ldots, r_T)'$ are the data, $D(r,\eta) = -2\log l(r;\eta)$ is the 'deviance', and $\{\eta^i\}_{i=1}^N$ a sample from the posterior (N is the sample size). In the limit, as the sample size tends to infinity, the model with the lowest expected deviance will have the highest posterior probability.

3.3 Empirical Results

In this section, we apply the approach of Section 2 to U.S. stock market returns. The emphasis is on demonstrating how the prior distribution of the intercept term μ_0 in equation (2) affects conclusions concerning the risk-return relationship. In particular, as discussed in Section 2, we assume a zero mean normal prior distribution for μ_0 and consider several alternative plausible values for the prior variance of μ_0 to see how the tightness of the prior assumption $\mu_0 = 0$ affects the estimation results. Although the equity premium puzzle indicates a positive μ_0, we rely on the underlying economic theory (ICAPM), which suggests zero mean on μ_0, as a basis of our prior knowledge[5]. Values of the prior variance of μ_0 close to zero are, thus, more in accordance with the ICAPM, while higher values of the prior variance of μ_0 allow for the possibility of the equity premium. Specifically, to study the effect of the prior of μ_0 on the estimate of μ_1, we let the prior standard deviation σ_μ of μ_0 vary from $6^{-1}\times 0.003$ to 10×0.003, where 0.003 is the standard error of the maximum likelihood estimate of μ_0 reported by Lanne and Saikkonen (2006). In practice, $\sigma_\mu = 10\times 0.003$ yields a noninformative prior distribution for μ_0, while $\sigma_\mu = 6^{-1}\times 0.003$ is small enough to force the posterior mean of μ_0 to be close to zero. Moreover, as explained in Section 2, the model fits of these competing models are compared by means of Bayes factors and estimated average discrepancies.

The data set consists of the monthly excess return on the value-weighted CRSP index (in excess of the 3-month Treasury bill rate) from 1928:1 to 2004:12.[6] The same data from the period 1928:1 – 2000:12 were used by Lanne and Saikkonen (2006). In addition to the entire sample, the models are estimated for

---

5 Since the equity premium puzzle indicates that a positive mean would be appropriate, we also considered several alternative values of μ_0 with prior standard deviation $\sigma_\mu = 6^{-1}\times 0.003$ around the premium estimates (≈6%) of Mehra and Prescott (1985). Specifically, we run nine regressions with prior mean of μ_0 varying from 0.02 to 0.10. The fit of these models turned out to be much worse than that of the models with 'prior variance adjustment'. For instance, the posterior probabilities for all these models were *virtually zero* (with noninformative model priors the posterior probability for model k can be calculated as $q(M_k \mid r) = q(r \mid M_k)/[q(r \mid M_1) + \ldots + q(r \mid M_K)]$, where $q(r \mid M_k)$ is the marginal likelihood of model k). Results were similar with different sample sizes. This indicates that models with 'mean adjustment' are clearly misspecified, since marginal likelihoods select the true model asymptotically. Natural exceptions were the models where the prior mean of μ_0 was set close to its posterior median (≈0.01). However, the model with high prior variance of μ_0 adequately takes these cases into account.

6 The data are downloaded from Kenneth French's homepage.

the 1928:1 – 1966:6 and 1966:7 – 2004:12 subsample periods of equal length. This serves as a check of robustness and parameter constancy. Furthermore, the results in the empirical finance literature suggest that the coefficient of relative risk aversion has increased, while the equity premium puzzle has diminished since the 1960's (see, e.g., Fama and French, 2002). That is, there may be some structural differences in these two subsample periods. In a similar subsample analysis, Lanne and Saikkonen (2006) also found the estimate of the slope coefficient μ_1 to be relatively stable once the intercept term μ_0 is restricted to zero, while the estimates from the different sample periods deviated considerably when μ_0 was unrestricted. Therefore, it is interesting to see whether our modeling approach leads to similar conclusions.

The estimation results[7] are presented in Table 1, in the topmost panel for the entire sample and in the lower panels for the two subsample periods, and in Figures 1-6. Each column corresponds to a different prior distribution of the intercept term μ_0, with the standard deviation ranging from 10×0.003 to 6^{-1}×0.003. As far as the GARCH parameters β_0, β_1 and β_2 and the degrees-of-freedom parameter ν are concerned, their posterior medians are insensitive to the prior. The parameters β_1 and β_2 are also accurately estimated, and the persistence in the volatility of the excess returns is high in all samples although the persistence tends to be somewhat lower in the latter subsample period. The estimated degrees of freedom indicate excess kurtosis, as expected. These findings are in accordance with the previous empirical results.

The models are compared by means of the Bayes factors and estimated discrepancy criteria. The Bayes factors reported in Table 1 are based on the benchmark of the model with the largest prior standard deviation of the intercept term (10×0.003). Note that small values of B_{1j} and $\hat{D}_{avg}(Y)$ in Table 1 favor a model. In the entire sample, the estimated average discrepancy criterion and Bayes factors seem to favor models with relatively large prior variance of the intercept term μ_0. In other words, the models allowing the intercept to deviate from zero, fit the data better. The Bayes factors lend the strongest support to the model with the prior standard deviation of the intercept term equal to 2×0.003. This model has Bayes factor 0.41 (relative to the benchmark model) and estimated average discrepancy -3044, while these figures are 6.89 and -3036, respectively, when the prior standard deviation of the intercept term is set at 6^{-1}×0.003. The results of the model with the full restriction of zero intercept (not reported) were very close to the latter. Thus, the cost of the full zero restriction is relatively high in this case. Furthermore, the preferred model gives moderate evidence in favor of a positive risk-return relationship. The posterior probability of μ_1 being positive exceeds 90% (see also Figure 2). Note that, the posterior medians of μ_1 vary considerably between the models. In the preferred model, the median equals 1.206 which is clearly smaller than 3.54, the estimate obtained by Lanne and Saikkonen (2006). However, such a small value

---

7 We used the Metropolis algorithm to generate a Monte Carlo sample from Eq. (5). Detailed description of the Metropolis algorithm and simulation routines are given in Appendix.

is better in accordance with the result (μ_1 = 2.6) that Ghysels et al. (2005) obtained using the far more sophisticated and data-demanding MIDAS method, as well as estimates of the coefficient of the relative risk aversion in the previous literature. Interestingly, the models with prior standard deviation of μ_0 above $2^{-1}\times0.003$ imply marginal posterior distributions of μ_0 with high probability of positive values, which is consistent with the equity premium puzzle. Note, however, that the posterior medians are very small compared to the equity premium estimate of approximately 6% obtained by Mehra and Prescott (1985).

In the first subsample period, the preferred model is the one with the prior variance of μ_0 equal to 4×0.003 according to the estimated discrepancy criteria and Bayes factors. The posterior median of μ_1 is as low as 0.506. In this period, the results lend only very weak support to the ICAPM, since only about 70% of the posterior mass of μ_1 lies above zero (see Figure 4).

The results for the latter subsample period suggest that the intercept term is very close to zero in accordance with the ICAPM. The most preferred model is the one with prior standard deviation of μ_0 equal to $6^{-1}\times0.003$, and, in this model, the coefficient of relative risk aversion is positive with high probability (see Figure 6). Furthermore, the results of the model with the full restriction of zero intercept (not reported) were practically equal to the most preferred model. All the models imply a considerably larger coefficient of relative risk aversion than in the first subsample period. There is also surprisingly little variation in the posterior medians of μ_1 between the models (see Figure 6). The differences between the subsample periods are consistent with the findings in the previous finance literature. In particular, the estimates of Ghysels et al. (2005) exhibit a similar change in the value of the coefficient of relative risk aversion. The fact that the intercept term was positive in the first but not in the second subsample period, probably reflects the moderation of the equity premium puzzle reported in the recent literature.

In sum, our results lend support to the robustness of the positive risk-return relationship in the U.S. stock market data, albeit the evidence is weaker in the first subsample period. In other words, this conclusion is shown not to be affected by (falsely) restricting the intercept term in the equation for the expected excess return equal to zero, when its true value is close to zero. The estimates of μ_1, however, were shown to be affected by imposing this restriction if untrue, as was the case in the full sample and the first subsample periods. As shown by Lanne and Saikkonen (2006), for power considerations, the restriction is needed in classical inference, and our analysis can be seen as a robustness check of the validity of their empirical results.

TABLE 1 Estimation results.

σ_μ	10×0.003	6×0.003	4×0.003	2×0.003	0.003	2^{-1}×0.003	4^{-1}×0.003	6^{-1}×0.003
				1928:1-2004:12				
β_0	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
	(4×1e-5)	(4×1e-5)	(4×1e-5)	(4×1e-5)	(5×1e-5)	(5×1e-5)	(5×1e-5)	(5×1e-5)
β_1	0.830	0.830	0.830	0.830	0.828	0.826	0.823	0.823
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.035)	(0.035)	(0.035)
β_2	0.127	0.127	0.127	0.127	0.125	0.121	0.118	0.117
	(0.028)	(0.028)	(0.029)	(0.028)	(0.027)	(0.027)	(0.026)	(0.026)
μ_0	0.007	0.007	0.007	0.006	0.005	0.002	6□1e-4	3□1e-4
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(7□1e-4)	(5□1e-4)
μ_1	0.886	0.895	0.951	1.206	1.838	2.713	3.205	3.328
	(0.977)	(0.960)	(0.953)	(0.940)	(0.865)	(0.752)	(0.664)	(0.638)
ν	8.124	8.141	8.177	8.261	8.475	8.740	8.805	8.862
	(2.188)	(2.149)	(2.178)	(2.310)	(2.364)	(2.440)	(2.510)	(2.532)
$P(\mu_0>0)$	99.9%	99.9%	99.9%	99.7%	99.1%	93.3%	79.3%	71.0%
$P(\mu_1>0)$	82.4%	82.0%	84.0%	90.3%	98.6%	100%	100%	100%
B_{1j}	-	0.64	0.49	0.41	0.82	2.82	5.64	6.89
$\hat{D}_{avg}(Y)$	-3044	-3044	-3044	-3044	-3043	-3039	-3037	-3036
				1928:1-1966:6				
β_0	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
	(5×1e-5)	(5×1e-5)	(5×1e-5)	(5×1e-5)	(5×1e-5)	(6×1e-5)	(6×1e-5)	(6×1e-5)
β_1	0.817	0.817	0.817	0.819	0.817	0.816	0.815	0.814
	(0.041)	(0.041)	(0.040)	(0.041)	(0.042)	(0.043)	(0.044)	(0.045)
β_2	0.157	0.159	0.158	0.157	0.155	0.148	0.145	0.144
	(0.045)	(0.044)	(0.044)	(0.044)	(0.043)	(0.041)	(0.040)	(0.040)
μ_0	0.011	0.010	0.010	0.009	0.006	0.002	0.001	3e-4
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(5□1e-4)
μ_1	0.392	0.429	0.506	0.820	1.562	2.400	2.823	2.906
	(0.949)	(0.930)	(0.935)	(0.916)	(0.872)	(0.814)	(0.773)	(0.772)
ν	8.7838	8.690	8.907	9.227	10.240	11.043	11.355	11.409
	(4.275)	(4.293)	(4.511)	(4.769)	(5.568)	(6.080)	(6.108)	(6.657)
$P(\mu_0>0)$	100.0%	100.0%	100.0%	100.0%	99.7%	95.6%	81.8%	73.0%
$P(\mu_1>0)$	66.5%	67.8%	70.4%	82.6%	96.8%	100.0%	100.0%	100.0%
B_{1j}	-	0.66	0.56	0.71	4.56	31	78	78
$\hat{D}_{avg}(Y)$	-1481	-1481	-1481	-1481	-1478	-1473	-1469	-1468
				1966:7-2004:12				
β_0	5e-4	5e-4	5e-4	4e-4	4e-4	4e-4	4e-4	4e-4
	(4×1e-4)	(4×1e-4)	(4×1e-4)	(4×1e-4)	(4×1e-4)	(4×1e-4)	(4×1e-4)	(4×1e-4)
β_1	0.677	0.684	0.687	0.693	0.696	0.701	0.702	0.700
	(0.199)	(0.195)	(0.194)	(0.190)	(0.187)	(0.183)	(0.185)	(0.186)
β_2	0.094	0.096	0.098	0.102	0.105	0.105	0.106	0.106
	(0.043)	(0.044)	(0.044)	(0.044)	(0.045)	(0.044)	(0.044)	(0.045)
μ_0	-0.003	-0.002	-0.002	-9e-4	-2e-4	-5e-5	-2e-5	-4e-6
	(0.009)	(0.008)	(0.007)	(0.005)	(0.003)	(0.001)	(0.001)	(0.001)
μ_1	4.734	4.507	4.335	3.883	3.615	3.522	3.495	3.515
	(4.440)	(3.876)	(3.412)	(2.416)	(1.599)	(1.196)	(1.031)	(1.012)
ν	8.321	8.264	8.302	8.314	8.267	8.326	8.315	8.346
	(3.599)	(3.435)	(3.373)	(3.412)	(3.320)	(3.315)	(3.399)	(3.337)
$P(\mu_0>0)$	36.6%	38.1%	38.6%	43.2%	46.5%	48.2%	48.6%	49.4%
$P(\mu_1>0)$	90.0%	90.7%	91.7%	95.4%	99.9%	99.9%	100.0%	100.0%
B_{1j}	-	0.54	0.35	0.25	0.22	0.21	0.21	0.19
$\hat{D}_{avg}(Y)$	-1559	-1559	-1559	-1559	-1560	-1560	-1560	-1560

Each column corresponds to a different prior distribution of the intercept term μ_0, with the standard deviation ranging from 10×0.003 to 6^{-1}×0.003. The posterior median and standard deviation (in parentheses) of each parameter are reported. The Bayes factors (B_{1j}) are based on the benchmark of the model with the largest prior variance of the intercept term (10×0.003).

3.4 Conclusion

We have studied the robustness of the risk-return relationship in the U.S. stock market. Three major conclusions stand out. First, the existence of a risk-return relationship is fairly robust in that it does not *strongly* depend on the prior beliefs concerning the intercept, especially when the true value of μ_0 is sufficiently close to zero. Second, although the existence of the relationship is relatively robust, the size of the estimated parameter measuring its magnitude depends on the specified prior of the intercept. Third, in accordance with a diminishing equity premium, there is evidence in favor of a positive intercept term only in the first half of the sample period.

FIGURES

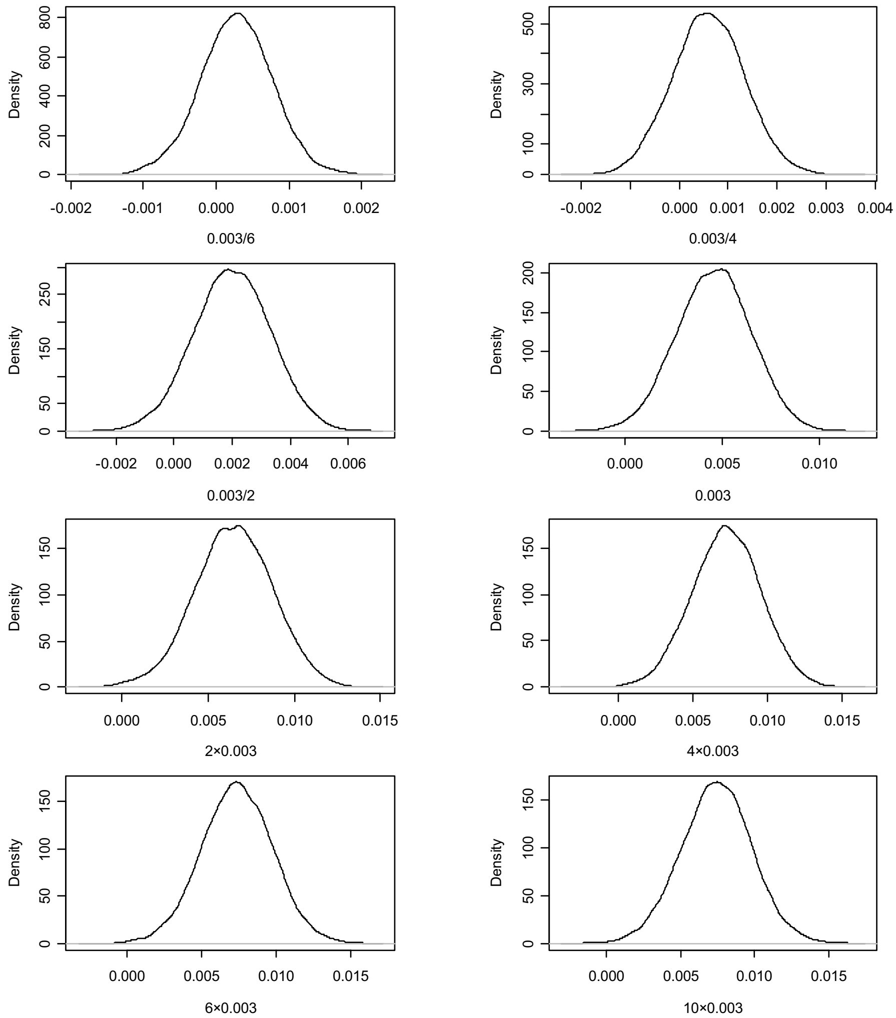

FIGURE 1 Estimated posterior densities of the parameter μ_0 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1928:1–2004:12.

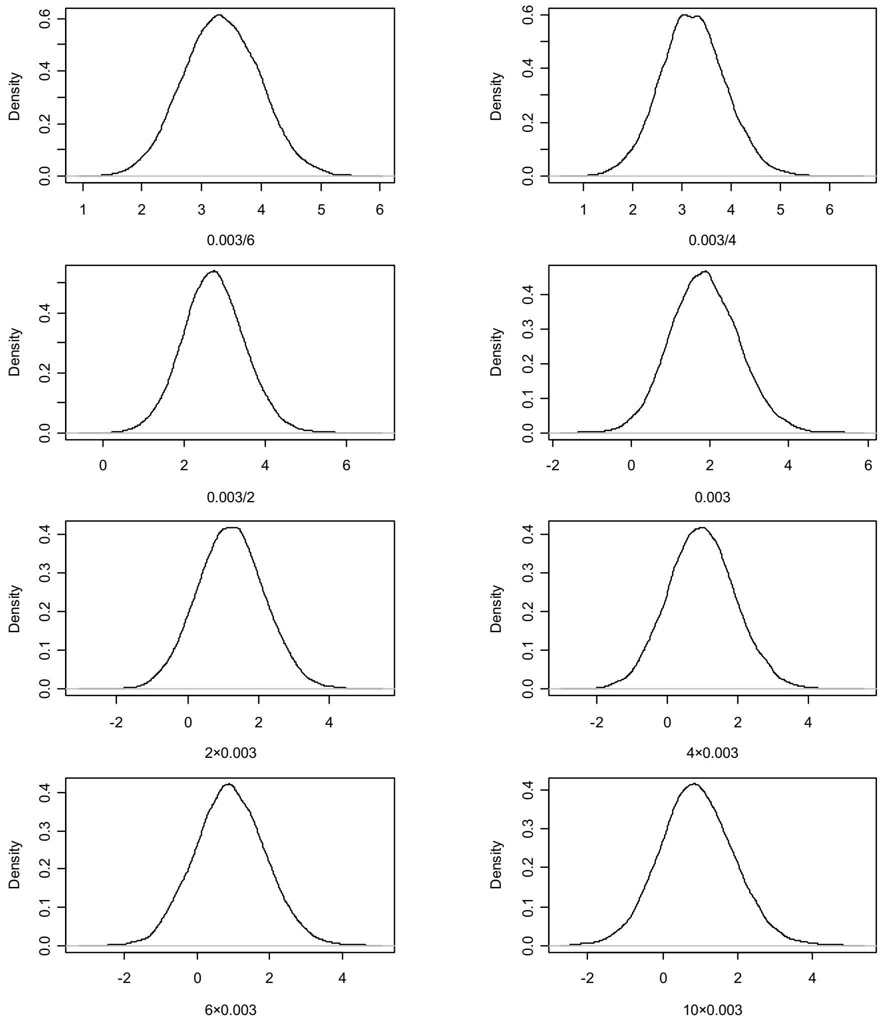

FIGURE 2 Estimated posterior densities of the parameter μ_1 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1928:1–2004:12.

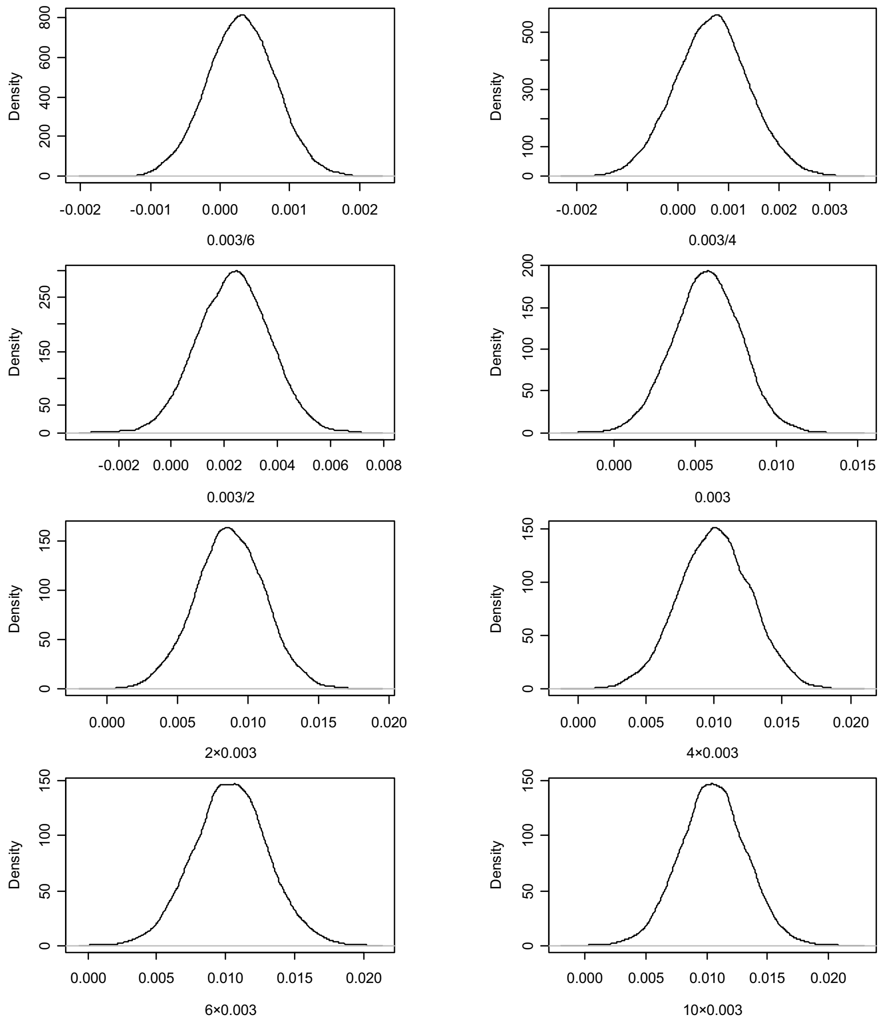

FIGURE 3 Estimated posterior densities of the parameter μ_0 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1928:1–1966:6.

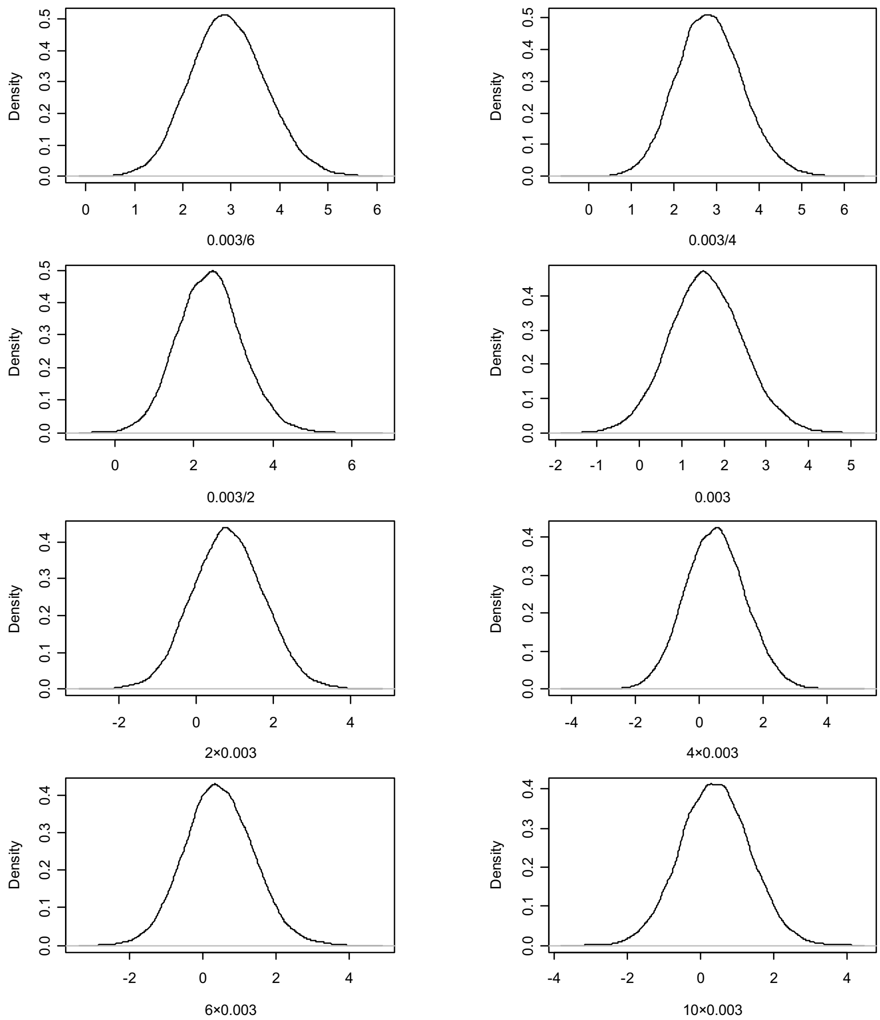

FIGURE 4 Estimated posterior densities of the parameter μ_1 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1928:1–1966:6.

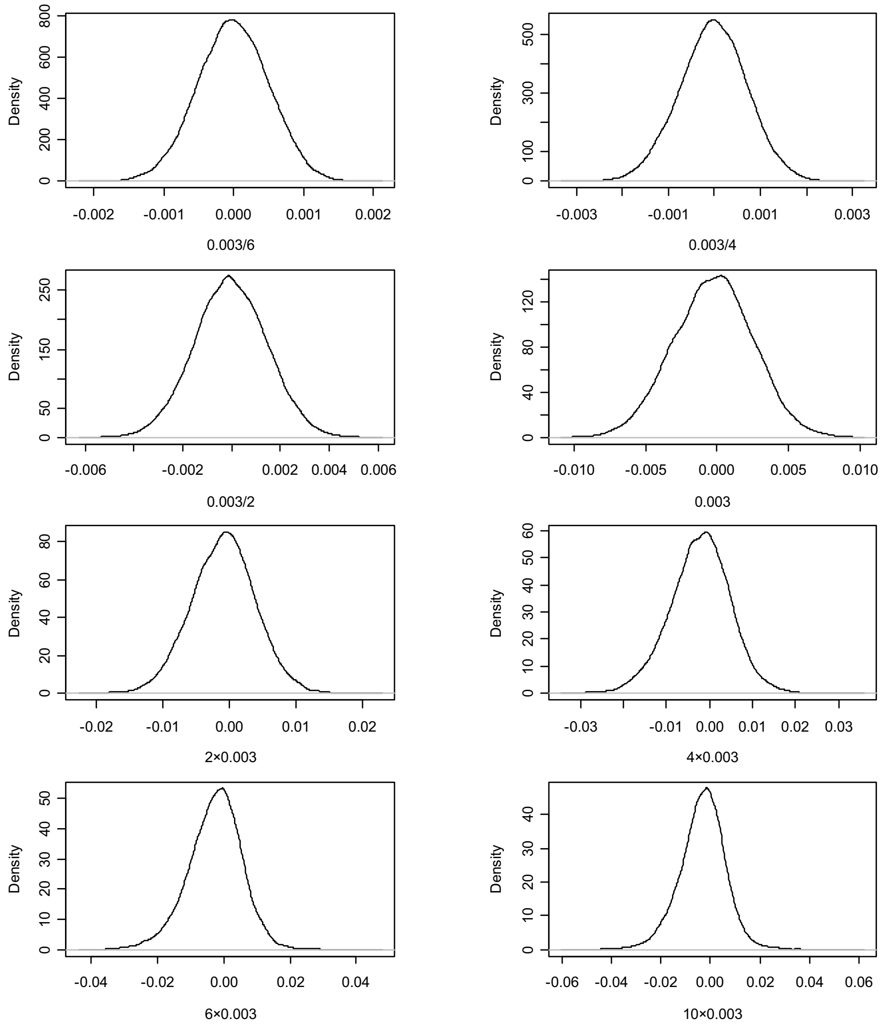

FIGURE 5 Estimated posterior densities of the parameter μ_0 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1966:7–2004:12.

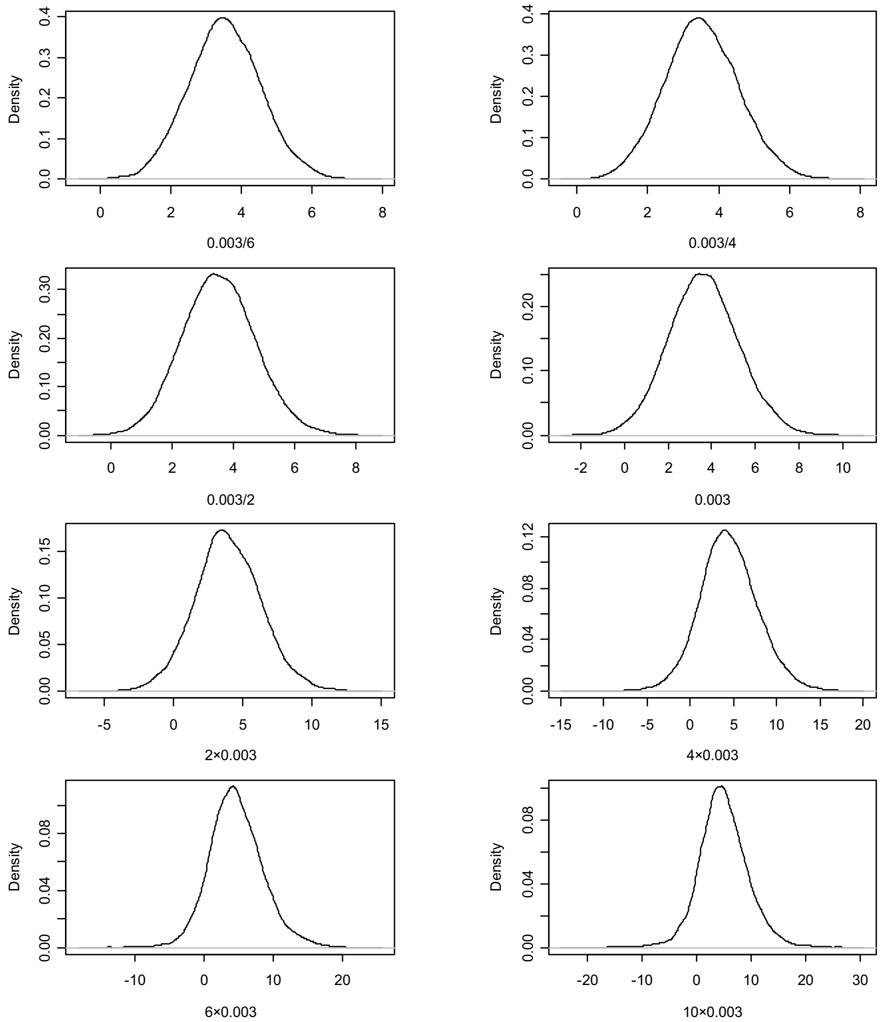

FIGURE 6. Estimated posterior densities of the parameter μ_1 with different prior standard deviations σ_μ = (0.003/6, 0.003/4, 0.003/2, 0.003, 2x0.003, 4x0.003, 6x0.003 and 10x0.003, respectively) of μ_0. Sample period 1966:7–2004:12.

REFERENCES

Bauwens, L. and Lubrano, M., 1998. Bayesian inference on GARCH models using Gibbs sampler. *Econometrics Journal* 1, 23–46.

Bauwens, L. and Lubrano, M., 2002. Bayesian option pricing using asymmetric GARCH models. *Journal of Empirical Finance* 9/3, 321-342.

Bauwens, L., Preminger, A. and Rombouts, J., 2006. Regime Switching GARCH Models. CORE Discussion Paper No. 2006/11.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Brooks, S.P. and Gelman, A., 1997. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434–455.

Engle, R.F., Lilien, D.M. and Robins, R.P., 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391–407.

Fama, E.F. and French, K.R., 2002. The equity premium. *Journal of Finance* 57, 637–659.

Gelfand, A. and Dey, D., 1994. Bayesian model choice: Asymptotic and exact calculations. *Journal of Royal Statistical Society, Ser. B* 56, 501 – 514.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004. Bayesian Data Analysis. 2nd edition, Chapman & Hall/CRC.

Gelman, A and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.

Geweke, J., 1989. Exact predictive densities in linear models with ARCH disturbances. *Journal of Econometrics* 40, 63-86.

Geweke, J., 1993. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8, 19–40.

Ghysels, E., Santa-Clara, P. and Valkanov, R., 2005. There is a risk-return tradeoff after all. *Journal of Financial Economics* 76, 509–548.

Hansen, P.R. and Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1). *Journal of Applied Econometrics* 20, 873–889.

Kaufmann, S. and Fruhwirth-Schnatter, S., 2002. Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis* 23, 425–458.

Kleibergen, F. and van Dijk, H., 1993. Non-stationarity in GARCH models: a bayesian analysis. *Journal of Applied Econometrics* 8, 41–61.

Lanne M. and Saikkonen , P., 2006. Why is it so difficult to uncover the risk-return tradeoff in stock returns? *Economic Letters* 92, 118–125.

Mehra, R. and Prescott, E.C., 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145 – 161.

Merton, R.C., 1973. An intertemporal capital asset pricing model. *Econometrica* 41, 867–887.

Miazhynskaia, T.S. and Dorffner, G., 2004. A comparison of Bayesian model selection based on MCMC with an application to GARCH-type models. *Statistical Papers* 47, 525-549.

Nakatsuma, T., 2000. Bayesian analysis of ARMA-GARCH models: a Markov chain sampling Approach. *Journal of Econometrics* 95, 57–69.

Vrontos, D., Dellaportas, P. and Politis, D.N., 2000. Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economics Statistics* 18, 187-198.

Vrontos, D., Dellaportas, P. and Politis, D.N., 2003. A full-factor multivariate GARCH model. *The Econometrics Journal* 6, 312-334.

APPENDIX

Since the posterior density function (p.d.f.) in Equation (5) and none of its full conditionals are in the form of any standard p.d.f., we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MMCMC) to generate a Monte Carlo sample from it. The algorithm uses a multivariate normal distribution for the jump distribution on changes in the transformed parameters $\theta = (\ln(\beta_0),\ \ln(\beta_1),\ \ln(\beta_2),\ \mu_0,\ \mu_1,\ \nu)'$. We apply logarithmic transformations to the parameters β_0, β_1 and β_2 to obtain approximate normality of their marginal posteriors, which makes the posterior simulations more efficient. Our simulation procedure was as follows: We first minimized the negative of the logarithm of the posterior density (5) numerically to obtain the posterior mode and evaluated the Hessian matrix at the minimum. We then used the inverse of the Hessian as an approximation to the posterior covariance matrix of θ and scaled it by the factor $2.4^2/d$, where d is the number of simulated parameters, to obtain an optimal covariance matrix for the jump distribution (see e.g. Gelman et al., 2004). We then added noise to the posterior mode to obtain overdispersed starting values and simulated three chains of length 400,000. We excluded the first 40,000 of simulations as a burn-in period in each chain and picked out every 90th draw. Thus, our final results are based on 12,000 draws. The convergence of the chains was checked using Gelman and Rubin's convergence diagnostic R (also called 'potential scale reduction factor') (see Gelman and Rubin, 1992). The diagnostic values close to 1 indicate approximate convergence and the values smaller than 1.1 are acceptable in most cases. The potential scale reduction factor of Gelman and Rubin (1992) was between 1 and 1.01 for each parameter. The multivariate version of Gelman and Rubin's diagnostic, proposed by Brooks and Gelman (1997), was 1 in all cases. Finally, the frequency of accepted jumps was between 0.18 and 0.26.[8] Finally, Figures 7-9 plots the sequences of the MCMC draws used to estimate some of parameters' posteriors.

---

[8] In the subsample 1966/72004/12 we use an adaptive Metropolis algorithm because the covariance matrix estimate based on local behaviour of the posterior at its highest peak turned out to give too optimistic a view of precision and thus failed to yield an efficient covariance matrix for the normal jump distribution. Specifically, we use a uniform prior between 0 and 1 for β_0, β_1 and β_2 and the following adaptive random walk Metropolis algorithm: We first simulate 10,000 draws using a diagonal covariance matrix with diagonal entries 0.00001 in the jump distribution. We then use these draws to estimate the posterior covariance matrix of η and scale it by the factor 2.42/*6*. We continue by simulating 10,000 draws and calculate a more accurate covariance matrix for η. We repeat this 2 – 5 times. Finally, we run 400,000 draws for three simulated chains using separate starting values and pick out every 90th draw after excluding the first 40,000 as a burnin period. The reader should note that this choice of the priors of the β's has no practical effect on the results; it only appears to be computationally more convenient to work with uniform rather than lognormal priors here.

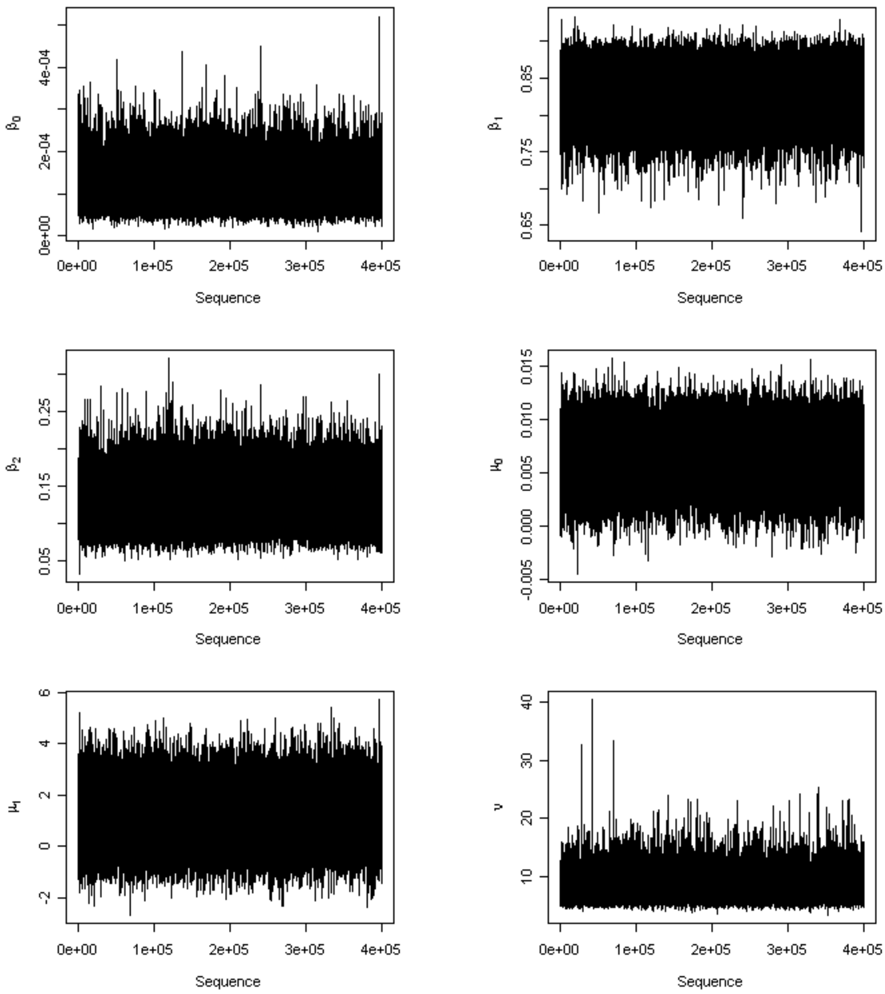

FIGURE 7 The sequences of MCMC draws of parameters' posteriors when σ_μ = 2x0.003. Sample period 1928:1–2004:12.

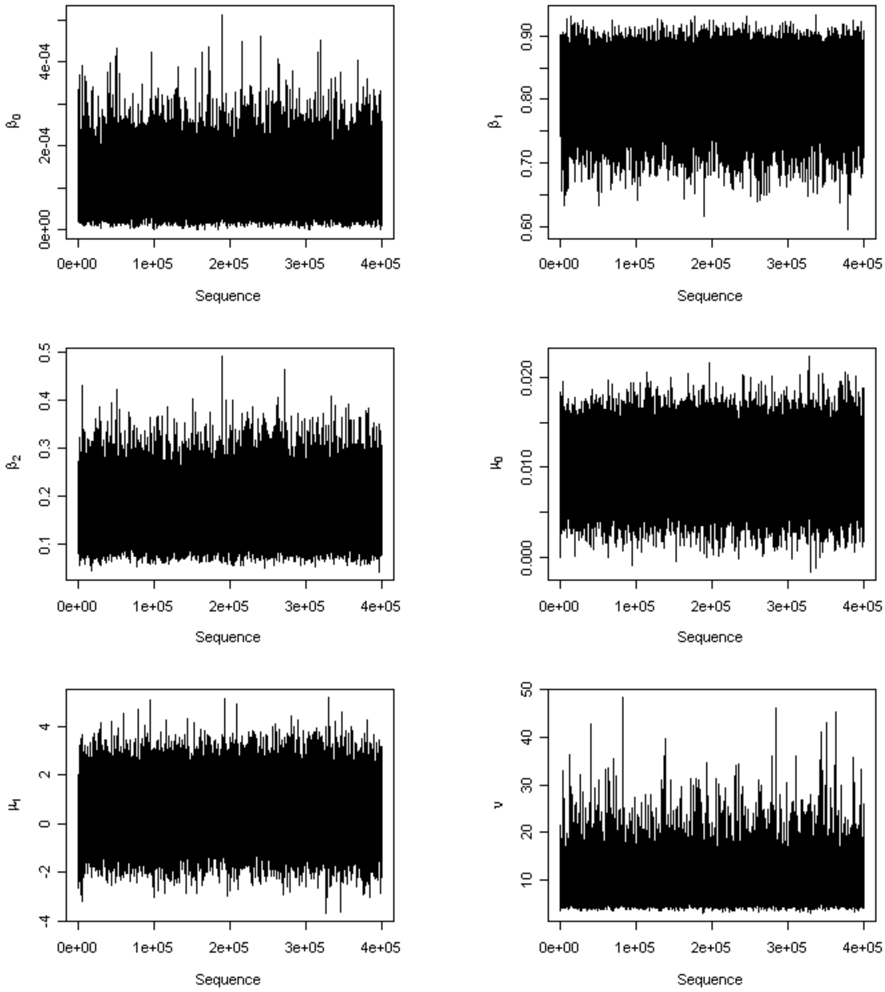

FIGURE 8 The sequences of MCMC draws of parameters' posteriors when σ_μ = 4x0.003. Sample period 1928:1–1966:6.

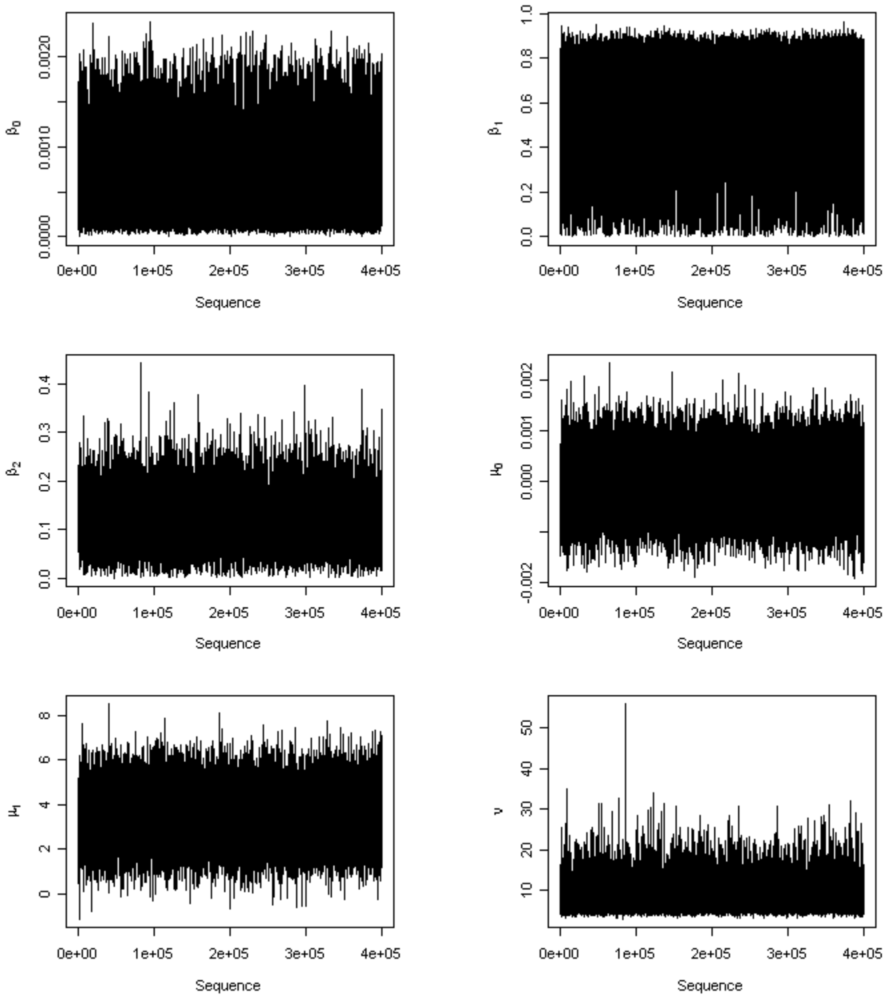

FIGURE 9 The sequences of MCMC draws of parameters' posteriors when $\sigma_\mu = 6^{-1}$x0.003. Sample period 1966:6–2004:12.

4 ESSAY 3: BAYESIAN TWO-STAGE REGRESSION WITH PARAMETRIC HETEROSCEDASTICITY[1]

ABSTRACT

In this paper we expand Kleibergen and Zivot's [2003. Bayesian and Classical Approaches to Instrumental Variable Regression. Journal of Econometrics 114, 29-72] Bayesian Two Stage (B2S) model by allowing for unequal variances. Our choice for modelling heteroscedasticity is a fully Bayesian parametric approach. As an application we present a cross-country Cobb-Douglas production function estimation.

Keywords: heteroskedasticity, Bayesian two-stage model, Cobb-Douglas production function

---

1 This paper has been accepted for publication in Advances in Econometrics 23 (Bayesian Econometrics, 2008). Jani Luoto is responsible for statistical estimation, interpretation of the results, and most of the writing. Arto Luoma assisted in some of the writing and derived most of the statistical equations.

4.1 Introduction

After Anderson and Rubin (1949) developed their limited information maximum likelihood (LIML) and Theil (1953) his two-stage least squares (2SLS) technique, instrumental variables (IV) regression has became a standard textbook approach in classical econometrics. The development of Bayesian analysis of such models started two decades later, being initiated by Drèze (1976) (see also e.g. Drèze and Morales, 1976, Drèze and Richard, 1983, and Bauwens and van Dijk, 1989). Drèze's idea was to equalize the classical and Bayesian analysis of IV models using suitable diffuse priors for the parameters. Unfortunately, his prior ignores important information concerning the near nonidentification of structural parameters due to weak instruments (see e.g. Kleibergen and Zivot, 2003, for discussion).

Mainly due to this undesirable property of the Drèze prior, recent research on Bayesian analysis of IV models has started to address the above mentioned problem of local non-identification (see e.g. Geweke, 1996, Kleibergen and van Dijk, 1998, and Chao and Phillips, 1998 and 2002). Following this tradition, Kleibergen and Zivot (2003) developed a new Bayesian two-stage (B2S) approach. In order to mimic classical 2SLS techniques, which essentially handle the problem of local non-identification, they constructed a prior for the parameters of the restricted reduced form specification and thus functionalized the steps used to obtain the 2SLS estimator.

In this paper we expand Kleibergen and Zivot's (2003) B2S model by allowing for unequal variances. In classical analysis, modelling heteroscedasticity improves the efficiency of estimation and enables the variance estimates to be consistent. Thus, not surprisingly, modelling heteroscedasticity has become standard in classical IV literature (see e.g. White, 1982, Cumby et al., 1983, and Davidson and MacKinnon, 1993). However, there is (to our knowledge) no single Bayesian study of IV models with unequal variances, although from the Bayesian point of view modelling heteroscedasticity should improve the precision of estimates and the quality of predictive inference. The latter follows from the fact that modelling heteroscedasticity allows predictive inferences to be more precise for some units and less so for other.

Our choice for modelling heteroscedasticity is a fully Bayesian parametric approach. Specifically, we assume that $\mathrm{var}(y_i) = \sigma^2 \zeta_i^{-\theta}$, where y is the response variable and ζ a variable explaining the variance. This specification requires only one unknown heteroscedasticity parameter (θ). Alternatively, we could follow, for example, Geweke (1993) and model heteroscedasticity using a nonparametric approach. This, however, would require estimation of several unknown parameters, which might give rise to identification and estimation problems in our relatively complex nonlinear model.

To give an empirical illustration of the properties of the heteroscedastic B2S model, we follow Benhabib and Spiegel (1994) and Papageorgiou (2003) and construct a simple exercise of aggregate production function estimation as an application.[2] We choose this example, since the problems of endogeneity and heteroscedasticity are well documented in the cross-country growth literature, see e.g. Benhabib and Spiegel (1994) and the surveys of Temple (1999a) and Durlauf et al. (2005).

The paper is organized as follows: In Section 2 we present a heteroscedastic Bayesian two-stage model (hereafter HB2S model). In Section 3 we give an example of estimating the empirical Cobb-Douglas aggregate production function. Section 4 concludes the paper.

4.2 The Bayesian Two-Stage Model with Parametric Heteroscedasticity

Consider the following limited information simultaneous equation model

$$y_1 = Y_2\beta + Z\gamma + \varepsilon_1, \tag{1}$$

$$Y_2 = X\Pi + Z\Gamma + V_2, \tag{2}$$

where $Y = (y_1\ Y_2)$ is an $N \times m$ matrix of endogenous variables, Z an $N \times k_1$ matrix of included exogenous variables, X an $N \times k_2$ matrix of excluded exogenous variables, that is, instruments, and ε_1 an $N \times 1$ vector of errors and V_2 an $N \times (m - 1)$ matrix of errors. Vectors β and γ contain the structural parameters of interest. The matrices Z and X are assumed to be of full column rank, uncorrelated with ε_1 and V_2, and weakly exogenous for the structural parameter vector β.

If the observation vectors y_i in the above simultaneous equation model have unequal covariance matrices, they are said to be heteroscedastic. In the following, we will model heteroscedasticity by assuming that the elements ε_{1i} of ε and the rows V_{2i} of V_2 are normally distributed with zero mean and the $m \times m$ covariance matrix

$$\Sigma_i = \mathrm{var}(\varepsilon_{1i} \quad V_{2i})' = f(\zeta_i, \theta)\begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{3}$$

where heteroscedasticity is captured by the function $f(\zeta_i,\theta)$, $\zeta_1,\ldots, \zeta_N$ being the known values of some positive-valued variable. Several alternative specifications of $f(\zeta_i,\theta)$ have been suggested in the literature (see e.g. Judge et al., 1985, Greene, 1990, Griffiths, 1999, and Tanizaki and Zhang, 2001). Here, we consider the following simple functional form

---

2 See also Barro (1999) and Temple (1999a, 1999b, 2001).

$$f(\zeta_i, \theta) = \zeta_i^{-\theta}, \qquad (4)$$

where $\theta \in [0,1]$, and the extreme of $\theta = 0$ corresponds to homoscedastic errors (see e.g. Greene, 1990, and Boscardin and Gelman, 1996). If we substitute the reduced form Equation (2) into the structural form Equation (1), we get the following nonlinearly restricted reduced form specification

$$y_1 = W\delta + v_1, \qquad (5)$$

$$Y_2 = UB + V_2, \qquad (6)$$

where $W = (UB\ Z)$, $\delta = (\beta'\ \gamma')'$, $U = (X\ Z)$, $B = (\Pi'\ \Gamma')'$ and $v_1 = \varepsilon_1 + V_2\beta$. Thus, U is an $N \times k$ matrix, where $k = k_1 + k_2$, and W is an $N \times (k_1 + m - 1)$ matrix. Denoting

$$\Omega_i = \mathrm{var}(v_{1i} \quad V_{2i})' = f(\zeta_i, \theta)\begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}; \ \omega_{11.2} = \omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}; \ \phi = \Omega_{22}^{-1}\Omega_{21},$$

we obtain that $e_1 = v_1 - V_2\phi$ is uncorrelated with V_2 and $\mathrm{var}(e_{1i}) = f(\zeta_i, \theta)\omega_{11.2}$.

From the reduced form equation (5) we can see the possible identification problem related to the two-stage approach. The parameter vector δ is identified when W has full column rank, which is equivalent to Π having full column rank, and is locally nonidentified when Π has a lower rank value. Therefore, the number of instruments has to be at least the number of endogenous regressors, that is, $k_2 \geq m - 1$. The model is called just-identified when $k_2 = m - 1$ and over-identified when $k_2 > m - 1$. In the case of weak instruments (that is, when the instruments are only weakly correlated with the endogenous regressors), identification problems may occur, since Π is then close to zero or close to having reduced rank (see e.g. Zivot et al., 1998, and Shea, 1997). Since in the Bayesian two-stage approach suggested by Kleibergen and Zivot (2003) the prior distribution is so constructed that it explicitly incorporates this kind of knowledge, we will choose it as a starting-point for our heteroscedasticity-corrected limited information model.

Using the restricted form specifications (5) and (6) we can write the likelihood in the form

$$p(Y|X, Z, \eta) = p(Y_2|X, Z, \eta)p(y_1|X, Z, \eta),$$

where

$$p(y_1|Y_2, X, Z, \eta) \propto \omega_{11.2}^{-0.5N}|\Lambda|^{-0.5}\exp\{-0.5\omega_{11.2}^{-1}(y_1 - W\delta - V_2\phi)'\Lambda^{-1}(y_1 - W\delta - V_2\phi)\}, \qquad (7)$$

$$p\left(Y_2 \middle| X, Z, \eta\right) \propto \left|\Omega_{22}\right|^{-0.5N} \left|\Lambda\right|^{-0.5(m-1)} \exp\left\{-0.5 tr \Omega_{22}^{-1} \left(Y_2 - UB\right)' \Lambda^{-1} \left(Y_2 - UB\right)\right\}. \qquad (8)$$

Here, η denotes the vector of all parameters and $\Lambda = diag\left(f\left(\zeta_1, \theta\right), \ldots, f\left(\zeta_N, \theta\right)\right)$.

As a prior distribution we will use a modification of the Jeffreys prior distribution. The Jeffreys prior is defined as $p(\eta) \propto \left|I(\eta)\right|^{1/2}$, where

$$I(\eta) = -E\left[\frac{\partial^2}{\partial \eta \partial \eta'} \log p(Y \mid \eta)\right]$$

is the Fisher information matrix for η. Our modifications will be as follows: firstly, we calculate the second-order derivative with respect to $vec(B)$ from the logarithm of the conditional density (8) instead of the full log likelihood. Secondly, we remove the prior dependence between B and δ by replacing the corresponding non-diagonal blocks in $I(\eta)$ with zero matrices. Then, if we assume that the geometric mean of ζ_i, i = 1,...,T, is unity, the joint prior is given by

$$p(\eta) \propto \left|\Omega_{22}\right|^{-0.5(m+k-1)} \omega_{11.2}^{-0.5(2m+k_1)} \left|U' \Lambda^{-1} U\right|^{0.5(m-1)} \left|W' \Lambda^{-1} W\right|^{0.5}. \qquad (9)$$

One can normalize the weight variable ζ by dividing it by its geometric mean. This has two advantages: firstly, one need not adjust the prior distribution, and secondly, the dispersion parameters $\omega_{11.2}$ and Ω_{22} have a consistent meaning under different values of θ. See also Boscardin and Gelman (1996), who discuss the issue in the context of one-stage regression models.

Our derivation of the prior distribution is somewhat arbitrary. However, it yields a prior with desirable properties. The presence of the term $\left|W' \Lambda^{-1} W\right|^{1/2}$ in the prior (9) reflects the fact that the model is not informative regarding β when Π has reduced rank, since this term tends to zero as Π tends to a reduced rank value. In the special case, when Z is not in the model and θ = 0, our prior reduces to that proposed by Kleibergen and Zivot (2003). The slightly simpler Drèze prior $\left|\Omega\right|^{-(1/2)(k+m+1)}$ has the drawback that the marginal posterior of Π has a non-integrable asymptote at Π = 0, when the model is just-identified (see Kleibergen and Zivot, 2003 for further discussion on the issue).

Multiplying the likelihood function by the joint prior (9) yields, after some tedious algebra, the following conditional and marginal posteriors

$$p\left(\delta \middle| Y,\phi,B,\theta,\omega_{11.2},\Omega_{22}\right) \propto \omega_{11.2}^{-0.5(m+k_1-1)}\left|W'\Lambda^{-1}W\right|^{0.5},$$
$$\exp\left\{-0.5\omega_{11.2}^{-1}\left(\delta-\hat{\delta}\right)'W'\Lambda^{-1}W\left(\delta-\hat{\delta}\right)\right\}, \quad (10)$$

$$p\left(\phi \middle| Y,B,\theta,\omega_{11.2},\Omega_{22}\right) \propto \omega_{11.2}^{-0.5(m-1)}\left|V_2'\Lambda^{-1}MV_2\right|^{0.5}$$
$$\exp\left\{-0.5\omega_{11.2}^{-1}\left(\phi-\hat{\phi}\right)'V_2'\Lambda^{-1}MV_2\left(\phi-\hat{\phi}\right)\right\}, \quad (11)$$

$$p\left(\omega_{11.2} \middle| Y,B,\theta,\Omega_{22}\right) \propto \omega_{11.2}^{-0.5(N+2)}\left(v'\Lambda^{-1}Mv\right)^{0.5N}\exp\left\{-0.5\omega_{11.2}^{-1}v'\Lambda^{-1}Mv\right\}, \quad (12)$$

$$p\left(\Omega_{22} \middle| Y,B,\theta\right) \propto \left|\Omega_{22}\right|^{-0.5(N+m+k-1)}\left|V_2'\Lambda^{-1}V_2\right|^{0.5(N+k-1)}\exp\left\{-0.5tr\Omega_{22}^{-1}V_2'\Lambda^{-1}V_2\right\}, \quad (13)$$

$$p\left(B,\theta \middle| Y\right) \propto \left(v'\Lambda^{-1}Mv\right)^{-0.5N}\left|U'\Lambda^{-1}U\right|^{0.5(m-1)}\left|\Lambda\right|^{-0.5m}\left|V_2'\Lambda^{-1}MV_2\right|^{-0.5}\left|V_2'\Lambda^{-1}V_2\right|^{-0.5(N+k-1)}, \quad (14)$$

where

$$\hat{\delta} = \left(W'\Lambda^{-1}W\right)^{-1}W'\Lambda^{-1}\left(y_1 - V_2\phi\right), \hat{\phi} = \left(V_2'\Lambda^{-1}MV_2\right)^{-1}V_2'\Lambda^{-1}My_1$$
$$M = I - W\left(W'\Lambda^{-1}W\right)^{-1}W'\Lambda^{-1}, \; v = y_1 - V_2\hat{\phi} \text{ and } V_2 = Y_2 - UB.$$

The distributions given in Equations (10)-(11) are multivariate normal, while those in (12) and (13) are Inverse Gamma and Inverse Wishart, respectively. The joint marginal posterior for B and θ in Equation (14) does not have a form of any standard p.d.f.

Kleibergen and Zivot (2003) discuss some properties of their B2S model and compare it to the original Drèze (1976) approach. We briefly review their discussion and make some comparison between our parametric heteroscedasticity-corrected model and their B2S model.

a) As with the Drèze and B2S approaches, the posteriors are not invariant to the ordering of the endogenous variables; that is, if y_1 is exchanged with some of the variables in Y_2, the results do not remain identical. See Drèze (1976) for the argument.
b) The mean of the conditional posterior of β in the B2S model is essentially $\hat{\beta}_{2SLS}$. However, this is not true for the HB2S model, since heteroscedasticity correction gives more weight to 'good' observations, while $\hat{\beta}_{2SLS}$ weights all observations equally. The difference between the heteroscedastic-corrected estimate of β and $\hat{\beta}_{2SLS}$ depends, of course, on the degree of heteroscedasticity.
c) As with the B2S approach, the marginal posterior (14) does not have the non-integrable asymptote at $\Pi = 0$ which appears in the Drèze approach. The last term in (14) may be written in the form

$\left|S+(B-\hat{B})'U'\Lambda^{-1}U(B-\hat{B})\right|^{-(N+k-1)/2}$, where $\hat{B}=(U'\Lambda^{-1}U)^{-1}U'\Lambda^{-1}Y_2$ and $S=(Y_2-U\hat{B})'\Lambda^{-1}(Y_2-U\hat{B})$, and is, for a fixed Λ, a kernel of a matric-variate Student-t density with N – 1 degrees of freedom. The other terms are, for a fixed Λ, bounded from zero and infinity, which implies that the posterior is integrable with respect to B. If the term $\left|W'\Lambda^{-1}W\right|^{1/2}$ were not present in the prior its inverse would appear in the posterior causing an infinite asymptote at the reduced rank values of Π.

d) As with the B2S approach (without heteroscedasticity correction), the form of the posterior of B is closely related to the marginal posterior which results from a standard diffuse prior analysis of the reduced form regression of Y_2 on U with heteroscedasticity correction.

4.3 Empirical Example

4.3.1 Estimated Model

To illustrate some of the properties of the HB2S model we construct a simple exercise of aggregate production function estimation with cross-country data. We chose this example, since problems of endogeneity and heteroscedasticity are well documented in the cross-country growth literature (see e.g. Benhabib and Spiegel, 1994, Papageorgiou, 2003, and the surveys of Temple, 1999a, and Durlauf et al., 2005). For example, Benhabib and Spiegel (1994) analyse the biases of coefficient estimates which result from the correlation between the accumulated physical and human capital series and the error term, and find that there is likely to be an upward coefficient bias in the input share of capital and human capital estimates, and a downward bias in estimates of the input share of labour. Our analysis is close to that of Benhabib and Spiegel (1994) or Papageorgiou (2003). However, we do not separate aggregate labour and human capital stocks; rather we follow Bils and Klenow (2000) and assume that individual human capital stock is related to individuals, years of schooling and years of experience. This implies that each individual has some degree of human capital and thus aggregate human-capital stock should be modelled as $H_t = h_tL_t$, where h_t is average human-capital stock per person and L_t is labour force.

We assume the Romer-type Cobb-Douglas production function $Y_{it}=A_{it}^{\alpha+\beta}K_{it}^{\alpha}H_{Y,it}^{\beta}\varepsilon_{it}$ (see Romer, 1990), where Y_{it} is the output, A_{it} productivity, K_{it} physical capital and $H_{Y,it}$ the human capital engaged in final-goods production in country i at time t. Taking log differences we obtain the following equation for long-run growth,

$$\log\left(\frac{Y_{iT}}{Y_{i0}}\right)=(\alpha+\beta)\log\left(\frac{A_{iT}}{A_{i0}}\right)+\alpha\log\left(\frac{K_{iT}}{K_{i0}}\right)+\beta\log\left(\frac{H_{Y,iT}}{H_{Y,i0}}\right)+\log\left(\frac{\varepsilon_{iT}}{\varepsilon_{i0}}\right). \qquad (15)$$

In Equation (15), we assume that the resource constraint $H_{it} = H_{A,it} + H_{Y,it}$, where $H_{A,it}$ is the human capital engaged in R&D activities, holds. One problem in estimating Equation (15) is that we should replace an unobservable $\log(A_T/A_0)$ by some function of observables. Otherwise the estimates of factor shares will be biased (see e.g. Temple, 1999a and b). We follow Papageorgiou (2003) and propose the following specification for the growth rate of technology,

$$\frac{A_{iT} - A_{i0}}{A_{i0}} = \delta H_{A,iT} + \mu H_{A,iT}\left(\frac{\overline{A}_0}{A_{i0}} - 1\right), \tag{16}$$

where $\overline{A}_0$ is the technology frontier, and δ and μ are the innovation and imitation parameters, respectively. In Equation (16), human-capital speeds technology growth through innovation and imitation. Using Equation (16) we can write Equation (15) in the estimation form

$$\log\left(\frac{Y_{iT}}{Y_{i0}}\right) = \psi \cdot d_i + (\alpha + \beta)\left[(\delta - \mu)H_{A,i0} + \mu\left(\frac{y_0^{\max}}{y_{i0}}\right)H_{A,i0}\right] + \alpha\log\left(\frac{K_{iT}}{K_{i0}}\right) + \beta\log\left(\frac{H_{Y,iT}}{H_{Y,i0}}\right) + u_{it}, \tag{17}$$

where d_i is a vector of deterministic components (constant and dummy variables), and u_{it} a normally distributed error term with zero mean and σ_{11} variance. We follow Benhabib and Spiegel (1994) and Papageorgiou (2003) in assuming that $(Y_0/L_0)^{\max}/(Y_{i0}/L_{i0}) = y_0^{\max}/y_{i0}$ approximates $\overline{A}_0/A_{i0}$.

Since human capital may also speed technology adoption and may be to some extent necessary for technology use, we propose the following two alternative specifications for technology growth with production technology, $Y_{it} = A_{it}^{\alpha+\beta}K_{it}^{\alpha}H_{it}^{\beta}\varepsilon_{it}$ and $Y_{it} = A_{it}^{\alpha+\beta}K_{it}^{\alpha}L_{it}^{\beta}\varepsilon_{it}$ (see e.g. Benhabib and Spiegel, 1994, and Bils and Klenow, 2000). With similar steps we obtain the corresponding empirical specifications

$$\log\left(\frac{Y_{iT}}{Y_{i0}}\right) = \psi \cdot d_i + (\alpha + \beta)\left[(\delta - \mu)H_{i0} + \mu\left(\frac{y_0^{\max}}{y_{i0}}\right)H_{i0}\right] + \alpha\log\left(\frac{K_{iT}}{K_{i0}}\right) + \beta\log\left(\frac{H_{iT}}{H_{i0}}\right) + u_{it}, \tag{18}$$

$$\log\left(\frac{Y_{iT}}{Y_{i0}}\right) = \psi \cdot d_i + (\alpha + \beta)\left[(\delta - \mu)H_{i0} + \mu\left(\frac{y_0^{\max}}{y_{i0}}\right)H_{i0}\right] + \alpha\log\left(\frac{K_{iT}}{K_{i0}}\right) + \beta\log\left(\frac{L_{iT}}{L_{i0}}\right) + u_{it}. \tag{19}$$

We estimate Equations (17)-(19) using our HB2S model. As a weight variable we use $\zeta_i = y_{i0}$, where

$$y_{i0} = \exp\left\{\log(Y_{i0} / L_{i0}) - \frac{1}{N}\sum_{i=1}^{N}\log(Y_{i0} / L_{i0})\right\}.$$

This corresponds to dividing the output per labour force (Y_{i0}/L_{i0}) by its geometric mean. We use the output per worker in the weight coefficients $f(\zeta_i,\theta)$, $i = 1,\ldots,N$ since we expect countries with higher initial income to have more stable growth paths due to developed institutional structures, which have the ability to reduce the overall risk in society. Alternatively, we could use some institutional indicator. However, since the choice of institutional indicators which approximate the 'true level' of institutional quality is somewhat difficult, and far from unique, we decided to abandon this approach.

4.3.2 Estimation Results

Since physical and human capitals are accumulated factors, they are endogenous. This causes the simple OLS estimator to be inconsistent. A common means of dealing with the issue of endogeneity is to instrument for endogenous regressors with variables correlated with them but exogenous to them and the regressed variable. Moreover, the validity of an instrument requires that it cannot be a direct growth determinant or correlated with omitted growth determinants (see e.g. Durlauf et al., 2005). Therefore, we instrument the growth rates of aggregate human and physical capital using the distance from the equator (Gallup, Sachs and Mellinger, 1998) and the following variables in 1970: age dependency ratio (dependents to working-age population), illiteracy rate (%) of people aged 15-24 from World Development Indicators (2002), and the level of physical capital per worker. The data and instruments are described in Appendix A.

To generate a Monte Carlo sample from the joint posterior of θ and B we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MMCMC). The algorithm uses a multivariate normal distribution for the jump distribution on changes in θ and B. Our simulation procedure was as follows: We first minimized the negative of the logarithm of the posterior density (14) numerically to obtain the posterior mode and evaluated the Hessian matrix at the minimum. We then used the inverse of the Hessian as an approximation to the posterior covariance matrix of $(\theta, vec(B)')'$ and scaled it by the factor $2.4^2/d$, where d is the number of simulated parameters, to obtain an optimal covariance matrix for the jump distribution (see e.g. Gelman et al., 2004).

We then added noise to the posterior mode to obtain overdispersed starting values and simulated three chains of length 100,000. We excluded the first half of simulations as a burn-in period in each chain and picked out every tenth draw. The convergence of the chains was checked using Gelman and Rubin's convergence diagnostic R (also called 'potential scale reduction factor') (see Gelman and Rubin, 1992). The diagnostic values close to 1 indicate approximate convergence and the values smaller than 1.1 are acceptable in most cases. In our case the diagnostic was estimated as 1.00 for all parameters and all models; the convergence was thus very good. Table 2 in Appendix B shows the simulation results for model (17). After simulating θ and B, the other parameter vectors and matrices were simulated from the conditional distributions (13) - (10). The B2S model could be estimated similarly, except that the covariance matrix of the classical first-stage regression (scaled by the factor $2.4^2/d$) could be used as the covariance matrix of the jump distribution.[3]

We use the posterior mean of deviance, $D_{avg}(Y) = E\{D(Y,\eta)|Y\}$, as a measure of model fit. This criterion is called 'average discrepancy' by Gelman et al. (2004) who recommend its use for model comparison[4]. It is estimated as $\hat{D}_{avg}(Y) = \frac{1}{N}\sum_{i=1}^{N} D(Y,\eta^i)$, where $D(Y,\eta) = -2\ln p(Y|\eta)$ is the deviance and $\eta^i, i = 1,...,N$, are posterior simulations. The average discrepancy is usually greater than $D_{\hat{\eta}}(Y) = D(Y,\hat{\eta})$, where $\hat{\eta}$ is a point estimate, such as posterior mode, mean or median. The difference $p_D = D_{avg}(Y) - D_{\hat{\eta}}(Y)$ is called the effective number of parameters and is in most cases approximately equal to the number of parameters in nonhierarchical models. The Bayesian equivalent (and also a generalization) of the Akaike information criterion (AIC) is the deviance information (DIC), defined as $DIC = D_{\hat{\eta}}(Y) + 2p_D$. The DIC has been suggested as a criterion of model fit when the goal is to select a model with best out-of-sample predictive power (see Spiegelhalter et al., 2002).

Figure 1 displays the residual plots for the first- and second-stage regressions, corresponding to Equations (8) and (7), respectively. The residuals have been obtained by replacing the unknown parameters by their posterior means and they have been plotted against the normalized initial output y_0. The approximate 95% probability belts, based on the normality assumption, are also shown. We see that the fit of the belts seems worse for the log differences of Y and K, since there are no points outside these bands for $y_0 > 2$. The reason for this is probably that we have only one heteroscedasticity parameter for all regressed variables and the fit cannot be equally good for all of them.

Table 1 shows the estimation results for Equations (17)-(19), obtained using the ordinary least squares (OLS) method and the Bayesian estimation of the B2S and HB2S models. On the basis of the figures and the posterior summaries of the heteroscedasticity parameter θ, we see that the data support heteroscedasticity in each model. Heteroscedasticity is especially obvious in the cases of output and physical capital growth, less so in human capital and only

3 The estimation was implemented using R, a statistical computing environment. R is freely available under the General Public Licence at www.R-project.org. The code and data sets are available at http://mtl.uta.fi/codes/HB2S/.

4 Gelman et al. (2004) prefer using discrepancy between data and model to using Bayes factors in model comparisons. They consider Bayes factors to be in most cases irrelevant, since they are used to compute the relative probabilities of the models conditional on one of them being true.

slight in labour growth. When we compare the estimated average discrepancies $\hat{D}_{avg}(Y)$ between the B2S and HB2S models, we see that the data lend strong support to the latter. If one would test the significance of θ using the likelihood ratio test, a significant result at the 5 % level would correspond to a difference greater than 3 in $\hat{D}_{avg}(Y)$. This follows from the fact that the nested model has one parameter less in this case and from the relation between $\hat{D}_{avg}(Y)$ and the number of parameters.

TABLE 1 Growth regressions for Equations 17-19

Par.	Equation (17)			Equation (18)			Equation (19)		
	OLS	B2S	HB2S	OLS	B2S	HB2S	OLS	B2S	HB2S
$(a+\beta)(\delta-\mu)$	-0.096	-0.090	-0.077	-0.016	-0.009	-0.006	-0.023	-0.021	-0.017
	(0.048)	(0.224)	(0.1496)	(0.0146)	(0.035)	(0.022)	(0.012)	(0.034)	(0.023)
$(a+\beta)\mu$	0.050[f]	0.067[a]	0.065	0.002[f]	0.002	0.001	0.002[f]	0.002	0.002[a]
	(0.013)	(0.048)	(0.065)	(0.0005)	(0.002)	(0.001)	(0.0004)	(0.002)	(0.001)
a	0.432[f]	0.283[b]	0.339[c]	0.433[f]	0.293[c]	0.347[c]	0.452[f]	0.313[c]	0.361[c]
	(0.073)	(0.107)	(0.107)	(0.074)	(0.097)	(0.103)	(0.070)	(0.101)	(0.101)
β	0.378[e]	0.537[c]	0.470[c]	0.365[e]	0.539[c]	0.483[c]	0.408[f]	0.618[c]	0.547[c]
	(0.137)	(0.152)	(0.147)	(0.144)	(0.150)	(0.150)	(0.162)	(0.185)	(0.169)
$a+\beta$	-	0.820	0.806	-	0.831	0.829	-	0.931	0.908
		(0.129)	(0.106)		(0.131)	(0.108)		(0.161)	(0.136)
δ	-	-0.036	-0.031	-	-0.012	-0.007	-	-0.022	-0.018
		(0.269)	(0.172)		(0.044)	(0.028)		(0.038)	(0.027)
μ	-	0.086[a]	0.077	-	0.002	0.002	-	0.002	0.002[a]
		(0.066)	(0.089)		(0.002)	(0.002)		(0.0018)	(0.0017)
θ	-	-	0.529	-	-	0.519	-	-	0.434
			(0.122)			(0.111)			(0.113)
Constant	0.275	0.335	0.305	0.255	0.270	0.235	0.351	0.417	0.376
	(0.064)	(0.133)	(0.099)	(0.081)	(0.150)	(0.109)	(0.062)	(0.133)	(0.105)
Dummy	-0.262	-0.348	-0.298	-0.235	-0.302	-0.254	-0.287	-0.387	-0.327
Africa	(0.112)	(0.102)	(0.105)	(0.106)	(0.098)	(0.097)	(0.121)	(0.107)	(0.107)
Dummy	-0.187	-0.261	-0.253	-0.174	-0.244	-0.237	-0.199	-0.285	-0.269
LatinAmerica	(0.089)	(0.104)	(0.090)	(0.088)	(0.097)	(0.088)	(0.096)	(0.105)	(0.097)
$\hat{D}_{avg}(Y)$	-	-378.7	-399.0	-	-373.0	-394.4	-	-417.8	-432.1
$\bar{R}^2$	0.62	-	-	0.62	-	-	0.61	-	-

a parameter > 0 with 90-94 % probability
b parameter > 0 with 95-98 % probability
c parameter > 0 with 99 -100 % probability
d p-value of one-sided hypothesis test < 0.10
e p-value of one-sided hypothesis test < 0.05
f p-value of one-sided hypothesis test < 0.01

For the OLS estimates, White's heteroscedasticity-corrected standard errors are given in parentheses. In the Bayesian models, the posterior means and standard deviations are reported. We do not report the OLS estimates of the imitation and innovation parameters, since exact standard errors are not available for them.

It seems that the margin between the estimated average discrepancies of B2S and HB2S depends on the degree of heteroscedasticity, since the gap is around 20 for models (17) and (18), while for model (19) it is about 14 only. Note that model (19) has the lowest heteroscedasticity parameter θ due to the small amount of heteroscedasticity in the labour growth series (see the third row in Figure 1). However, the differences in θ are not very significant between the models, since they are smaller than the posterior deviation of the parameter. Furthermore, the model defined by (19) has the smallest estimated average discrepancy (-432). However, in the 'economic theory' sense, this result does not necessarily indicate that Equation (19) is more preferable than the other models, since the first-stage regression could be more informative in this model, increasing the overall model fit.

We also find that the IV regression estimates of a are, in general, higher, and the estimates of β lower, than the corresponding OLS estimates[5]. Thus our results confirm the finding of Benhabib and Spiegel (1994) that there is an upward coefficient bias in the OLS estimates of α and human capital share β (Equations 17-18), and a downward bias in the OLS estimates of the labour share parameter β (Equation 19)[6] (see also Griliches and Mairesse, 1995, for a discussion of endogeneity of regressors in the aggregate production function approach).

Finally, based on the results reported in Table 1, the data are not consistent with the innovation parameter δ being positive (see e.g. Benhabib and Spiegel, 1994, who obtained similar results in their analyses). On the other hand, there is weak (or moderate) support in the data for the imitation parameter μ being positive. Thus, contrary to Papageorgiou (2003), our results slightly favour catch-up progress over country-specific technological progress as the channel through which accumulation of human capital affects output growth. This is quite sensible, since only about 15 per cent of the countries in our sample have economically meaningful innovation activities (see also Benhabib and Spiegel, 1994).

4.4 Conclusion

In this paper we have presented a relatively straightforward way to model unequal variances in Bayesian two-stage instrumental variable regression. We have done this using a fully Bayesian parametric approach. As noted, modelling heteroscedasticity is important also in the Bayesian instrumental variable context, since it improves the precision of estimates and the quality of predictive inference.

We used a simple production function approach as a tool to provide an empirical illustration of some properties of the heteroscedastic B2S model. On

5 See also the OLS results, where the dummies are excluded from the analysis, in Table 3 (Appendix B). Our OLS results are, in general, quite similar to those in previous studies. Specifically, our estimates for physical capital share α lie between 0.432 and 0.53 and are positive at the 1% significance level. The estimates of human capital/labour shares are relatively low and positive at the 5% level.

6 Note that multiplying L_t by h_t seems to reduce β more in the IV models than in the ordinary regression model.

the basis of residual plots and estimated discrepancies between the data and the models, we have shown that the data lend strong support to the use of the HB2S model instead of the homoscedastic B2S model.

Because our modelling of heteroscedasticity is relatively limited, we suggest that future research on Bayesian IV regression under unequal variances should focus on multiplicative heteroscedasticity, which is flexible and includes most of the useful formulations for heteroscedasticity as special cases (see e.g. Tanizaki and Zhang, 2001).

FIGURES

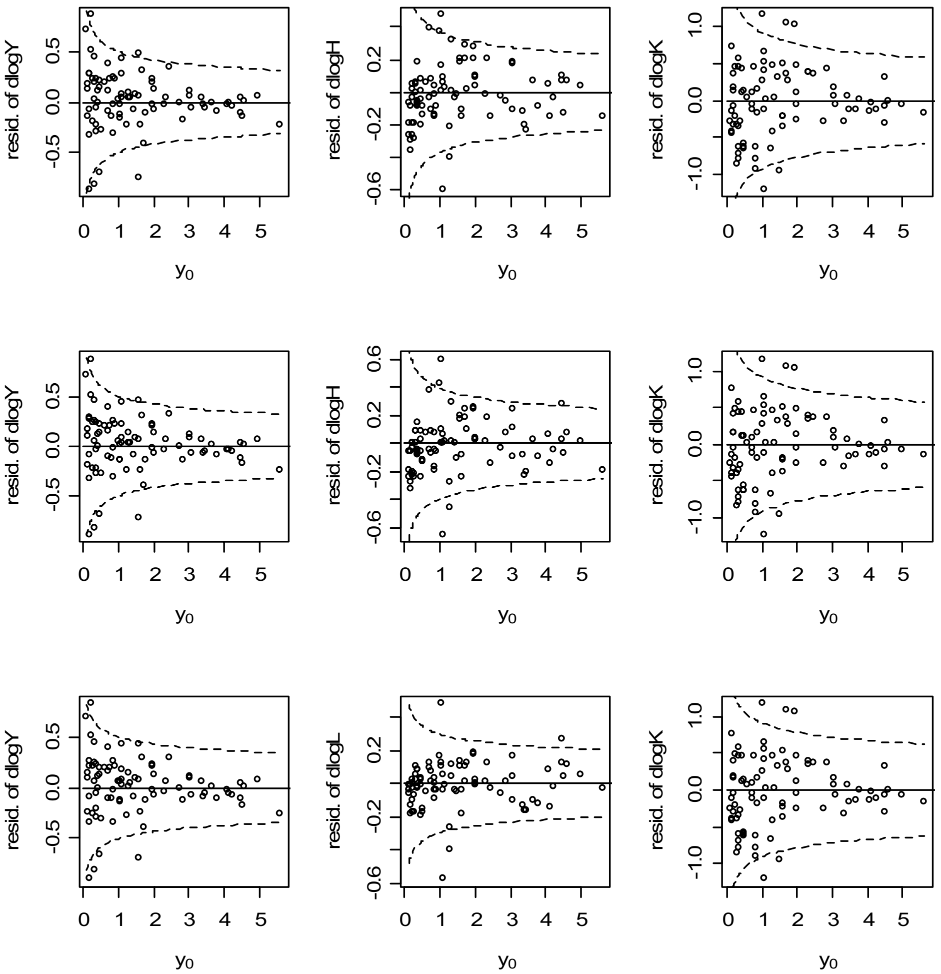

FIGURE 1 Residual plots of the first- and second-stage regressions, corresponding to equations (8) and (7), respectively. The dotted lines are approximate 95% probability intervals, based on the normality assumption. The first row gives the residual plots against y_{i0}, when $\log(Y_T/Y_0)$, $\log(H_{Y,T}/H_{Y,0})$ and $\log(K_T/K_0)$ of model (17) are regressed on the instrumental variables. The second row gives the corresponding residual plots for $\log(Y_T/Y_0)$, $\log(H_T/H_0)$, and $\log(K_T/K_0)$ when model (18) is used. Finally, the third row gives the residual plots for $\log(Y_T/Y_0)$, $\log(L_T/L_0)$ and $\log(K_T/K_0)$, corresponding to model (19). The residuals are obtained when the unknown parameters are replaced by their posterior means.

REFERENCES

Anderson, T.W. and Rubin, H., 1949. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics* 21, 570-582.

Barro, R.J., 1999. Notes on Growth Accounting, *Journal of Economic Growth* 4, 119-137.

Bauwens, L. and van Dijk, H.K., 1989). Bayesian Limited Information Analysis. in J.J. Gabszewich, et al. (Ed.), Economic Decision-Making: Games, Econometrics and Optimisation. North-Holland Publishing Co., Amsterdam.

Benhabib, J. and Spiegel, M.M., 1994. The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data, *Journal of Monetary Economics* 34, 143-73.

Bils, M. and Klenow, P.J., 2000. Does Schooling Cause Growth?. *American Economic Review* 90, 1160-182.

Boscardin, J.W. and Gelman, A., 1996. Bayesian Computation for Parametric Models of Heteroscedasticity in the Linear Model. in *Advances in Econometrics* 11, A 87-110.

Chao, J.C. and Phillips, P.C.B., 1998. Posterior Distribution in Limited Information Analysis of the Simultaneous Equations Models Using the Jeffreys' Prior. *Journal of Econometrics* 87, 49-86.

Chao, J.C. and Phillips, P.C.B., 2002. Jeffreys' Prior Analysis of the Simultaneous Equations Model in the Case of n+1 Endogenous Variables. *Journal of Econometrics* 111, 251-283.

Cohen, D. and Soto, M., 2001. Growth and Human Capital: Good Data, Good Results. OECD Development Centre Technical Papers No.179.

Cumby, R.E., Huizinga, J. and Obstfeld, M., 1983. Two-Step Two-Stage Least Squares Estimation in Models With Rational Expectations. *Journal of Econometrics* 21, 333-355.

Davidson, R., and MacKinnon, J.G., 1993. Estimation and Inference in Econometrics, 2nd ed. New York: Oxford University Press.

Dreze, J.H., 1976. Bayesian Limited Information Analysis of the Simultaneous Equation Model. *Econometrica* 44, 1045-1075.

Drèze, J.H. and Morales, J.A., 1976. Bayesian full information analysis of simultaneous equations, *J. Amer. Statist. Assoc*. 71, 919–923.

Drèze, J.H. and Richard, J. F., 1983. Bayesian analysis of simultaneous equation systems. in: Griliches, Z. and Intriligator, M., eds. Vol. 1. *Handbook of Econometrics*. Amsterdam: North-Holland.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W., 2005. Growth Econometrics. in *Handbook of Economic Growth* 1, 555-677.

Gallup, J.A., Sachs, J. and Mellinger, A., 1998. Geography and Economic Development. NBER Working Paper No: 6849.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457-511.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004. *Bayesian Data Analysis*, 2nd edition. Chapman & Hall/CRC.

Geweke, J., 1993. Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics* 8, S19-S40.

Geweke, J., 1996. Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75, 121-146.

Greene, W.H., 1990. *Econometric Analysis*. New York: Macmillian.

Griffiths, W.E., 1999. Heteroscedasticity. Working Papers in Econometrics and Applied Statistics, Department of Econometrics, University of New England, No. 104.

Griliches, Z. and Mairesse, J., 1995. Production Functions: The Search for Identification. NBER manuscript.

Judge, G.G., Griffiths, W.E., Hill, R.C. and Lee, T.C., 1985. *The Theory and Practice of Econometrics*. New Your: John Wiley and Sons.

Kleibergen, F. and Zivot, E., 2003. Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics* 114, 29-72.

Kleibergen, F. and van Dijk, H.K., 1998. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory* 14, 701-743.

Papageorgiou, C., 2003. Distinguishing Between the Effects of Primary and Post-Primary Education on Economic Growth. *Review of Development Economics* 7, 622-635.

Romer, P., 1990. Endogenous Technological Change. *Journal of Political Economy* 98, 71-96.

Shea, J., 1997. Instrument Relevance in Multivariate Linear Models: A Simple Measure. *The Review of Economics and Statistics* 79, 348-352.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 64, 583–639.

Stock, J.H. and Yogo, M., 2002. Testing For Weak Instruments in Linear IV Regression. NBER Technical Working Paper No. 284.

Tanizaki, H. and Zhang, X., 2001. Posterior Analysis of the Multiplicative Heteroscedasticity Model. *COMMUN. STATIST.-THEORY METH.* 30, 855-874.

Theil, H., 1953. Estimation and Simultaneous Correlation in Complete Equation Systems. *Mimeographed Memorandum of the Central Planning Bureau, The Hague*.

Temple, J., 1999a. The New Growth Evidence. *Journal of Economic Literature* XXXVII, 112-156.

Temple, J., 1999b. A Positive Effects of Human Capital on Growth. *Economic Letters* 65, 131-134.

Temple, J., 2001. Generalizations that Aren't? Evidence on Education and Growth. *European Economic Review* 45, 905-918.

White, H., 1982. Instrumental Variable Regression with Independent Observation. *Econometrica* 50, 483-499.

Zivot, E., Startz, R. and Nelson, C.R., 1998. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review* 39, 1119-1144.

APPENDIX A

Data and Instruments

Our estimation involves data on 85 countries (see Table 4 in Appendix B). The stock of physical capital is estimated using each country's investment rates from Penn World Tables 6.1 and perpetual inventory methods. The capital stock in 1960 is estimated using $K_i = I_i/(g_i+d+n_i)$, where I denotes the investments, g the growth rate of GDP per worker, d the depreciation rate and n the growth rate of the population, calculated as the average growth rate from 1961 to 1970. The depreciation rate d is assumed to be 0.07.[7]

In the case of human capital we follow Bils and Klenow (2000), who approximate the human capital per person using the years of schooling per person and the experience of each age group. Specifically, we assume that the log of human capital stock of a worker of age a is

$$\ln h(a) = f(s) + \gamma_1(a-s-6) + \gamma_2(a-s-6)^2, \qquad (20)$$

where γ_1 and γ_2 are parameters of return to experience, s is average years of schooling and $f(s) = \theta \cdot s^{1-\psi}/(1-\psi)$; $\psi > 0$, $\theta > 0$. Equation (20) is of the same form as that of Bils and Klenow (2000); however we assume that the influence of a teacher on human capital is zero. Using Equation (20) we calculate the average human capital stock for all age groups between 20 and 59 in 1970 and 2000 by weighting the human capital of the age group by its proportion of the country's total population. In Equation (20) we set γ_1 = 0.0512 and γ_2 = -0.00071, which corresponds to the average estimates across 52 countries as reported in Bils and Klenow (2000). We set ψ at 0.28 and set θ so that the mean of $f'(s) = \theta / s^{\psi}$ equals the mean Mincerian returns across 56 countries, which is 0.099 (see Bils and Klenow, 2000). Finally, $H_{A,it}$ is determined as H_{it} times the percentage of the population aged 15 or over with some higher education (complete + incomplete).

The education series are from Cohen and Soto (2001). The population data are from the International Data Base of the U.S. Census Bureau (Population Division of the International Programs Center (IPC)) and the United Nations population data (1995). The labour stock (L) in each country is obtained from World Development Indicators (2002). The output series have been taken from Penn World Tables 6.1.

---

7 We have predicted the missing GDP and investment values for some countries in our data. The missing GDP values are predicted using the linear trend model while the investment share (I/GDP) values are predicted using the latest available data points. These countries (and missing years) are Angola (1997-2000), Central Africa (1999-2000), Cyprus (1997-2000), Fiji (2000), Guyana (2000), Haiti (1999-2000), Sierra Leone (1997-2000) and Singapore (1997-2000).

Since physical and human capitals are accumulated factors, they are endogenous. This causes the simple OLS estimator to be inconsistent. A common means of dealing with the issue of endogeneity is to instrument for endogenous regressors with variables correlated with them but exogenous to them and the regressed variable. Moreover, the validity of an instrument requires that it cannot be a direct growth determinant or correlated with omitted growth determinants (see e.g. Durlauf et al., 2005). Therefore, we instrument the growth rates of aggregate human and physical capital using the distance from the equator (Gallup, Sachs and Mellinger 1998) and the following variables in 1970: age dependency ratio (dependents to working-age population), illiteracy rate (%) of people aged 15-24 from World Development Indicators (2002), and the level of physical capital per worker.

We make the assumption that the distance of a country from the equator, the initial (year 1970) values for age dependency ratio and youth illiteracy rate are not direct growth determinants; rather they influence the environment and investment culture, where individuals accumulate physical and human capital. For a more detailed discussion on these topics see e.g. Durlauf et al. (2005).

Since one may question the validity of our instruments, we check the consistency of the IV estimators using two specification tests. Firstly, Hansen's test for over-identification restrictions is used to see whether the model specification is correct and the instruments are uncorrelated with the error process. The second test is for weak instruments. We follow Stock and Yogo (2002), who propose quantitative definitions of weak instruments based on the maximum IV estimator bias or the maximum Wald test size distortion. The smallest p-value of the Hansen's test for over-identification restrictions for the regression models in this paper is 0.88 and the smallest test statistic of the Stock and Yogo test for weak instruments is 9.17. Thus, we can reject the null of weak instruments and can not reject the null of appropriate instruments at the 5% level. Note that we use classical tests here, since these are readily available and do not demand extra programming effort. Finally, we used African and Latin American country dummies since, based on the above test results, our instruments behave much more appropriately when these dummies are included in the analysis. The reason may be that these dummies approximate some omitted growth determinants which may be correlated with some of our instruments.

APPENDIX B

TABLE 2 Summary of the posterior simulation of θ and B when model (17) was used. The elements of B are listed by columns. These results were obtained using the summary function of the R package MCMCpack.

	Empirical mean and standard deviate-ons, plus standard errors of the means:				Quantiles for each variable				
Parameter	Mean	SD	Naiv.SE	T-S SE	2.5%	25%	50%	75%	97.5%
θ	0.5286	0.1166	0.0010	0.0024	0.3032	0.4485	0.5279	0.6066	0.7612
b_{11}	-0.6697	0.2886	0.0024	0.0060	-1.2414	-0.8623	-0.6662	-0.4807	-0.1021
b_{12}	0.4412	0.4261	0.0035	0.0085	-0.4085	0.1574	0.4451	0.7266	1.2684
b_{13}	-0.0096	0.0036	2.9e-05	6.6e-05	-0.0167	-0.0120	-0.0095	-0.0072	-0.0025
b_{14}	-0.3303	0.0690	0.0006	0.0014	-0.4674	-0.3758	-0.3293	-0.2846	-0.1947
b_{15}	4.7937	0.8125	0.0066	0.0162	3.2129	4.2394	4.7897	5.3363	6.3840
b_{16}	-1.1493	0.1901	0.0016	0.0039	-1.5192	-1.2758	-1.1507	-1.0225	-0.7689
b_{17}	-0.6680	0.1530	0.0012	0.0030	-0.9682	-0.7687	-0.6686	-0.5667	-0.3633
b_{18}	0.1390	0.2506	0.0020	0.0051	-0.3560	-0.0280	0.1406	0.3074	0.6310
b_{19}	-0.0040	0.1059	0.0009	0.0021	-0.2108	-0.0760	-0.0042	0.0664	0.2024
b_{21}	-0.5981	0.1200	0.0010	0.0026	-0.8356	-0.6786	-0.5986	-0.5172	-0.3663
b_{22}	1.4091	0.1778	0.0015	0.0035	1.0535	1.2914	1.4107	1.5285	1.7531
b_{23}	-0.0004	0.0015	1.2e-05	3.1e-05	-0.0032	-0.0014	-0.0004	0.0006	0.0026
b_{24}	0.0326	0.0280	0.0002	0.0006	-0.0212	0.0136	0.0325	0.0516	0.0876
b_{25}	-0.1891	0.3326	0.0027	0.0068	-0.8405	-0.4106	-0.1909	0.0383	0.4631
b_{26}	-0.1773	0.0767	0.0006	0.0017	-0.3289	-0.2286	-0.1775	-0.1260	-0.0269
b_{27}	-0.1368	0.0614	0.0005	0.0013	-0.2558	-0.1783	-0.1359	-0.0952	-0.0165
b_{28}	-0.1324	0.1001	0.0008	0.0020	-0.3321	-0.1982	-0.1314	-0.0664	0.0646
b_{29}	0.0122	0.0422	0.0003	0.0008	-0.0725	-0.0158	0.0128	0.0409	0.0946

TABLE 3 OLS results of growth regressions for Equations 17-19 (dummies excluded)

Parameters	Equation (17)	Equation (18)	Equation (19)
Constant	0.197	0.159	0.256
	(0.073)	(0.094)	(0.068)
$(a+\beta)(\delta-\mu)$	-0.03	0.005	-0.003
	(0.054)	(0.017)	(0.015)
$(a+\beta)\mu$	0.057[c]	0.001[b]	0.001[b]
	(0.018)	(0.0005)	(0.0004)
β	0.231[b]	0.267[b]	0.202[b]
	(0.104)	(0.124)	(0.113)
a	0.513[c]	0.500[c]	0.534[c]
	(0.060)	(0.064)	(0.056)
$\overline{R}^2$	0.587	0.592	0.577
OBS	85	85	85

b p-value of one-sided hypothesis test < 0.05
c p-value of one-sided hypothesis test < 0.01

White's heteroscedasticity-corrected standard errors in parentheses.

TABLE 4 Sample of 85 countries

Algeria	Malawi	Colombia	Fiji	Germany
Egypt	Mali	Costa Rica	Indonesia	Greece
Jordan	Mauritius	Dominican Rep	Korea South	Ireland
Morocco	Mozambique	Ecuador	Malaysia	Italy
Syria	Niger	El Salvador	Philippines	Japan
Tunisia	Nigeria	Guatemala	Thailand	Netherlands
Angola	Senegal	Guyana	Bangladesh	New Zealand
Benin	Sierra Leone	Haiti	India	Portugal
Burkina Faso	South Africa	Honduras	Nepal	Singapore
Burundi	Tanzania	Jamaica	Australia	Spain
Cameroon	Uganda	Mexico	Austria	Sweden
Cen. African Rep.	Zambia	Nicaragua	Belgium	Switzerland
Ethiopia	Zimbabwe	Panama	Canada	United Kingdom
Gabon	Argentina	Paraguay	Cyprus	United States
Ghana	Bolivia	Peru	Denmark	Hungary
Kenya	Brazil	Uruguay	Finland	Romania
Madagascar	Chile	China	France	Turkey

5 ESSAY 4: A NAIVE STICKY INFORMATION MODEL OF HOUSEHOLDS' INFLATION EXPECTATIONS[1]

ABSTRACT

This paper provides a simple epidemiology model where households, when forming their inflation expectations, rationally adopt the past release of inflation with certain probability rather than the forward-looking newspaper forecast as suggested in Carroll [2003, Macroeconomic Expectations of Households and Professional Forecasters, Quarterly Journal of Economics, 118, 269-298]. The posterior model probabilities based on the Michigan survey data strongly support the proposed model. We also extend the agent-based epidemiology model by deriving for it a simple adaptation, which is suitable for estimation. Our results show that this model is able to capture the heterogeneity in households' expectations very well.

Keywords: Inflation expectations; Heterogeneous expectations; Survey expectations; Sticky information; Bayesian analysis

---

1 The first and very preliminary version of this paper has been accepted for publication in Applied Economics. Jani Luoto is responsible for statistical estimation, writing the theory, interpreting the results, and for most of the writing. In this new paper Markku Lanne and Arto Luoma assisted in interpreting the results and in some of the writing.

5.1 Introduction

In recent years there has been an increasing interest in explaining agents' inflation expectations formation process (see, e.g., Mankiw and Reis, 2006 and 2007, Sims, 2006, Trabandt, 2007, and Branch, 2004 and 2007). This is mainly due to observed failure of the rational expectation hypothesis. Within this literature, Mankiw and Reis (2002) propose a simple sticky information model where agents know the true probability distribution of the economy, but update their information set each period with certain probability. Carroll (2003, 2006) and Reis (2006a,b) seek microfoundations for sticky information models, while Mankiw et al. (2003) and Carroll (2003, 2006) find evidence based on survey data supporting these models (see also Khan and Zhu, 2006, Andres et al., 2005, Kiley, 2007, Coibion, 2006 and 2007, and Doepke et al., 2008). Finally, Branch (2007) bridges the sticky information and heterogeneous expectations literatures by presenting empirical evidence in favor of both, model heterogeneity and limited information flows (see Branch, 2007, and references therein).

Closest to our work, Carroll (2003) develops and estimates an expectation formation model, where the general public adopt professionals' forecast with certain probability, rather than form their own rational forecasts. The structure of his model was inspired by simple models of disease spread from the epidemiology literature, and it provides promising microfoundations for sticky information models. To the best of our knowledge, it is also unique in relaxing the assumption that an ordinary person either knows the true probability distribution of the economy or can estimate some sophisticated econometric model[2] when forming expectations. This relaxation is, however, important, since although trained economists might have this kind of knowledge, it would probably be an overwhelming task for an ordinary consumer (producer) (see Shiller, 1997). According to Carroll (2003), it might require, for example, obtaining a Ph.D. degree in economics first.

Despite the virtues of Carroll's (2003) model, recent work has cast doubt on the reliability of professionals' inflation forecasts, and, in general, on traditional approaches to inflation forecasting (see e.g. Atkeson and Ohanian, 2001, Fisher et al., 2002, Sims, 2002, Stock and Watson, 2002 and 2007), and Brave and Fisher (2004). In particular, Atkeson and Ohanian (2001) found that since 1984 the one-year-ahead inflation forecast of professionals[3] has not been better than the "naïve" forecast given by the inflation rate over the previous year. Thus, it is natural to question the rationale of searching for relatively rare newspaper forecasts (or to form one's own rational forecast), when the most recently reported past inflation statistic provides a competitive forecast 'model' for future inflation. In this paper, we propose an epidemiology model, where agents simply adopt with certain probability the past release of the annualized

---

2 That is, agents are assumed to be 'boundedly' rational; see e.g. Evans and Honkapohja (2001).

3 Specifically, Federal Reserve's Greenbook forecasts.

monthly inflation figure, *the most commonly reported figure in the news* coverage of inflation. We refer to this model as the naïve sticky information model, and test it empirically against Carroll's sticky information approach using quarterly U.S. data. Specifically, we compare posterior probabilities of the alternative models in which households update their expectations either to the forward-looking newspaper forecast or to the most recently reported past inflation statistic. As will be seen, U.S. data strongly support the latter.

Based on our empirical findings, we extend the agent-based epidemiology model, proposed by Carroll (2006), by deriving a relative simple adaptation of that model, suitable for estimation. The model assumes a constant personal probability for each agent to read a newspaper article on inflation. This variation in their newspaper reading propensities could explain differences in survey expectations across demographic groups, documented in Bryan and Venkatu (2001a, b) and Souleles (2004). The model differs from that of Carroll (2003, 2006) in that it no longer assumes agents to be 'infected' by rare newspaper forecasts. Rather, the source of 'infection' is the past release of annualized monthly inflation. The model is estimated with classified household-level survey data from 1981/3 to 2001/4 constructed by the Survey Research Center (SRC) at the University of Michigan. The results indicate that people on average update their expectations roughly once a year, which is in accordance with the previous literature, while their updating probabilities vary from 0.12 to 0.42.

An agent-based epidemiology model also captures the overall heterogeneity between agents' expectations fairly well, in the sense that the variance of unexplained heterogeneity (σ^2), i.e. heterogeneity in agents' expectations which the underlying model cannot explain, is quite small (approximately 1.1) relative to the high degree of heterogeneity observed in the actual micro level data. For example, in Branch's (2007) Rationally Heterogeneous Expectations (RHE) sticky information model, σ^2 was 36. Although our result is not fully comparable to that of Branch (2007), we note that in the RHE model most variation in agents' expectations is attributed to unexplained heterogeneity[4].

This paper is organized as follows. In Section 2, we discuss and estimate two alternative models where agents update their expectations either to the forward-looking newspaper forecast or to the most recently reported past inflation statistic. Section 3 provides an adaptation of the agent-based epidemiology model and estimates it using classified household-level survey data. Finally, Section 4 concludes the paper.

---

4 The empirical standard deviation of Branch's (2007) sample was 12.7010. However, according to Branch (2007), the large empirical standard deviation is accounted for by a few outliers that expect inflation to be greater than 40%. Since his estimate of standard deviation of unexplained heterogeneity is 6 we state that in his model most variation in agents' expectations is attributed to unexplained heterogeneity.

5.2 Population Mean Analysis

In this section, we introduce two alternative versions of an epidemiological expectation formation model and test them against each other. In both models agents face a constant probability of reading an article on inflation, and they believe that inflation follows a random walk (see Carroll, 2003). However, in the first model, proposed by Carroll (2003), agents also believe that a forecast from a professional forecaster is more accurate than a forecast that they could construct themselves. In the alternative model, we give up this potentially invalid assumption.

5.2.1 Naïve Sticky Information Model

Epidemiological information structure provides promising microfoundations for sticky information models. In an epidemiology model, each individual at any time point faces a constant probability (say λ) of observing an article on inflation, which Carroll assumes to consist of professional forecasters' forecasts. Individuals who do not observe such an article simply continue to believe the last forecast they read about. This information structure leads to the following relationship between the mean inflation expectation of the general public $M_t[\cdot]$ and the newspaper forecast $N_t^f[\cdot]$ of professional forecasters at time t (see Carroll, 2003, for a systematic treatment and references),

$$M_t[\pi_{t+1}] = \lambda N_t^f[\pi_{t+1}] + (1-\lambda)\{\lambda N_{t-1}^f[\pi_{t+1}] + (1-\lambda)(\lambda N_{t-2}^f[\pi_{t+1}] + \ldots)\}. \tag{1}$$

where $M_t[\cdot]$ is an operator that yields the population-mean value of the people's inflation expectations at time t and π_{t+1} is measured as the quarterly mean of monthly inflation of the seasonally adjusted Consumer Price Index (CPI) for all urban consumers.[5]

In what follows, model (1) can be written, using certain assumptions on the public's beliefs about the process of inflation, as

$$\begin{aligned} M_t[\pi_{t,t+4}] &= \lambda N_t^f[\pi_{t,t+4}] + (1-\lambda)\{\lambda N_{t-1}^f[\pi_{t-1,t+3}] + (1-\lambda)(\lambda N_{t-2}^f[\pi_{t-2,t+2}] + \ldots)\} \\ &= \lambda N_t^f[\pi_{t,t+4}] + (1-\lambda) M_{t-1}[\pi_{t-1,t+3}], \end{aligned} \tag{2}$$

5 In this paper we use the quarterly series, since the only relevant candidate series for the views of professional forecasters which has the same forecasting horizon as the Michigan survey of households is the four-quarter inflation forecast from the Survey of Professional Forecasters (SPF). Furthermore, we use the quarterly means of monthly inflation series rather than quarterly inflation of the quarterly means of the monthly CPI series because a twelve-month forecast for annual inflation is the series that is asked from the respondents of the University of Michigan's Survey Research Center.

where $\pi_{\tau,\tau+4} = \pi_{\tau+1} + \pi_{\tau+2} + \pi_{\tau+3} + \pi_{\tau+4}$ is the annual inflation from τ to $\tau + 4$. Equation (2) provides a testable implication for the public's mean inflation expectations. It is based on two fundamental assumptions. Firstly, people believe that the economy has an underlying 'fundamental' inflation rate, π_t^f, and that the future changes in this fundamental rate are unforecastable. That is, they believe that $\pi_t = \pi_t^f + e_t$ and $\pi_{t+1}^f = \pi_t^f + \eta_{t+1}$, where e_t is a transitory shock to the inflation rate, unforecastable beyond period t, while η_{t+1} is a permanent innovation in the fundamental inflation rate, unforecastable beyond period t + 1. Secondly, people also believe that professional forecasters have some deeper knowledge on how the economy works, and, therefore, are capable to estimate the past and present values of e and η through periods t and $t + 1$, respectively. These assumptions are in line with the near-unit-root behavior of the inflation rate, which is well documented in the empirical literature, and with the result of Shiller (1997) that ordinary persons do not know the causes of inflation. However, recent work has cast doubt on the reliability of professionals' inflation forecasts and the reliability of traditional approaches to forecasting inflation, suggesting that it might be rational for agents to use frequently reported actual inflation figures, rather than rare newspaper forecasts, as their inflation expectations (see e.g. Atkeson and Ohanian, 2001, Fisher et al., 2002, Sims, 2002, Stock and Watson, 2002 and 2007, and Brave and Fisher, 2004).

To this end, we propose an alternative version of an epidemiological expectation formation model, which we call a naïve sticky information model. In the model, we follow Carroll and assume that people believe inflation to follow a random walk.

Equation (1) is based on the assumptions that every inflation article contains a complete forecast of the inflation rate for all future periods and that the agent who reads an article can recall the entire forecast. However, when these newspaper forecasts are on average no better than the naïve forecasts, people would eventually note this and rationally expect that $F_{t-j}[\pi_{t+1}] = N_{t-j}[\pi_{t-j}] = N_{t-j}^f[\pi_{t+1}]$, where j = 0,1,2,…, $N_t[\pi_t]$ is the actual inflation reported in the news media ($N[\cdot]$) at time t, and F is a forecast operator. Furthermore, no newspaper article contains an inflation forecast into infinite future. Rather, newspaper forecasts are rare compared to the frequently presented actual inflation figures. That is, the most recently reported inflation statistic $N_t[\pi_t]$ is the figure which people most probably observe when being influenced by the news media. Under these assumptions Equation (1) can be rewritten as

$$M_t[\pi_{t+1}] = \lambda F_t[\pi_{t+1}] + (1-\lambda)\{\lambda F_{t-1}[\pi_{t+1}] + (1-\lambda)(\lambda F_{t-2}[\pi_{t+1}] + \ldots)\}$$

$$= \lambda N_t[\pi_t] + (1-\lambda)\{\lambda N_{t-1}[\pi_{t-1}] + (1-\lambda)(\lambda N_{t-2}[\pi_{t-2}] + \ldots)\}. \quad (3)$$

Unfortunately, Equation (3) can be tested empirically only with annual data, since the available survey data only provide households' inflation expectations over the next year (i.e. their forecasts of annual inflation). Given the available dataset, we have roughly 20 annual observations, which are too few for valid inference. In addition, we prefer using quarterly data to keep the results comparable to those in Carroll (2003, 2006).

To derive a testable implication from Equation (3), let us note that under the random walk assumption, $\pi_{t,t+4} = \pi_{t,t+1} + v_{t+4}$, where $\pi_{t,t+1} \equiv 4\pi_{t+1}$ is annualized inflation and v_{t+4} a zero mean i.i.d. shock. Applying a lagged forecast operator on both sizes of this equation yields $F_{t-j}\pi_{t,t+4} = F_{t-j}\pi_{t,t+1}$ for any j (j = 0,1,2,…). This holds since people believe that future values of v_t are unforecastable. Furthermore, according to the random walk assumption, people also believe that the best predictor for the t + 1 period annualized inflation, $F_{t-j}\pi_{t,t+1}$, is $N_{t-j}[\pi_{t-j-1,t-j}]$. Thus, we have the following empirically testable equation for the population mean of inflation expectations,

$$
\begin{aligned}
M_t\left[\pi_{t,t+4}\right] &= \lambda F_t\left[\pi_{t,t+4}\right] + (1-\lambda)\left\{\lambda F_{t-1}\left[\pi_{t,t+4}\right] + (1-\lambda)\left(\lambda F_{t-2}\left[\pi_{t,t+4}\right] + \ldots\right)\right\} \\
&= \lambda N_t\left[\pi_{t-1,t}\right] + (1-\lambda)\left\{\lambda N_{t-1}\left[\pi_{t-2,t-1}\right] + (1-\lambda)\left(\lambda N_{t-2}\left[\pi_{t-3,t-2}\right] + \ldots\right)\right\} \\
&= \lambda N_t\left[\pi_{t-1,t}\right] + (1-\lambda) M_{t-1}\left[\pi_{t-1,t+3}\right], \qquad (4)
\end{aligned}
$$

where $N_t[\pi_{t-1,t}]$ is the annualized inflation reported in the news media at time t. Equation (4) predicts the mean inflation expectation of the next year as a weighted average of the latest annualized inflation figure and the lagged mean inflation expectation. In our paper, annualized inflation is measured as (a quarterly mean of) annualized monthly inflation, the figure which the ordinary person most probably observes when being influenced by the news media.

5.2.2 Econometric Approach

Since the models in Equations (2) and (4) are non-nested, conventional tests cannot be used to the test them against each other. We therefore apply posterior model probabilities to explore if either of the models (2) or (4) is the true data generating process of the public's inflation expectation formation.

Given the data y, and competing models $M_1,\ldots,M_K$ with parameter vectors θ_k, $k = 1,\ldots,K$, the posterior model probability for model M_k is given by

$$
p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{i=1}^{K} p(y|M_i)p(M_i)}, \qquad (5)
$$

where

$$
p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k, \qquad (6)
$$

is the marginal likelihood of model k, $p(\theta_k | M_k)$ the prior density of θ_k under model M_k, $p(y | \theta_k, M_k)$ the likelihood, and $p(M_k)$ the prior probability of M_k. The explored models are standard linear regression models corresponding to theoretical models such as specifications (2) and (4). We will assume uniform independent prior distributions on given intervals for the regression coefficients. For simplicity, we also assume that the error variance σ is, a priori, distributed uniformly on the interval [0,2]. Furthermore, we give the models equal prior probabilities, that is, $p(M_k) = 1/K$, $k = 1,..,K$. Given these priors it is straightforward to derive analytical solutions for Equations (5) and (6).

However, since these uniform priors are not necessarily non-informative, we control our results using approximate posterior model probabilities based on the Schwarz Bayesian Information Criterion (BIC) (see Schwarz, 1978, and Garratt et al., 2007). Specifically,

$$\ln p(y|M_k) \approx \ln p_{BIC}(y|M_k) \equiv l - \frac{\ln(T) \times p}{2}, \tag{7}$$

where l denotes the log of the likelihood function evaluated at the maximum likelihood estimates, p denotes the number of parameters in the model, and T is the sample size. We use the previous marginal likelihood approximation, since it selects the same model as BIC, familiar to non-Bayesians.

5.2.3 Posterior Model probabilities

We estimate Equations (2) and (4) using a monthly survey of inflation expectations of approximately 500 households, conducted by the Survey Research Center (SRC) at the University of Michigan, and the mean four-quarter inflation forecast from the Survey of Professional Forecasters[6] (SPF) as proxies for households' expectations and the newspaper forecast, respectively. The quarterly average series of annualized monthly inflation is based on seasonally adjusted real-time CPI (for all urban consumers) data[7]. The sample period ranges from 1981/3 to 2001/4, for which all the series are available. Specifically, we run the following regressions,

$$M_t[\pi_{t,t+4}] = \alpha_1 N_t[\pi_{t-1,t}] + \alpha_2 M_{t-1}[\pi_{t-1,t+3}] + \varepsilon_{1,t}, \tag{8}$$

$$M_t[\pi_{t,t+4}] = \alpha_3 N_t^f[\pi_{t,t+4}] + \alpha_4 M_{t-1}[\pi_{t-1,t+3}] + \varepsilon_{2,t}, \tag{9}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are assumed to be zero mean normally distributed errors with variances σ_1^2 and σ_2^2, respectively. We refer to Equations (8) and (9) as models M_1 (naïve model) and M_2 (Carroll's model), respectively. In these models $a_1 + a_2$ and $a_3 + a_4$ are set at unity, since we are interested in situations where the epidemiology model can be treated as a structural description of the true

---

6 Data set is compiled by the Federal Reserve Bank of Philadelphia and is available at http://www.phil.frb.org/econ

7 Data set is available at http://econweb.rutgers.edu/nswanson/realtime.htm

process of the public's inflation expectations formation. We assume for the a_1 and a_3 uniform independent prior distributions on the interval [0,1].

The upper panel of Table 1 gives the summary statistics for the models M_1 and M_2. According to the posterior model probabilities, there is a strong support in the data for model M_1. Specifically, the posterior probability of the naïve sticky information model being the true model exceeds 99.9 percent. Furthermore, the point estimate of a_1 (0.18) is of sensible magnitude and in accordance with earlier empirical results. Conversely, the data do not lend support to model M_2; its posterior probability of being the true model is very close to zero.

TABLE 1 Posterior Model Probabilities for Mean Models

Estimated Models

$M_t[\pi_{t,t+4}] = a_1 N_t[\pi_{t-1,t}] + a_2 M_{t-1}[\pi_{t-1,t+3}] + \varepsilon_{1t}$, and $M_t[\pi_{t,t+4}] = a_3 N^f_t[\pi_{t,t+4}] + a_4 M_{t-1}[\pi_{t-1,t+3}] + \varepsilon_{2t}$

	Estimates and Posterior Model Probabilities					
Models	a_1	a_2	a_3	a_4	$p(M_k \mid y)$	$p_{BIC}(M_k \mid y)$
M_1	0.18	0.82	-	-	99.993%	99.997%
	(0.02)	(0.02)				
M_2	-	-	0.35	0.65	0.006%	0.002%
			(0.07)	(0.07)		

Estimated Model

$M_t[\pi_{t,t+4}] = a_5 N^f_t[\pi_{t,t+4}] + a_6 M_{t-1}[\pi_{t-1,t+3}] + a_7 N_t[\pi_{t-4,t}] + \varepsilon_{3t}$

	Estimates and Posterior Model probabilities				
Models	a_5	a_6	a_7	$p(M_k \mid y)$	$p_{BIC}(M_k \mid y)$
M_3	0.45	0.64	-0.09	0.001%	0.001%
	(0.09)	(0.07)	(0.07)		
M_4	0.54	0.54	-0.05	0.000%	0.000%
	(0.11)	(0.09)	(0.07)		

All Equations are estimated over the period 1981:3 to 2001:4 for which real-time inflation, Michigan and SPF expectations series are available. Posterior standard errors (in parenthesis). There is no evidence of serial correlation or heteroscedasticity in the residuals of any reported regression.

Carroll (2003) suggests expanding model (9) to include the recently published *annual* inflation to test for the possibility that (a fraction of) individuals form their own forecast, the 'adaptive expectations' model, rather than use the forward-looking newspaper forecast. Using real-time CPI data, we therefore run the following regression

$$M_t[\pi_{t,t+4}] = \alpha_5 N_t^f[\pi_{t,t+4}] + \alpha_6 M_{t-1}[\pi_{t-1,t+3}] + \alpha_7 N_t[\pi_{t-4,t}] + \varepsilon_{3t}, \qquad (10)$$

where $N_t[\pi_{t-4,t}]$ is the annual inflation reported in the news media at time t. We estimate this model with and without the constraint $a_5 + a_6 + a_7 = 1$ and refer to these models as M_3 and M_4, respectively. We assume for the a_5 and a_6 uniform independent prior distributions on the interval [0,1] and for the a_7 on [–1,1]. The results presented in the lower panel of Table 1 provide strong evidence against the adaptive expectations model. The posterior probabilities of the models M_3 and M_4 are virtually zero. Furthermore, the estimated coefficients of annual inflation are negative, which is counterintuitive, but their posterior intervals include zero. Carroll (2003), in contrast, obtained negative and statistically significant estimates. The fact that our results deviate from his probably follows from the differences in the data: we used real-time inflation figures, while he uses final vintage data. Finally, including a constant in the previous regressions does not alter the qualitative results. We will not present such results here, since, according to Carroll (2003), the presence of a positive constant term could reflect the effect of social transmission of inflation expectations (e.g. conversations with neighbors), in addition to the news-media channel explored in our paper.

In sum, these results are quite impressive in supporting the naïve sticky information model. However, there is likely to be heterogeneity in households' expectations that cannot be captured by the model. We will therefore, in the next section, estimate an agent-based version of the naïve sticky information model, where each individual has his own constant newspaper-reading propensity λ and a fraction of population are allowed to round their inflation forecasts to the closest 0%, 5%, 10% or 15%.

5.3 An Agent-Based Epidemiology Model

In this section we provide a relatively simple adaptation of an agent-based epidemiology model. We estimate it using classified household-level survey data of approximately 500 households, constructed by the Survey Research Center (SRC) at the University of Michigan. We use a quarterly dataset from 1981/3 to 2001/4 to keep results comparable to those presented in the previous section and in Carroll (2003, 2006). Furthermore, this dataset is easily available for the public. In it, the inflation expectation of each household is classified into one of seven categories.

5.3.1 The Model

Inspired by the preliminary results presented in the previous section, we estimate an agent-based version of the epidemiological model, where, in any given period t, each agent faces a constant personal probability λ_i ($i = 1,\ldots,P$, where P is the size of population) of reading a newspaper article on the latest inflation figure. If he does not encounter such an article, his probability to use the views of period $t - 1$ about inflation is $(1-\lambda_i)\lambda_i$. Generally, the probability that he uses the newspaper view of period $t - j + 1$ in period t is

$$p(j \mid \lambda_i) = \lambda_i (1 - \lambda_i)^{j-1}, \qquad j = 1, 2, \ldots \qquad (11)$$

This is the probability function of the geometric distribution and we can easily compute an individual's j with given λ_i. In the population level we need to choose a proper probability density function (p.d.f.) for individuals' λ_i. The beta distribution is a natural candidate for this purpose, since it is very flexible and assumes various shapes with different parameter values of a and β. Thus, we assume that across agents the p.d.f of λ is

$$p(\lambda) = \frac{1}{B(\alpha,\beta)} \lambda^{\alpha-1}(1-\lambda)^{\beta-1}, \qquad a > 0, \beta > 0, \tag{12}$$

where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the beta function and $\Gamma(\cdot)$ is the gamma function. The density (12) indicates that the higher the values of a and β, the more homogenous are agents' updating probabilities. Note also that functions (11) and (12) form the basis of the likelihood of the agent based naïve sticky information model.

Combining functions (11) and (12) yields the following joint density function for j and agents' updating probabilities λ:

$$p(\lambda, j) = \frac{1}{B(\alpha,\beta)} \lambda^{\alpha}(1-\lambda)^{\beta+j-2}. \tag{13}$$

Integrating over *all agents'* updating probabilities λ, in turn, yields the following marginal probability function for j,

$$p(j) = \frac{1}{B(\alpha,\beta)} \int_0^1 \lambda^{\alpha}(1-\lambda)^{\beta+j-2} d\lambda = \frac{1}{B(\alpha,\beta)} B(\alpha+1, \beta+j-1)$$

$$= \frac{\alpha\Gamma(\alpha+\beta)\Gamma(\beta+j-1)}{\Gamma(\beta)\Gamma(\alpha+\beta+j)}, \qquad j = 1, 2, \ldots \tag{14}$$

Thus, $p(j)$ gives the probability of a randomly picked agent using the newspaper inflation view of period $t - j + 1$ when forming his expectation at time t, assuming that λ follows the beta distribution in population level. Note that Eq. (14) is not in the form of any standard probability mass function. However, we can use numerical methods (classical or Bayesian) to estimate the mean, $E(\lambda) = \mu_\lambda$, and the variance, $Var(\lambda) = \sigma_\lambda^2$, of all agents' updating probabilities λ, using the properties of the beta distribution. That is, according to the beta distribution, the parameters a and β can be solved as functions of μ_λ and σ_λ^2 as follows: $a = \mu_\lambda^2(1 - \mu_\lambda)/\sigma_\lambda^2 - \mu_\lambda$ and $\beta = a(1/\mu_\lambda - 1)$.

To complete the model we assume that each survey respondent reports

$$\pi^e_{t,t+4} = N_{t-j+1}[\pi_{t-j,t-j+1}] + \varepsilon_{t+4}, \tag{15}$$

where his j takes values from 1 to K (we have truncated the distribution in Equation 14 such that K = 28), and ε_{t+4} is a normally distributed error term with mean zero and variance σ^2. In Eq. (15) we assume that, after observing the newspaper article on inflation, an individual makes adjustments to the data and reports the figure that corresponds to his perception of future inflation. This is the standard modeling approach in the RHE literature where the stochastic terms ε_t are interpreted as individual idiosyncratic shocks representing unobserved heterogeneity, i.e. heterogeneity in individuals' expectations which underlying model can not explain (see Branch, 2007, for discussion and references therein).

Given Equations (14) and (15) and the classified household-level survey data, the likelihood function for the sample of $T \times C$ observations, $n = (n_{11}, n_{12}, \ldots, n_{TC})'$, can be written as

$$L(\theta;n) = \prod_{t=l}^{T}\prod_{c=1}^{C}\left[P_t(c;\theta)\right]^{n_{tc}} , \tag{16}$$

where $\theta = (\mu_\lambda, \sigma_\lambda, q, \sigma)'$ is the vector of parameters, n_{tc} the number of individuals in class c at time t, and $P_t(c; \theta)$ the probability that an individual belong the class c at time t. The classes of individuals' inflation expectations, c = 1,…,7, are defined as follows: 1 = –0%, 2 = 1–2%, 3 = 3–4%, 4 = 5%, 5 = 6–9%, 6 = 10–14%, 7 = 15%+. The probabilities $P_t(c;\theta)$ are given by

$$P_t(c=1;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{-\infty}^{0.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right] + q\left[\sum_{k=1}^{K} p(k)\int_{-\infty}^{2.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=2;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{0.5}^{2.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=3;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{2.5}^{4.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=4;\theta) = (1-q)\left[\sum_{k=k}^{K} p(k)\int_{4.5}^{5.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right] + q\left[\sum_{k=1}^{K} p(k)\int_{2.5}^{7.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=5;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{5.5}^{9.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=6;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{9.5}^{14.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right] + q\left[\sum_{k=1}^{K} p(k)\int_{7.5}^{12.5} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

$$P_t(c=7;\theta) = (1-q)\left[\sum_{k=1}^{K} p(k)\int_{14.5}^{\infty} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right] + q\left[\sum_{k=1}^{K} p(k)\int_{12.5}^{\infty} p\left(\pi^e_{t,t+4} \middle| j=k\right) d\left(\pi^e_{t,t+4}\right)\right]$$

where

$$p\left(\pi^e_{t,t+4} \middle| j=k\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\pi^e_{t,t+4} - N_{t-j+1}\left[\pi_{t-j,t-j+1}\right]}{\sigma}\right)^2\right\}. \tag{17}$$

As we can see, the probability $P_t(c;\theta)$ is calculated by integrating the conditional probability density $p\left(\pi^e_{t,t+4} \middle| j=k\right)$ over the interval of class c and summing over the possible updating intervals k = 1,…,K. After observing actual household level survey data, we have also allowed for the possibility that a fraction q of the total population round their inflation expectations to the closest 0%, 5%, 10%, or 15%. This kind of behavior may be typical of those agents who have no special interest in the economy (see, e.g., Bryan and Palmqvist, 2004). Note that the extreme of q = 0 corresponds to the situation where there is no such behavior in the population.

5.3.2 Results

We will estimate the previous model using Bayesian methods. The starting point of the Bayesian analysis is to determine the prior density function of the parameters, $p(\theta)$, which together with the likelihood function (16) yields the posterior density

$$q(\theta|n) = \frac{p(\theta)L(\theta;n)}{\int p(\theta)L(\theta;n)d\theta}. \tag{18}$$

The prior density reflects the researcher's prior beliefs concerning plausible parameter values. Table 2 lists the marginal prior distributions of the parameters. A standard assumption on prior independence is used (see e.g. Zellner, 1971). Reported marginal priors reflect the following parameter constraints, $0 < \mu_\lambda < 1$, $0 < q < 1$, $0 < \sigma_\lambda$, and $0 < \sigma$. The prior mean of μ_λ is set at 0.25, which is a common result in previous studies. The prior mean of σ_λ is set at 0.06, indicating moderate variation in individuals updating probabilities (see Carroll, 2006). We have not a clear idea about the value of q, hence the uniform prior (Beta(1,1)) is entitled in this case. Furthermore, the prior mean of σ is set at 5, slightly below the estimate of Branch (2007). Finally, with given prior variances, these marginal prior distributions turned out to be practically noninformative.

TABLE 2 Priors and Posteriors for Agent-Based Epidemiology Models

The Likelihood

$$L(\theta;n)=\prod_{t=l}^{T}\prod_{c=1}^{C}\left[P_t(c;\theta)\right]^{n_{tc}}$$

	Prior Distributions			Posterior Distr. Model (17)		Posterior Distr. Model (19)	
	Distr.	Mean	St.Dev.	Median	St.Dev.	Median	St.Dev.
	1981/3-2001/4						
μ_λ	Beta	0.25	0.38	0.379	0.006	0.378	0.005
σ_λ	Invgamma	0.06	0.19	0.116	0.004	0.099	0.004
σ	Invgamma	5.00	11.00	4.337	0.014	2.171	0.027
q	Beta	0.50	0.38	0.217	0.002	0.207	0.002
a	Gamma	1.00	1.00	-	-	0.548	0.012
σ_0	-	-	-	-	-	1.607	0.017
$p(M_k \mid y)$				0.000		1.000	
	1984/1-2001/4						
μ_λ	Beta	0.25	0.38	0.193	0.002	0.220	0.005
σ_λ	Invgamma	0.06	0.19	0.008	0.002	0.024	0.007
σ	Invgamma	5.00	11.00	3.876	0.015	1.295	0.020
q	Beta	0.50	0.38	0.208	0.002	0.204	0.002
a	Gamma	1.00	1.00	-	-	0.669	0.019
σ_0	-	-	-	-	-	1.061	0.021
$p(M_k \mid y)$				0.000		1.000	

We set K at 28 (i.e. seven years) to have a sufficiently good approximation for Equation (14). The results of the regressions with shorter and longer lag lengths are quite similar. The inflation data before 1978:2 were based on the quarterly means of monthly inflation of the lagged CPI series (lagged by one month), which was acquired from Norman R. Swanson's home page.

The estimation results of model (17) are reported in Table 2.[8] As can be seen, the median of the updating probability, $\mu_\lambda = 0.38$, is higher than the point estimates of λ reported in Table 1, but it is in accordance with previous studies. The truncation of Equation (14) is most likely the reason for the difference between

8 We used the Metropolis algorithm to generate a Monte Carlo sample from the posteriors. The algorithm uses the multivariate normal distribution for the jump distribution on changes in the parameters θ. The inverse of the Hessian of the log posterior density at the posterior mode, scaled by the factor $2.4^2/4$ ($2.4^2/5$), is used to obtain an optimal covariance matrix of the multivariate normal jump distribution; see e.g. Gelman et al. (2004). We use 10,000 draws, discarding the first 2,000 as a burn-in period. As a convergence check, three chains with different randomly selected starting values are simulated. The potential scale reduction factor of Gelman and Rubin (1992) was between 1 and 1.04 for each parameter. The multivariate version of Gelman and Rubin's diagnostic, proposed by Brooks and Gelman (1997), were between 1.00 – 1.01 for each model. Finally, the frequency of accepted jumps was roughly 0.28.

the results. For comparison, we also estimated a truncated version of Equation (4) with K = 28,[9] based on Michigan mean expectations and obtained an estimate of λ quite close to that obtained for model (17).[10]

The median of the standard deviation of λ is estimated to be 0.12, indicating strong heterogeneity in agents' newspaper reading habits. The updating probability λ varies from 0.11 to 0.71 among agents according to our simulation experiment (not reported here)[11]. This strong variation in their newspaper reading propensities could explain the differences in survey expectations across demographic groups, documented in Bryan and Venkatu (2001a, b) and Souleles (2004). Furthermore, roughly 1/5 (q = 0.22) of the population round their inflation forecasts to the closest 0%, 5%, 10% or 15%. This result is consistent with the shape of the empirical distribution of individual level data. The standard deviation of idiosyncratic shocks, σ, is estimated to be 4.3.

According to specification (17), a typical person is able to remember past inflation figures correctly. We are, however, rather skeptical about this, as it seem more likely that individuals' ability to remember the contents of news articles decreases over time. To allow for this, we assume a very simple linear model for the error variance, related to recalling the newspaper figure. Specifically, the variance in (17) is parametrized as $\sigma^2(j) = \sigma^2 (3 (j - 1) + a)$ and, thus, increases by σ^2 every month. Furthermore, we can interpret $\sigma_0^2 \equiv \sigma^2 a$ as the variance in individuals' expectations not explained by the memory weakening process. With time-varying variance, model (17) can be written in the form

$$p\left(\pi^e_{t,t+4} \middle| j = k\right) = \frac{1}{\sqrt{2\pi}\sigma(j)} \exp\left\{ -\frac{1}{2}\left(\frac{\pi^e_{t,t+4} - N_{t-j+1}\left[\pi_{t-j,t-j+1}\right]}{\sigma(j)} \right)^2 \right\}. \tag{19}$$

The results in Table 2 lend strong support to the specification (19). In particular, the probability of the time-varying variance model being the true model is virtually one.[12] Furthermore, the results concerning the median value of λ and the heterogeneity of λ between agents are close to those obtained for model (17). This model can explain agents' overall heterogeneity fairly well, in the sense that the standard deviation of unexplained heterogeneity ($\sigma_0 \approx 1.6$) is small

---

9 The results of the regressions with shorter and longer lag lengths are quite similar.

10 We also estimated model (17) with the SPF series and found the estimate of the updating probability to be very high (≈ 1). The probabilities λ were also very homogenous (σ_λ was close to zero). When estimating a truncated version of Equation (2) based on Michigan and SPF mean expectations, an estimate of λ close to 1 was also obtained. These implausibly high estimates lend support to the naïve sticky information model.

11 The updating frequencies λ_i are simulated from the beta distribution (12), given the posterior means of a and β (the size of the population P = 1500).

12 The marginal likelihoods are estimated from the simulated posterior samples using the reciprocal importance estimator (see Gelfand and Dey, 1994) with a truncated multivariate normal importance density proposed by Geweke (1999).

relative to the high degree of heterogeneity observed in the actual data. According to our simulation experiment (not reported) this model can easily explain the long tails observed in individual level data. In Branch's (2007) RHE sticky information model, unexplained heterogeneity explains these long tails. In particular, the empirical standard deviation of Branch's (2007) sample was 12.7010. However, according to Branch (2007), the large empirical standard deviation is accounted for by a few outliers with expected inflation to be greater than 40%. Since his estimate for the standard deviation of unexplained heterogeneity was 6, we state that most variation in agents' expectations is in his model attributed to unexplained heterogeneity.

Since the forecasts based on standard econometric methods have not on average been better than the naïve forecast since 1984 (see Atkeson and Ohanian, 2001), we may expect that a fraction of individuals have used newspaper forecasts before that date. This might disturb the previous results. Therefore, the results should be checked using a post 1984 sample. Table 2 shows the estimation results of a sample from 1984/1 to 2001/4. The median estimate of λ is now close to the estimate of the corresponding parameter in the model of Section 2.4 (0.18), which is quite good news for the naïve sticky information model. The heterogeneity of λ between agents (σ_λ = 0.024), obtained for model (19), is lower than that obtained for the full sample, but markedly higher than that obtained for the constant variance model (σ_λ = 0.008). In this sample, the updating probability λ varies from 0.15 to 0.31 among agents according to our simulation experiment (not reported), indicating a moderate degree of heterogeneity. This evidence is in accordance with the 'implicit' evidence of Carroll (2006). Finally, the standard deviation of unobserved (or unexplained) heterogeneity between individuals ($\sigma_0 \approx 1.1$) is close to that obtained for the full sample and indicates that an agent-based naïve sticky information model does a fairly good job in capturing the heterogeneity in individuals' expectations.

5.4 Conclusion

Mankiw and Reis (2002) have proposed sticky information as an alternative to the sticky prices of Calvo (1983). Carroll (2003) has provided microfoundations for the aggregate inflation expectations equation of Mankiw and Reis (2002). The model presented in this paper can be interpreted as an extension of Carroll's (2003) model. We have proposed that agents, when forming their inflation expectations, adopt the past release of annualized monthly inflation with certain probability rather than the forward-looking newspaper forecast as suggested in Carroll (2003). The model is motivated by recent empirical work, which has cast doubt on the reliability of professionals' inflation forecasts, and, in general, on traditional approaches to inflation forecasting. We have shown that this simple model is able to fit the inflation expectations data of the

Michigan survey very well. In particular, the model can capture not only aggregate inflation expectations, but also the observed heterogeneity in households' expectations. The latter finding is based on a relatively simple adaptation, which we have derived in this study for Carroll's (2006) agent-based epidemiology model.

REFERENCES

Andres, J., Lopez-Salido, D. and Nelson, E., 2005. Stick-Price Models and the Natural Rate Hypothesis. *Journal of Monetary Economics* 52, 1025-1053.

Atkeson, A. and Ohanian, L.E., 2001. Are Phillips curves useful for forecasting inflation?. *Quarterly Review Federal Reserve Bank of Minneapolis* 25, 2-11.

Branch, W.A., 2004. The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *Economic Journal* 114, 592-621.

Branch, W.A., 2007. Sticky information and model uncertainty in survey data on inflation expectations. *Journal of Economic Dynamics and Control* 31, 245-276.

Brave, S. and Fisher, J.D., 2004. In search of a robust inflation forecast. *Economic Perspectives* 28, 12–31.

Brooks, S.P. and Gelman, A., 1997. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434-455.

Bryan, M. and Palmqvist, S., 2004. Testing near-rationality using survey data. Sveriges Riksbank Working Paper No. 183.

Bryan, M. and Venkatu, G., 2001a. The demographics of inflation opinion surveys. Economic Commentary, Federal Reserve Bank of Cleveland.

Bryan, M. and Venkatu, G., 2001b. The curiously different inflation perspectives of men and women. Economic Commentary, Federal Reserve Bank of Cleveland.

Calvo, G., 1983. Staggered Prices in a Utility Maximizing Framework. *Journal of Monetary Economics* 12, 383–98.

Carroll, C.D., 2003. Macroeconomic Expectations of Households and Professional Forecasters. *Quarterly Journal of Economics* 118, 269-298.

Carroll, C.D., 2006. The Epidemiology of Macroeconomic Expectations, in: Blume, L., Durlauf, S. (Eds.), The Economy as an Evolving Complex System, III, Oxford University Press.

Coibion, O., 2006. Inflation Inertia in Sticky Information Models. *Contributions In Macroeconomics* 6, Iss. 1, Article 1.

Coibion, O., 2007. Testing the Sticky Information Phillips Curve. Manuscript.

Doepke, J., Dovern, J., Fritsche, U. and Slacalek, J., 2008. The Dynamics of European Inflation Expectations. *The B.E. Journal of Macroeconomics* 8, Iss. 1 (Topics), Article 12.

Evans, G.W. and Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press, Princeton, NJ.

Fisher, J.D., Liu, C. and Zhou, R., 2002. When can we forecast inflation?. Economic Perspectives, Federal Reserve Bank of Chicago, First Quarter, 30–42.

Garratt, A., Koop, G., Mise, E. and Vahey, S.P., 2007. Real-time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty. Birkbeck Working Papers in Economics and Finance, No 714.

Gelfand, A. and Dey, D., 1994. Bayesian model choice: Asymptotic and exact calculations. *Journal of Royal Statistical Society* Ser. B 56, 501 – 514.

Gelman, A and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457 – 472.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2004. Bayesian Data Analysis 2nd edition. Chapman & Hall/CRC, Boca Raton.

Geweke, J., 1999. Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication. *Econometric Reviews* 18, 1-126.

Khan, H. and Zhu, Z., 2006. Estimates of the Sticky-Information Phillips Curve for the United States. *Journal of Money, Credit and Banking* 38, 195-207.

Kiley, M.T., 2007. A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting. *Journal of Money, Credit and Banking* 39, 101-125.

Mankiw, N.G. and Reis R., 2002. Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *Quarterly Journal of Economics* 117, 1295-1328.

Mankiw, N.G. and Reis R., 2006. Pervasive Stickiness. *American Economic Review Papers and Proceedings* 96, 164-169.

Mankiw, N.G. and Reis R., 2007. Sticky Information in General Equilibrium. *Journal of the European Economic Association* 5, 603-613.

Mankiw, N.G. and Reis, R., Wolfers, J., 2003. Disagreement about inflation expectations. in: Gertler, M., Rogoff, K. (Eds.). NBER Macroeconomics Annual 2003.

Reis, R., 2006a. Inattentive Consumers. *Journal of Monetary Economics* 53, 1761-1800.

Reis, R., 2006b. Inattentive Producers. *Review of Economic Studies* 73, 793–821.

Schwarz, G., 1978. Estimating the dimension of a model. *Annals of Statistics* 6, 461-464.

Shiller, R.J., 1997. Why Do People Dislike Inflation. in: Romer, C.D., Romer, D.H. (Eds.), Reducing Inflation: Motivation and Strategy, Chicago: University of Chicago Press.

Sims, C., 2002. The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity* 2, 1–62.

Sims, C.A., 2006. Rational Inattention: A Research Agenda. Manuscript.

Souleles, N.S., 2004. Expectations, Heterogeneous Forecast Errors, And Consumption: Micro Evidence Form The Michigan Consumer Sentiment Surveys. *Journal of Money, Credit and Banking* 36, 39-72.

Stock, J.H. and Watson, M., 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J.H. and Watson, M., 2007. Why Has U.S. Inflation Become Harder to Forecasts?. *Journal of Money, Credit and Banking* 39, 3–33.

Trabandt, M., 2007. Sticky Information vs. Sticky Prices: A Horse Race in a DSGE Framework. Kiel Working Papers No. 1369.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, Inc., New York.

6 ESSAY 5: AGGREGATE INFRASTRUCTURE CAPITAL STOCK AND LONG-RUN GROWTH: EVIDENCE FROM FINNISH DATA[1]

ABSTRACT

In this paper, the Bayesian structural vector autoregressive model and the Finnish aggregate infrastructure capital series from 1860 to 2003 are used to explore how government infrastructure policy affects long-run output growth. We use Finnish data, since to our knowledge the Finnish land and water construction investments series is the best available sufficiently long time series on *aggregate* infrastructure investments. We find strong and robust support in the Finnish data to indicate that permanent changes in government infrastructure policy can have permanent effects on the growth rate of output.

Keywords: aggregate infrastructure capital stock, Bayesian SVAR, long-run output growth

---

1 The first draft of this paper has been published in the University of Jyväskylä Working Paper Series No. 306 (2006).

6.1 Introduction

There is a growing body of research exploring how government policies influence long-run output growth[2]. Exogenous and endogenous growth theories give substantially different predictions on this relationship. The conventional wisdom of the exogenous neoclassical model has been that government actions can have an effect on the income level but only a temporal effect on the growth rate, while the endogenous growth theory predicts that permanent changes in government policy, e.g. investments in infrastructure capital or human capital, can also have a permanent effect on the growth rate of output.

It would appear that the relationship between public expenditure and output growth is relatively complex, suggesting that it is necessary to differentiate among the various components of governmental expenditure (see e.g. Fölster and Henrekson, 1999). For example, productive public expenditure in the areas of infrastructure and human capital are typically less than one fifth of the total public expenditure in OECD countries. In other words, more than 80 percent of public expenditure consists of expenditure which is claimed to have no positive growth effects. It is also important to note that public investment, which is generally used to construct the public capital stock, does not fully correspond to the concept of infrastructure. For example, telecommunications, electricity and water inputs used in the production process of nearly every sector are typically produced by private firms.

In this paper, we use the Finnish land and water construction investment series[3] from 1860 to 2003 to construct the long-run aggregate infrastructure capital series. This capital series covers all important factors of infrastructure investments in Finland during that period. To our knowledge, this is the first study on the dynamic effect of infrastructure to use a 'proper' approximation of aggregate infrastructure capital stock[4]. The choice of infrastructure series is, however, particularly important, since only about half of Finnish infrastructure investments were made by the public sector. This suggests that public capital stock may be a very poor guide to the amount of infrastructure capital

---

2 See for example: Jones (1995), Kocherlakota and Yi (1997), Canning and Pedroni (1999), Kneller et al. (1999), Karras (1999), Bleaney et al. (2001), Romero de Avila and Strauch (2003), Kamps (2008), Perotti (2005), and Fedderke et al. (2006). See also the survey by Romp and de Haan (2007) and references therein.

3 A full description of the elements in the land and water construction investments series can be found at http://www.stat.fi/tk/tt/luokitukset/toimiala_00_index.html. The source of the series is Statistics Finland.

4 The reader should note that since Aschauer (1989), a wide range of studies have addressed the question of the significance of the effect of infrastructural investment on aggregate output. This literature is reviewed in the World Bank's World Development Report (1994); see also the surveys by Boarnet (1995), Garcia-Mila et al. (1996), and Fernald (1999). However, these studies (to our knowledge) have used some subcomponent of aggregate infrastructure stock and/or public capital as an approximation for aggregate infrastructure stock.

produced. We decided to use Finnish data, since to our knowledge the Finnish land and water construction investments series is the best available sufficiently long time series on *aggregate* infrastructure investments. Furthermore, Finland had the highest growth rate in GDP per capita in Europe in the 20th century. During this period it also developed from a relatively backward agricultural society to a modern post-industrial state. It is thus important to learn something about this 'miraculous' growth process[5] (see further discussion on the special features of the Finnish economy in Jalava et al., 2006).

In the estimation, we used the Bayesian structural vector autoregressive (SVAR) model and identified it using a Barro (1990) type growth model. As opposed to the rest of the VAR studies on the dynamic effect of infrastructure (public) capital, which have based their analyses on impulse responses[6], we use the SVAR model to obtain a consistent parameter estimate for the long-run multiplier of output w.r.t. aggregate infrastructure capital.

Our analysis is close to that of Kocherlakota and Yi (1997); however, using the SVAR model rather than the Distributed-Lag approach gives us some advantages. Firstly, when we analyse a country where public infrastructure investments are typically less than 5% of total public expenditures (as is typical in OECD countries), we should not include a tax variable, but instead a private capital variable in the regression. If we estimate the model with infrastructure capital and the tax variable on the right-hand side, as is typical in a distributed-lag approach, only the revenue part of the government's budget is controlled. This biases the result towards exogenous growth. However, including a private capital variable allows us to control the effect of the agents' decisions on the allocation of investment between the infrastructure and other forms of capital. Thus, the advantage of the VAR approach is that all endogenous variables are explicitly included in the analysis. Secondly, unlike the distributed-lag approach, the VAR approach makes no specific assumptions as to the form of long-run relations between 'private' capital, infrastructure capital, and output. The assumption that there is a one-to-one relationship between infrastructure capital and output severely restricts the dynamics of the distributed-lag model. Thirdly, the VAR approach gives us a natural means of control for the influence of business cycles due to lagged output variables. Even if we can add a capacity utilisation rate to the distributed-lag regression to control for the influence of business cycles, there are problems in finding theoretical justification for this.

---

5 We are aware of certain specific problems, e.g. the effects of World War II and the slowdown years 1990-1993, in the regression analysis with long-run Finnish growth data. In this paper we use different sets of dummies to control the effects of unique shocks and find that the results with and without these dummies are fairly similar. This is not surprising since the use of long enough time series (100-140 years) should reduce the effect of unique shocks (and business cycles) in the regression. Finally, because there may be problems with using monetary values to calculate the infrastructure capital stock, we also run regressions using the kilometres of paved roads (see e.g. Pritchett, 1996, and Canning and Pedroni, 1999). We use the transport sector here in view of the reported importance of the highway infrastructure and the transport infrastructure in general (see e.g. Fernald, 1999).

6 See Kamps (2008) and Romp and de Haan (2007) for further discussion of VAR studies of public capital.

New empirical evidences presented in this paper are based on posterior analysis, since it allows us to draw an exact inference on the parameters with near non-stationary data[7]. When we estimate the SVAR model on the dynamic effect of infrastructure, we find strong support in the Finnish data for endogenous growth. However, failing to include a private capital variable in the regression biases the result towards exogenous growth.

The paper is organised as follows: Chapter 2 presents the Barro (1990) type growth model and the joint posterior p.d.f. of estimated parameters. Chapter 3 presents the data and results. Chapter 4 gives concluding remarks.

6.2 Estimated Model

Here we briefly review the simple stylised growth model initially developed by Barro (1990) and 'evolved' by Kocherlakota and Yi (1997) and Canning and Pedroni (1999), among others. In this model, aggregate production Y_t depends on aggregate infrastructure capital stock G_{t-1} accumulated through the end of period t-1, other forms of capital K_{t-1} accumulated through the end of period t-1, and the level of technology A_t. The log of the production function takes the form

$$y_t = \alpha \cdot k_{t-1} + \beta \cdot g_{t-1} + a_t , \tag{1}$$

where, from now on, lowercase letters refer to logs. As Kocherlakota and Yi (1997) argue, it is easy to modify the above technology to include labour, inelastically supplied, without changing any results derived from the underlying model. The production technology of Equation (1) is, however, fully appropriate for our purpose[8]. We assume that the log level of technology follows a stochastic exogenous process $a_t = \bar{a} + \rho \cdot a_{t-1} + \eta_t$, where the error term η_t can be broadly interpreted as i.i.d. technology shock. 'Private' capital and infrastructure capital accumulates according to the following technologies:

$$k_t = b + (1-\delta)k_{t-1} + \delta \cdot i_t \text{ and} \tag{2}$$

$$g_t = b_g + (1-\delta_g)g_{t-1} + \delta \cdot i_{gt} , \tag{3}$$

where δ and δ_g are the depreciation rates and I_t and I_{gt} are the aggregate investments on 'private' capital and infrastructure capital, respectively. As in Canning and Pedroni (1999), it is irrelevant for our purposes whether we think

7 See e.g. Sims and Uhlig (1991), Sims and Zha (1999), and Bauwens, Lubrano and Richard (1999: 136) for a discussion of posterior analysis of near non-stationary data.

8 Moreover, as our sensitive analysis shows, omitting labour from our empirical analysis has no influence on the estimation results.

of the decision on how much to invest in infrastructure as being made by the public sector and financed out of taxes, or whether it is being made by the private sector as a decision on the allocation of investment between different sectors. However, we assume for simplicity that the proportion of investment going to infrastructure investment ($0 < \tau_t < 1$) follows a stochastic exogenous process $\tau_t = \overline{\tau} + v_t$, where v_t is a zero mean stationary i.i.d. series which can be broadly interpreted as a shock to infrastructure investments.

The agent can use 'after-tax' output to consume or invest in 'private' capital, suggesting that

$$C_t + I_t = (1 - \tau_t)Y_t . \tag{4}$$

Finally, a representative agent seeks to maximise the following periodic utility function:

$$E_0 \sum_{t=1}^{\infty} \rho_c^t \ln(C_t) . \tag{5}$$

In what follows, the representative agent takes the processes generating $\{A_t, \tau_t, G_{t-1}\}$ as given and chooses C and I to maximise Equation (5), subject to constraints 1, 2, 4 and a requirement that capital be non-negative in each period. The derivations of the optimal path of 'private' capital can be found in Kocherlakota and Yi (1997; 253-255). They show that there exists the constant $S = a\rho_c\delta/(1 - \rho_c(1 - \delta))$ such that if

$$I_t = S(1 - \tau_t)Y_t \text{ and} \tag{6}$$

$$C_t = (1 - S)(1 - \tau_t)Y_t , \tag{7}$$

then the individual's intertemporal Euler equation and transversality condition are satisfied.

Equations (1)-(3) and (6)-(7) together with the exogenous processes of a_t and τ_t, the infrastructure investment relationship

$$I_{gt} = \tau_t Y_t , \tag{8}$$

and the initial values of K and G form the solution to the model. To characterise the dynamics of the economy we use (6) to write the capital accumulation equation (2) in the form

$$k_t = \overline{b} + (1 - \delta) \cdot k_{t-1} + \delta \cdot y_t + \delta \log(1 - \overline{\tau} - v_t), \tag{9}$$

$$\Leftrightarrow$$

$$k_t - y_t = \overline{b}\,\delta^{-1} - (1 - \delta)\delta^{-1}\Delta k_t + \log(1 - \overline{\tau} - v_t),$$

where Δ is the difference operator and $\bar{b} = b + \delta \cdot s$. Given (8), the infrastructure accumulation equation (3) follows

$$g_t = b_g + (1-\delta_g)g_{t-1} + \delta_g \cdot y_t + \delta_g \log(\bar{\tau} + v_t), \tag{10}$$
$$\Leftrightarrow$$
$$g_t - y_t = b_g \delta_g^{-1} - (1-\delta_g)\delta_g^{-1} \Delta g_t + \log(\bar{\tau} + v_t).$$

Substituting (9) and (10) in the production function equation (1), we have

$$y_t = \phi_0 + (\alpha + \beta)y_{t-1} + \alpha \log(1 - \bar{\tau} - v_{t-1}) + \beta \log(\bar{\tau} + v_{t-1}) + \phi_\alpha \Delta k_{t-1} + \phi_\beta \Delta g_{t-1} + a_t, \tag{11}$$

where $\phi_0 = \bar{\alpha}\bar{b}\,\delta^{-1} + \beta b_g \delta_g^{-1}$, $\phi_\alpha = -\alpha(1-\delta)\delta^{-1}$, and $\phi_\beta = -\beta(1-\delta_g)\delta_g^{-1}$.

We follow Canning and Pedroni (1999) and assume that one of the following two mechanisms determines the balanced growth path of our economy where all endogenous variables grow at the same constant growth rate, i.e. variables have a unit root and are co-integrated.

Firstly, let us suppose that $a + \beta < 1$ and assume that infrastructure capital and 'private' capital have a unit root. Then, as long as technology grows at a constant rate (i.e. $\rho = 1$), aggregate output will have a constant *exogenous* growth rate in the balanced growth path. This occurs, since all terms in Equation (11) are stationary except log output y_t and log productivity a_t. Moreover, because exogenous technological progress is the driving force of the economy, shocks to infrastructure investments have no long-run effects on the level of output. According to Equations (9) and (10), infrastructure capital and 'private' capital stocks grow at the same constant rate as output, i.e. variables have a unit root and are co-integrated, since Δg_t and Δk_t are stationary, as are the remaining terms in the relationships. Finally, given Equations (6)-(8), consumption and investments also grow at this constant rate.

Let now us suppose that $a + \beta = 1$ and assume that infrastructure capital and 'private' capital have a unit root. Then, under stationary technological progress ($\rho < 1$), output has a constant *endogenous* growth rate in the balanced growth path. That is, according to Equation (11), permanent changes in the proportion of investment going to infrastructure (τ) can have a permanent effect on growth rates of output. The sign of this effect may be positive or negative, depending on the level of the proportion of investment going to infrastructure. Finally, according to Equations (6)-(10), all endogenous variables in the economy grow at the same constant rate as output.

Thus, in our economy, the key difference between endogenous and exogenous growth lies in the values of a, β, and ρ. We may estimate Equation (11) directly using nonlinear regression. Estimates of a, β and ρ, and the analysis based on these parameters, are consistent provided that the model is correctly specified. However, since this approach may ignore some relevant long-run information, we prefer to use a more flexible modelling strategy where long-run relationships of key endogenous variables in the model are estimated without restrictions.

To be more concrete, let us substitute $a_t = \bar{a} + \rho \cdot a_{t-1} + \eta_t$ into Equation (1) and define a vector of endogenous variables as $x_t = (y_t\ k_t\ g_t)'$. We can then easily write the system of equations (1)-(3) and (6)-(8) in the following matrix form:

$$\Gamma_0 x_t = \gamma + \Gamma_1 \cdot x_{t-1} + \Gamma_2 \cdot x_{t-2} + e_t, \tag{12}$$

where the vector of stationary i.i.d. shocks e_t, parameter vector γ, and parameter matrices Γ_0, Γ_1, and Γ_2 can be written as

$$\Gamma_0 = \begin{pmatrix} 1 & 0 & 0 \\ -\delta & 1 & 0 \\ -\delta_g & 0 & 1 \end{pmatrix}, \Gamma_1 = \begin{pmatrix} \rho & \alpha & \beta \\ 0 & 1-\delta & 0 \\ 0 & 0 & 1-\delta_g \end{pmatrix}, \Gamma_2 = \begin{pmatrix} 0 & -\rho\alpha & -\rho\beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \gamma = \begin{pmatrix} \bar{a} \\ \bar{b} \\ b_g \end{pmatrix},$$

$$e_t = \begin{pmatrix} \eta_t \\ \delta(1-\bar{\tau}-v_t) \\ \delta_g(\bar{\tau}+v_t) \end{pmatrix}.$$

Our empirical analysis is based on the above model. We analyse the SVAR model of the following form

$$A_0 x_t = a + \sum_{i=1}^{p} A_i x_{t-i} + \varepsilon_t\ . \tag{13}$$

Equation (13) is of the same form as Equation (12), except that we assume the vector of structural errors ε_t to be normally distributed with zero mean and Λ diagonal covariance matrix, matrices A_i ($i = 0,\dots,p$) to be estimated without any restrictions, and the data to be used to estimate proper lag length (p) for the model. According to Equation (12), the contemporaneous relation matrix A_0 follows

$$A_0 = \begin{pmatrix} 1 & 0 & 0 \\ a_{0ky} & 1 & 0 \\ a_{0gy} & 0 & 1 \end{pmatrix}, \tag{14}$$

suggesting that A_0 is a non-singular matrix such that the model provides a complete description of the p.d.f. for the data conditional on the initial observations.

To proceed, we reparametrized the reduced-form version of the SVAR model in Equation (13) as follows:

$$\Delta x_t = \pi + \sum_{i=1}^{p-1} \Pi_i \Delta x_{t-i} + \Pi x_{t-1} + u_t \ , \tag{15}$$

where the reduced-form errors are normally distributed with zero mean and

$$\Omega = \left(A'_0 \, \Lambda^{-1} A_0\right)^{-1} \tag{16}$$

covariance matrix. Parameter matrices Π and Π_i capture the long-run and short-run effects, respectively. Even if the unconstrained least squares estimates of Π and Π_i are consistent, and we can draw an exact posterior inference on the parameters when some of the series are non-stationary and/or co-integrated[9], the major problem in the analysis of (15) lies in interpreting the long-run coefficient in the Π matrix. Specifically, the element π_{yg} in this Π matrix is the long-run multiplier of output with respect to aggregate infrastructure capital (g_t) given that only g_t moves contemporaneously following a shock to g_t. This identifying assumption is fulfilled by the simple theoretical model above. That is, given the theory-based restriction on the contemporaneous relation, matrix A_0, π_{yg} is a long-run multiplier in Equation (15).

We may alternatively parametrise Π as $\Pi = a\beta'$, where a and β are ($m\square r$) matrices of rank r, to incorporate rank restrictions in the model of Equation (15) (see e.g. Bauwens and Lubrano, 1996). The columns of β correspond to the co-integrating vectors, and the rows of a to their adjustment coefficient (weights). Given that there exists a positive long-run relationship between log output and log infrastructure capital, we may basically test whether innovations to infrastructure have a long-run effect on the level of output using the estimates of the elements of a (see e.g. Canning and Pedroni, 1999). The problem in the analysis lies in the possible pseudoidentification of the co-integrating coefficient (see e.g. Strachan and Inder, 2004, and Villani, 2005 and 2006). To be more concrete, it is only the cointegration space that is identified, not particular cointegrating vectors. Moreover, a contains reduced-form coefficients, which implies that we cannot fully trust the analysis based on them.

Thus, to keep our analysis as simple as possible, we rewrite Equation (15) in the following condensed matrix form

$$X = ZB + U \ , \tag{17}$$

where the *tth* rows of the matrices $X(T\times m)$, $Z(T\times k)$ and $U(T\times m)$ are given by $\Delta x'_t$, $(1 \ \Delta x'_{t-1} \cdots \Delta x'_{t-p+1} \ x'_{t-1})$, and u'_t, respectively. Note also that matrix B is obtained by stacking π', Π_i':s, and Π' ($k = 1 + m\times p$).

Let the subscript i denote the *ith* column vector. We can then write the equation for variable i in the multivariate regression model (17) as $x_i = Z\beta_i + u_i$ (see e.g. Kadiyala and Karlsson, 1997). Given the contemporaneous restrictions

---

9 See e.g. Luetkepohl (1991), Sims et al. (1990), Sims and Uhlig (1991), Sims and Zha (1999), and Bauwens et al. (1999: 158-196).

in matrix A_0, the reduced-form equation for output (i = 1) in Equation (17) is (after slight reparametrisation) equal to the structural-form equation for output. This suggests that we can write the likelihood function for a reparametrised version of the structural form equation for output as

$$L\left(x_y \middle| \beta_y, \lambda_y, Z\right) \propto \lambda_y^{-0.5T} \exp\left\{-0.5\lambda_y^{-1}\left(x_y - Z\beta_y\right)'\left(x_y - Z\beta_y\right)\right\}, \tag{18}$$

where we let subscript y denote the first column vector, and the last row of vector β_y is the long-run multiplier of output w.r.t. infrastructure capital (π_{yg}). With the standard Jeffreys' prior (Zellner, 1971)

$$p(B, \Lambda) \propto \lambda_y^{-0.5}, \tag{19}$$

the joint posterior for parameters is obtained as

$$q\left(\beta_y, \lambda_y \middle| Z, x_y\right) \propto \lambda_y^{-0.5(T+1)} \exp\left\{-0.5\lambda_y^{-1}\left(x_y - Z\beta_y\right)'\left(x_y - Z\beta_y\right)\right\}. \tag{20}$$

Integrating λ_y out of the joint posterior (20), the marginal posterior of β_y is a k-variate Student-t distribution with location parameter vector

$$\hat{\beta}_y = (Z'Z)^{-1} Z' x_y, \tag{21}$$

and scale matrix

$$\frac{\left(x_y - Z\hat{\beta}_y\right)'\left(x_y - Z\hat{\beta}_y\right)}{T-k}(Z'Z)^{-1}. \tag{22}$$

This enables us to make posterior inferences regarding the long-run multiplier of output w.r.t. aggregate infrastructure capital (π_{yg}) using standard t-tables (see e.g. Zellner, 1971).

6.3 Data and Estimation Results

In this section we first describe the data, then continue our analysis of the dynamic effect of aggregate infrastructure capital using the distributed-lag approach. We show that the distributed-lag approach biases the results in favour of exogenous growth. Finally, we present the SVAR results, which correct this bias.

Our data source is Statistics Finland's historical series of Finnish national accounts. The base year of the non-residential gross domestic product (GDP) and investment series is 2000. Using data on gross land and water construction investments from 1860-2003, we use the perpetual-inventory method with a depreciation rate of 1.98 per cent per year (Fernald, 1999, and Boskin et al., 1989, use a similar depreciation rate in their roads stock estimates) to estimate constant-euro values of the infrastructure stock for each year. Like Fernald (1999), we assume that the infrastructure input in a given year depends on the stock of infrastructure at the beginning of the year[10].

TABLE 1 Augmented Dickey-Fuller Tests and FML-based rank estimates

Variables ln GDP = y, ln K = k, and ln G = g

Sample period	ln Variable	t-adf	FML-based rank estimates (r = rank) between *y* and *g* (*k*)	FML-based rank estimates (r = rank) between *y*, *k*, and *g*
1860-2003	GDP	-2.1485	-	-
1860-2003	'Private' Capital	-1.8179	r = 1	-
1860-2003	Infrastructure Capital	-2.0426	r = 1	r = 0
1948-2003	Paved Roads	-1.0148	r = 2	r = 1

Alternative hypothesis is stationary. Trend is included in all regressions.

In 2000, the value of estimated infrastructure capital stock was about 70 billion euro. The shares of traffic investments and the share of traffic, energy and water investments were 63% (53% if telecommunications is excluded) and 81% of total land and water construction investments on average over the period 1960-2003, respectively. Additionally, public investments are about half the total land and water construction investments over the given period (all reported shares remain relatively constant over time).

Figure 1 shows volume indices of gross domestic product (GDP) and the land and water investment series over the period 1860-2003. We can see from the figure that the marked industrialisation and urbanisation of the Finnish economy in the post-war period required (and made possible) high investments in infrastructure. Construction of infrastructure capital stock was particularly rapid during the period 1946-1970, and slowed down thereafter.

To see whether the data confirm the existence of a co-integrating relationship between log infrastructure capital, log physical capital and log GDP series, we follow Corander and Villani (2004) and compute approximate fractional marginal likelihoods (FML) from Equation (15) with different lags, ranks and sets of series ($x_t = (y_t\ k_t)'$, $x_t = (y_t\ g_t)'$, and $x_t = (y_t\ k_t\ g_t)'$). Here we use classical augmented Dickey-Fuller tests (the ADF test) in the preliminary data analysis, since this is readily available and does not require extra programming effort.

---

10 We estimate non-infrastructure capital stock using a similar method, but using 5% depreciation per year. The capital stocks in 1860 are estimated using I(1860)/(g+d) where I denotes the investments, g the growth rate of GDP (calculated as the average growth rate from 1860 to 1870), and d the depreciation rate.

Table 1 reports estimated ranks and the results from the ADF tests. According to the table, all the series are form I(1) processes. We found some support in the data for long-run relationships between the infrastructure capital and GDP series and between the non-infrastructure capital and GDP series. However, estimated relationships are relatively weak, since they seem to vanish when both k_t and g_t variables are included in the regression.

6.3.1 Distributed-Lag Results

In our first set of regressions we follow Kocherlakota and Yi (1997), Kneller, Bleaney and Gemmell (1999), Karras (1999), Bleaney et al. (2001), and Romero de Avila and Strauch (2003), among others, and use the distributed-lag approach to analyse the relationship between infrastructure capital and output growth[11]. Note that this approach is consistent with our theoretical model above, which predicts a one-to-one long-run relationship between log aggregate infrastructure and log output.

Table 2 contains the results of regressions with different time periods, which employ eight lags of explanatory variables. As will be seen form the table, estimated parameters are statistically insignificant, favouring exogenous growth. These empirical estimates are robust, as compared to those in previous studies. The results simply indicate that one should include the marginal tax variable in the above distributed-lag regression to control both sides of the governmental budget.

Table 3 displays our results for distributed-lag regressions based on the infrastructure capital and the income tax series[12]. Looking at the results of the first row, we see that the long-run coefficients of the aggregate infrastructure capital and income tax series are close to zero and statistically insignificant. Regression with the paved roads series and the income tax series also gives similar results. These results contradict previous literature on the effect of public capital. This is not surprising, since public infrastructure investment was less than 7% of total public expenditure on average over the period 1948-2003. This indicates that only the revenue part of the government's budget is controlled in the above regression.

---

11 We will use the standard frequentist approach in order to maintain comparability with previous literature.

12 The proper income tax series is available for the period 1948-2003, and can be found in Turkkila (2006).

TABLE 2 The Estimation Results of Distributed-Lag Models

The estimation results of model

$$\Delta y_t = \lambda \cdot D_t + \sum_{i=1}^{p-1} \gamma_i \cdot \Delta(g_{t-i} - y_{t-i}) + \alpha \cdot (g_{t-1} - y_{t-1}) + \varepsilon_t$$

The regression was run with eight lags (k = 8). The Newey-West (1987) standard errors (with 5 lags) are shown in parentheses. The results of the regressions with shorter and longer lag lengths are similar. D_t includes a constant. Results are not sensitive to the different sets of dummy variables. With dummies, we control the influence of World War I, the slowdown years 1917-1920, World War II, and the slowdown years 1990-1993.

Sample period	variable (g)	α
1860-2003	Infrastructure Capital	0.013 (0.014)
1860-2003	Public Expenditure	0.013 (0.015)
1900-2003	Infrastructure Capital	0.016 (0.029)
1900-2003	Public Expenditure	0.015 (0.030)
1948-2003	Infrastructure Capital	-0.011 (0.030)
1948-2003	Public Expenditure	-0.104 (0.038)
1948-2003	Paved Roads	-0.007 (0.009)

To obtain a closer view of this, we regress GDP growth with the income tax and public expenditure series using the above distributed-lag approach. Estimation results are given in the third row of Table 3 (see also Table 2). The statistically and economically significant long-run coefficients are now consistent with the results of previous studies. The value of α is 0.15, suggesting that a one per cent permanent increase in public expenditure share will permanently boost long-run growth by 0.15 percentage points. Thus, when controlling both sides of the governmental budget, we find that permanent changes in government policy have effects on the growth rate of output. This gives some support for endogenous growth.

According to the results shown in Tables 2 and 3, we conclude that since agents must make decisions on the allocation of the investments of infrastructure capital and other forms of capital, we should not control income taxation, but rather non-infrastructure investments when exploring the relationship between aggregate infrastructure capital and output. This suggests that the above results may be misleading due to omitted-variable bias.

TABLE 3 The Estimation Results of Distributed-Lag Models with Revenue and Expenditure Variable

The estimation results of model

$$\Delta y_t = \lambda \cdot D_t + \sum_{i=1}^{p-1} \gamma_i \cdot \Delta(g_{t-i} - y_{t-i}) + \sum_{i=1}^{p-1} \delta_i \cdot \Delta \ln(1-\tau_{t-i})$$
$$+ \alpha \cdot (g_{t-1} - y_{t-1}) + \beta \cdot \ln(1-\tau_{t-1}) + \varepsilon_t$$

where D_t includes a constant and dummy variable, which controls the slowdown years 1990-1993. The trend is statistically insignificant. The regression was run with eight lags (k = 8). The results of the regressions with shorter and longer lag lengths are not similar. The Newey-West (1987) standard errors (with 5 lags) are shown in parentheses. Finally, τ_t is the income tax variable.

Sample period	Variable (g)	α	β
1948-2003	Infrastructure Capital	-0.011	-0.022
		(0.062)	(0.084)
1948-2003	Paved Roads	-0.001	-0.065
		(0.010)	(0.161)
1948-2003	Public Expenditure	0.154**	0.313**
		(0.058)	(0.071)

** p-value of one-sided hypothesis test < 0.01

6.3.2 SVAR Results

In the above distributed-lag approach, the OLS estimates of the long-run parameters are consistent only if the model is correctly specified. We, however, found several reasons for misspecification. Firstly, as our empirical result indicates, we should not include a tax variable, but non-infrastructure capital in the regression. Secondly, the distributed-lag approach assumes that there is a one-to-one long-run relationship between infrastructure capital and the output variable, which severely restricts the dynamics of the underlying model. Thirdly, the above distributed-lag approach fails to control the influence of the business cycles.

In order to avoid these problems, we report the point estimates (means) and standard deviations of the long-run multiplier of output growth w.r.t. aggregate infrastructure capital (π_{yg}). The estimation results shown are based on Equations (21) and (22). We estimate proper lag lengths for the models using the estimated average discrepancy[13] reported in Table 5 (Appendix).

13 The estimated average discrepancy approximates expected deviance. In the limits of a large sample size, the model with the lowest expected deviance will have the highest posterior probability. We prefer using the discrepancy between the data and the model to using Bayes factors in model comparisons. We consider Bayes factors to be irrelevant in most cases, since they are used to compute relative probabilities of the models conditional on one of them being true; see Gelman et al (2004). The estimated average discrepancy is defined as

$$\hat{D}_{avg}(y) = \frac{1}{N} \sum_{i=1}^{N} D(y, \theta^i),$$

where y is the data, $D(y,\theta) = -2 \log l(y;\theta)$ is the 'deviance', and $\{\theta^i\}_{i=1}^{N}$ a sample

Table 4 shows the results for three time periods. In addition to the entire sample 1860-2003, models are estimated for the 1900 – 2003 and 1948 – 2003 (post World War II period) subsample periods. This serves as a check of robustness and parameter constancy. Looking at the results in Table 4, we see that aggregate infrastructure capital has positive effects on economic growth with a probability of over 95 per cent. Thus, the data give strong support for endogenous growth in Finland during the past century. The estimation results based on the paved roads series confirm such a conception. The point estimate of the long-run multiplier of output w.r.t. aggregate infrastructure capital (π_{yg} = 0.10) suggests that a one per cent permanent increase in the aggregate infrastructure capital stock raises GDP growth permanently by 0.10 percentage points. As we see in Table 4, estimated relations remain relatively robust over time. The results of the regression (to save space not reported here) with shorter and longer lag lengths are similar. Moreover, the results shown are not sensitive to the use of different sets of dummies. It is thus fair to say that in Finland, governmental actions have had an *economically significant* effect on output growth during the past 140 years.

Table 4 also reports the point estimates and standard errors for the long-run multiplier of 'private' capital w.r.t. infrastructure capital (π_{kg})[14]. According to the table, there is strong support in the data for the long-run multiplier (π_{kg}) being positive. Results indicate that 'private' capital and infrastructure capital are complements in the long run. We may expect that there are two opposite forces determining the long-run effect of infrastructure capital on 'private' capital (see e.g. Baxter and King, 1993, and Kamps, 2008). The first concerns the allocation of investments in infrastructure capital and other forms of capital. This cost reduces the recourses available to the private sector, since higher levels of investment in infrastructure are obtained at the cost of lowered investments in other forms of capital. The second is the positive effect of the infrastructure on the marginal product of other forms of capital. As a consequence, higher levels of infrastructure investment should induce a rise in private investment. According to the results shown in Table 4, the latter effect has dominated in Finland over the past 140 years. This finding is in line with that in Kamps's (2008) VAR study of public capital. He, however, found only moderate evidence for capital being complementary in Finland. The major reason for the difference between his and our results could be that we use an aggregate infrastructure capital series rather than a public capital series. Furthermore, Kamps' time series covers a considerably shorter time period (from 1960 to 2001) than our data.

---

from the posterior (N is the sample size). In the limit, as the sample size tends to infinity, the model with the lowest expected deviance will have the highest posterior probability. The DIC is defined as $DIC = 2\hat{D}_{avg}(y) - D_{\hat{\theta}}(y)$, where $D_{\hat{\theta}}(y) = D(y, \hat{\theta}(y))$ and $\hat{\theta}$ is the parameter mean.

14 Estimates shown are reduced-form estimates, so that the reader should be careful in the interpretation of these results.

TABLE 4 The means and standard deviations of the Parameter's Posteriors of the VAR Model

Model

$$\begin{pmatrix} \Delta y_t \\ \Delta k_t \\ \Delta g_t \end{pmatrix} = \psi D_t + \sum_{j=1}^{k-1} \Pi_j \begin{pmatrix} \Delta y_{t-j} \\ \Delta k_{t-j} \\ \Delta g_{t-j} \end{pmatrix} + \begin{pmatrix} \pi_{yy} & \pi_{yk} & \pi_{yg} \\ \pi_{ky} & \pi_{kk} & \pi_{kg} \\ \pi_{gy} & \pi_{gk} & \pi_{gg} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ k_{t-1} \\ g_{t-1} \end{pmatrix}$$

In samples 1860-2003 and 1948-2003, the data do not support the use of the trend variable (results are similar with or without trend); thus in these cases, ψ includes a constant. In 20th regressions, ψ includes a constant and trend (results are similar with or without it). The results are not sensitive to using different dummies and/or different ancillary variables. The results of the regressions with shorter and longer lag lengths are similar.

Sample Period	Variable (g)	π_{yg}
1860-2003	Infrastructure Capital	0.096* (0.045)
1900-2003	Infrastructure Capital	0.127* (0.057)
1948-2003	Infrastructure Capital	0.141* (0.066)
1948-2003	Kilometres of Paved Roads	0.017* (0.009)
Sample Period	**Variable (g)**	π_{kg}
1860-2003	Infrastructure Capital	0.020* (0.010)
1900-2003	Infrastructure Capital	0.032* (0.015)
1948-2003	Infrastructure Capital	0.022* (0.013)
1948-2003	Kilometres of Paved Roads	0.003* (0.002)

* parameter > 0 at 95% probability.

In sum, our estimation results shown in Table 4 are robust, as compared to those of previous studies on the dynamic effects of public (infrastructure) capital (see e.g. Kocherlakota and Yi, 1997, Kneller, et al., 1999, Karras, 1999, Canning and Pedroni, 1999, Bleaney et al., 2001, Romero de Avila and Strauch, 2003, Kamps, 2008, and Fedderke et al., 2006). However, as we have shown, when the dynamic effect on aggregate infrastructure capital (not public capital) are studied, we should put effort into modelling to ensure consistent parameter estimates for the long-run coefficient of output w.r.t. infrastructure capital.

6.3.3 Robustness of the Results

To check the robustness of the VAR results shown in Section 3.2, two types of sensitive analysis are needed.

a) Most studies on the dynamic effect of public (infrastructure) capital have used four-variables VAR models (public capital (infrastructure capital), 'private' capital, employment (e_t), and output) in their analyses. Panel (A) in Table 6 (Appendix) reports the results of our four-variable VAR model. As will be seen from the table, there is strong support in the data for the long-run multiplier of output w.r.t. infrastructure capital being positive. The point estimates of the parameters are higher than those shown in Section 3.2.; however, our analysis is not critical for this.

   One may, of course, prefer to include the tax variable in the analysis. We therefore estimated the VAR model with five variables: infrastructure capital, 'private' capital, employment, income tax rate, and output. Results for the long-run multipliers are shown in Panel (B) in Table 6. They show that adding the income tax variables in the regression has no influence on the long-run results. In addition, including both a public expenditure variable and an income tax variable in the analysis also does not alter the results (results not given here).

b) In addition to checking ancillary variables, we also examine the sensitivity of our results using an alternative identification of the VAR models. As has been widely reported in the VAR literature, estimation results may be sensitive to the choice of identification. Specifically, we estimate impulse responses of GDP (and private capital) to a shock to aggregate infrastructure capital using two precisely identified VAR models with alternative orderings of variables. In addition, we estimate Equation (15) with a diagonal covariance matrix structure. In the first model, we use a recursive approach with the following ordering of variables: x_t = (y_t k_t e_t g_t). This ordering corresponds to the VAR model shown in Section 2. In the second model, we follow Kamps (2008), among others, in using the following ordering of variables: x_t = (g_t k_t e_t y_t). In both analyses, A_0 is restricted to be triangular and we solve for $A_0^{-1}\Lambda^{0.5}$ by taking a Choleski decomposition of Ω (see Equation, 15). In the third model, A_0 is the identity matrix. This identification implies that Equation (15) is a structural form model allowing direct interpretation for all long-run parameters.

   Figures 2-5 show the impulse responses of GDP (*yg*) (and private capital, *kg*) to a shock to infrastructure capital for a horizon of 25 years achieved by three alternative identification schemes. We find positive responses of GDP (and private capital). In general, impulse responses are very similar, suggesting that we can trust the results given in Section 3.2. This is in stark contrast to Kamps (2008), who finds highly sensitive impulse responses in the Finnish data. The major reason for the difference between his and our results could be that we use an aggregate infrastructure capital series rather than a public capital series. Furthermore, Kamps' time series covers a considerably shorter time period (from 1960 to 2001) than our data.

Figures 2-5 also show the impulse responses of infrastructure capital to a shock to GDP (*gy*) (and private capital, *gk*) for a horizon of 25 years. The results indicate that impulse responses of infrastructure capital are sensitive to the ordering of variables. Note, however, that the VAR results given in Section 3.2 are based on an equation for output, not an equation for infrastructure. Thus, the results are not critical for this.
Finally, according to the impulse responses shown, we find that the data give moderate support for reverse causation, i.e. feedback effects from the output to infrastructural investments (at least in the short-run).

In sum, the main results of this study are robust to the given ancillary variables and the identification of VAR models.

6.4 Conclusion

In this paper, we have explored how investment in infrastructure capital influenced long-run output growth over the long-run period 1860-2003. Our analysis is based on the Bayesian SVAR model, since this gives reliable parameter estimates for the long-run multiplier of output with respect to aggregate infrastructure capital.

Based on the estimation results presented here, we find that investment in infrastructure capital has had a positive effect on output growth over the long run. The results shown are robust, suggesting that the Finnish data are consistent with endogenous growth. Our results also indicate that, when output growth is regressed with aggregate infrastructure capital stock (rather than public capital stock), the parameter estimates of the standard Distributed-Lag approach are inconsistent. The reason for this bias is the lack of a proper control variable. As was explained above, it seems appropriate to include 'private' capital rather than distortionary taxes in the analysis when the effect of infrastructure capital on output growth is explored.

Finally, based on our impulse response analysis, we find moderate evidence for a feedback effect from the output to infrastructural investments, suggesting that it is important to treat all model variables as endogenous.

FIGURES

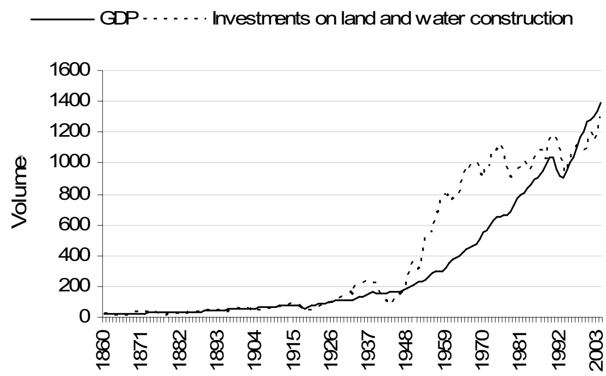

FIGURE 1 Real GDP and infrastructure investments series

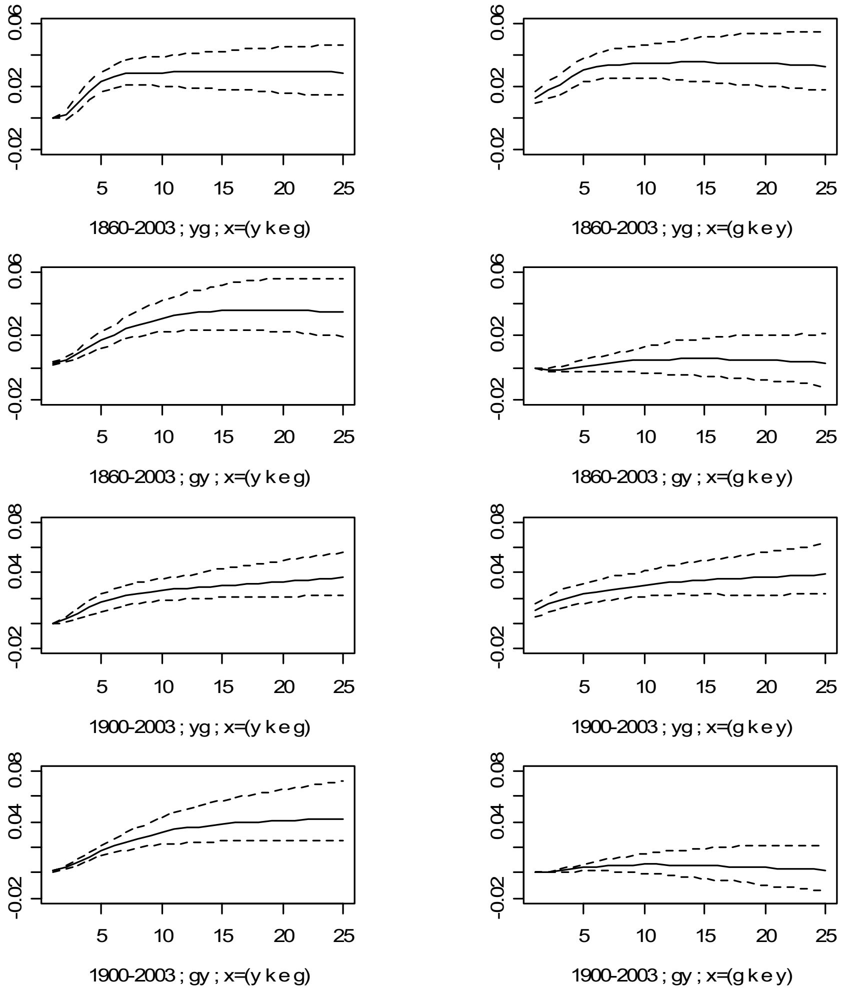

FIGURE 2 Impulse responses of GDP to a shock to infrastructure capital, and impulse responses of infrastructure capital to a shock to GDP. Identification of the models is based on a recursive approach with a different ordering of variables. The time horizon of the impulse responses is 25 years. ------ represents 68% credible intervals, and —— median.

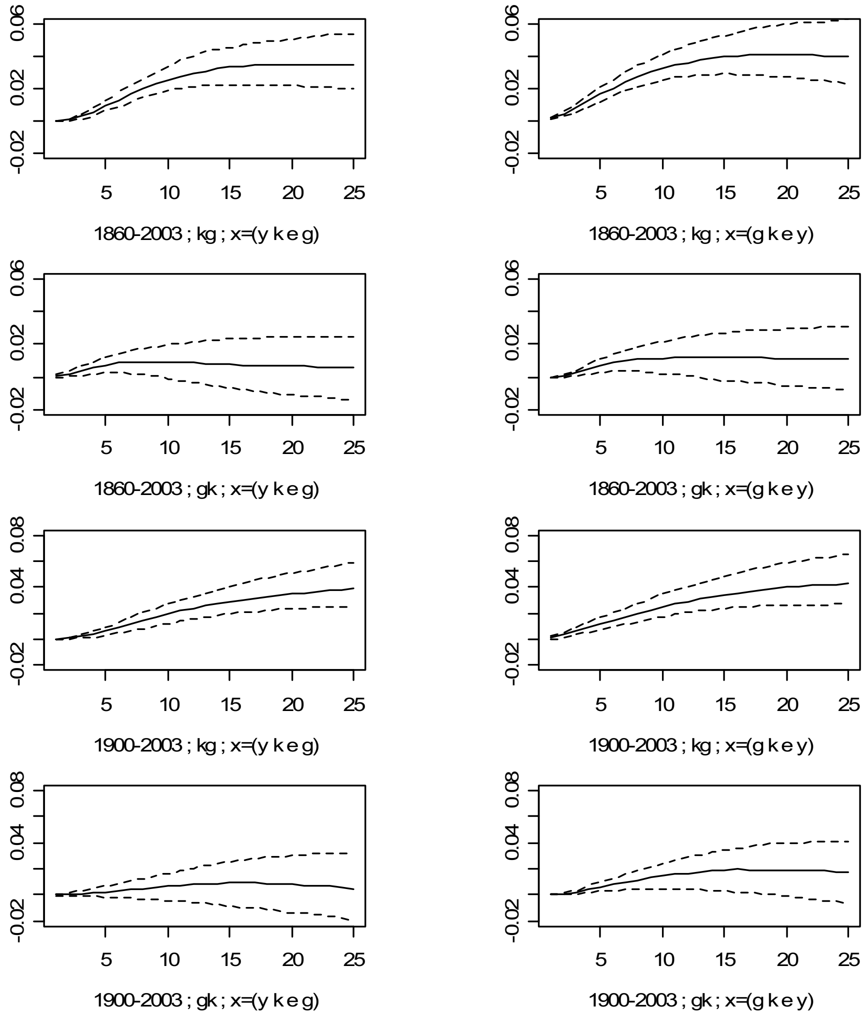

FIGURE 3 Impulse responses of private capital to a shock to infrastructure capital and the impulse responses of infrastructure capital to a shock to private capital. Identification of the models is based on a recursive approach with a different ordering of variables. The time horizon of the impulse responses is 25 years. ------ represents 68% credible intervals, and —— median.

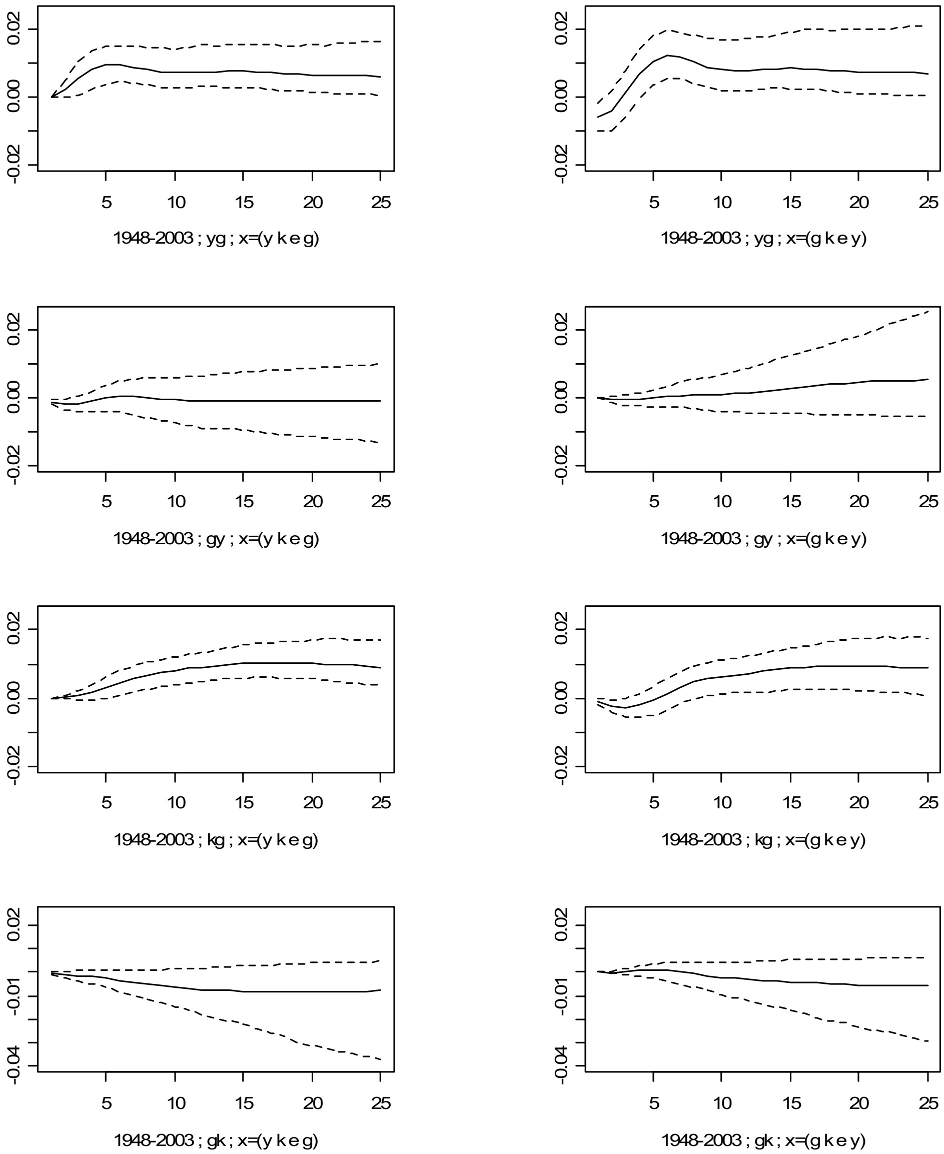

FIGURE 4 Impulse responses of GDP (private capital) to a shock to infrastructure capital, and impulse responses of infrastructure capital to a shock to GDP (private capital). Identification of the models is based on a recursive approach with a different ordering of variables. The time horizon of the impulse responses is 25 years. ------ represents 68% credible intervals, and —— median.

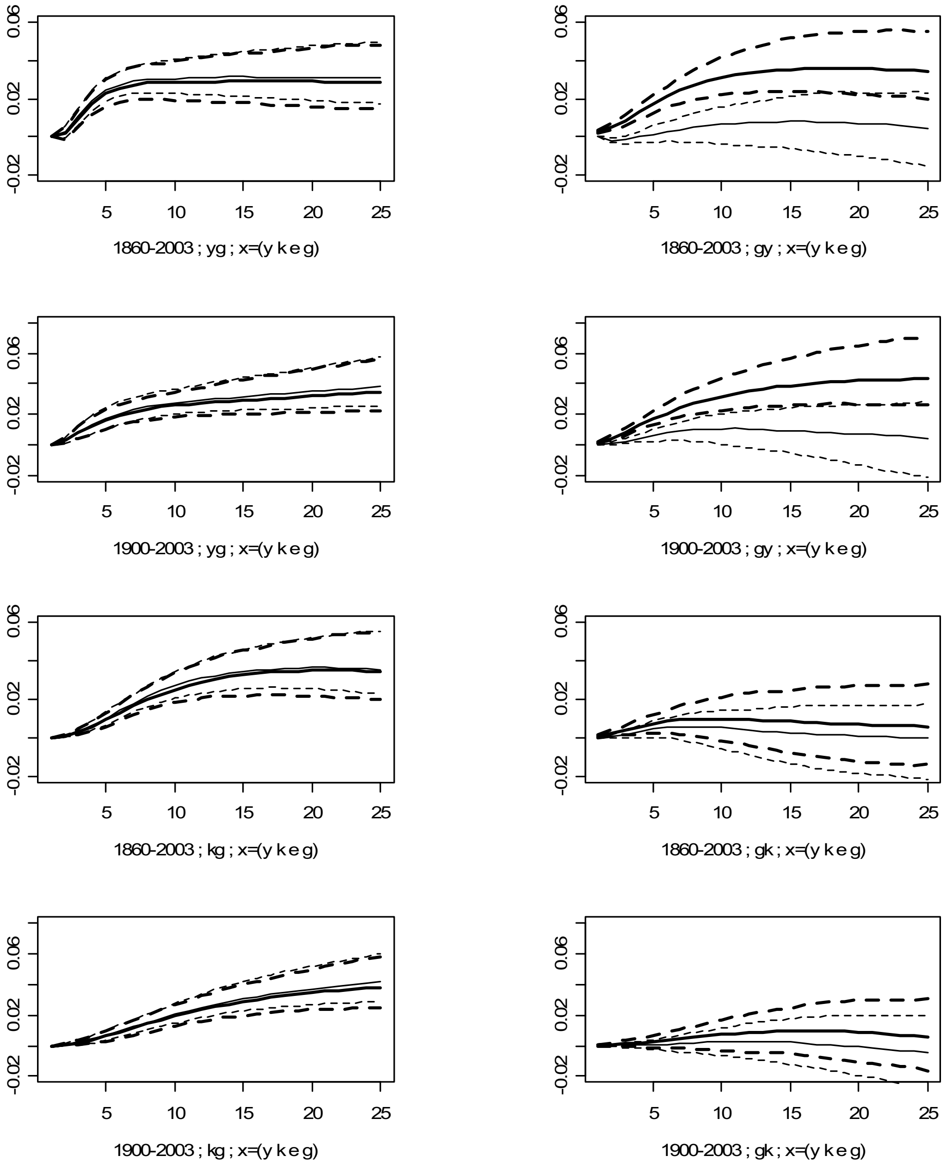

FIGURE 5 Impulse responses of GDP (private capital) to a shock to infrastructure capital, and impulse responses of infrastructure capital to a shock to GDP (private capital). Identification of the models is based on a recursive and diagonal approach. The time horizon of the impulse responses is 25 years. ------ represents 68% credible intervals, and —— median. The thinner lines represent the impulse responses of VAR with diagonal restriction, and the fatter lines the impulse responses of VAR with lower triangular restriction.

REFERENCES

Aschauer, D.A., 1989. Is Public Expenditure Productive?. *Journal of Monetary Economics* 23, 177-200.

Barro, R.J., 1990). Government Spending in a Simple Model of Endogenous Growth. *Journal of Political Economy* 98, 103-125.

Bauwens, L. and Lubrano, M. (1996), Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR System, *Advances in Econometrics* 11, 29-49.

Bauwens, L. Lubrano, M. and Richard, J.-F., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University Press, Oxford.

Baxter, M. and King, R.G., 1993. Fiscal Policy in General Equilibrium, *American Economic Review* 83, 315-333.

Bleaney, M., Gemmell, N. and Kneller, R., 2001. Testing the Endogenous Growth Model: Public Expenditure, Taxation, and Growth over the Long-Run. *Canadian Journal of Economics* 34, 36-57.

Boarnet, M.G., 1995. Highways and Economic Productivity: Interpreting Recent Evidence, University of California School of Social Ecology Working Paper No 291.

Boskin, M.J. Robinson, M.S. and Huber, A.M., 1989. Government Saving, Capital Formation and Wealth in the United States, 1947-85. in R.E. Lipsey and H.S. Tice, eds., The measurement of savings, investment, and wealth. Chicago: University of Chicago Press, 287-353

Canning D. and Pedroni P., 1999. Infrastructure and Long Run Economic Growth. Center for Analytical Economics Working Paper No. 99-09, Cornell University.

Corander, J. and Villani, M., 2004. Bayesian Assessment of Dimensionality in Reduced Rank Regression. *Statistica Neerlandica* 58, 255-270.

Fedderke, J.W., Perkins, P. and Luiz, J.M., 2006. Infrastructural Investment in Long-run Economic Growth: South Africa 1875-2001. *World Development* 34, 1037-1059.

Fernald, J.G., 1999. Roads to Prosperity? Assessing the Link Between Public Capital and Productivity. *The American Economic Review* 85, 619-638.

Fölster, S. and Henrekson, M., 1999. Growth and the Public Sector: A Critique of the Critics, *Journal of Political Economy* 15, 337-358.

Garcia-Mila, T., McGuire, T.J. and Porter, R.H., 1996. The Effect of Public Capital in State-Level Production Functions reconsidered. *The Review of Economic and Statistics* 78, 177-180.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2004. Bayesian Data Analysis 2nd edition. Chapman & Hall/CRC, Boca Raton.

Jalava, J., Ripatti, A., Pohjola, M. and Vilmunen, J., 2006. Biased Technical Change and Capital-Labour Substitution in Finland, 1902-2003. *The B.E. Journal of Macroeconomics*, 6 (Topics), Article 8.

Jones, C.I., 1995. Time Series Properties of Endogenous Growth Models, *Quarterly Journal of Economics* 110, 495-525.

Kamps, C., 2008. The Dynamic Effects of Public Capital: VAR Evidence for 22 OECD Countries. *International Tax and Public Finance*, Forthcoming.

Karras, G., 1999. Taxes and Growth: Testing the Neoclassical and Endogenous Growth Models. *Contemporary Economic Policy* 17, 177-188.

Kneller, R., Bleaney, M.F. and Gemmell, N., 1999. Fiscal Policy and Growth: Evidence form OECD Countries. *Journal of Public Economics* 74, 171-190

Kadiyala, K.R. and Karlsson, S., 1997. Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics* 12, 99-132.

Kocherlakota, N.R. and Yi, K.M., 1997. Is There Endogenous Long-Run Growth? Evidence from the United States and the United Kingdom. *Journal of Money, Credit, and Banking* 29, 235-262.

Perotti, R., 2005. The Empirical Case Against the Golden Rule of Government Investment. OECD, ECO/GEN.

Pritchett, L., 1996. Mind your P's and Q's. World Bank Policy Research Paper No. 1660.

Romero de Avila, T. D. and Strauch R., 2003. Public Finances and Long-Term Growth in Europe – Evidence From A Panel Data Analysis. European Central Bank, Working Paper No. 246.

Romp, W. and de Haan, J., 2007. Public Capital and Economic Growth: A Critical Survey. *Perspektiven der Wirtschaftspolitik* 8, 6–52.

Sims, C.A, Stock, J.H, and Watson, M.W., 1990. Inference in Linear Time Series Models with some Unit Roots. *Econometrica* 58, 113-144.

Sims, C. A. and Uhlig, H., 1991. Understanding Unit Rooters: A Helicopter Tour. *Econometrica* 59, 1591-1599.

Sims, C.A. and Zha, T., 1999. Error Bands for Impulse Responses. *Econometrica* 67, 1113-1156.

Strachan, R. and Inder, B. , 2004. Bayesian analysis of the error correction model. *Journal of Econometrics* 123, 307-325.

Turkkila, J., 2006. Valtion tuloveroasteikkojen reaalinen kiristyminen vuosina 1948-1997. unpublished manuscript.

Villani, M., 2005. Bayesian reference analysis of cointegration. *Econometric Theory* 21, 326-357.

Villani, M., 2006. Bayesian point estimation of the cointegration space. *Journal of Econometrics* 127, 645-664.

World Bank, 1994. World Development Report, Infrastructure for Development. Washington D.C.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, Inc., New York.

APPENDIX

TABLE 5 Estimated average discrepancies and deviance information criterions (DIC)

Summary statistics are based on the VAR model in Equation (13).

	Infrastructure Capital (1860-2003)		Infrastructure Capital (1900-2003)	
Lag	DIC	$\hat{D}_{avg}(y)$	DIC	$\hat{D}_{avg}(y)$
1	-2445	-2460	-1772	-833
2	-2698	-2723	-2025	-962
3	-2737	-2772	-2012	-932
4	2720	-2763	-1980	-905
5	-2685	-2740	-1942	-882
6	-2653	-2718	-1920	-870
	Infrastructure Capital (1948-2003)		**Paved Roads (1948-2003)**	
Lag	DIC	$\hat{D}_{avg}(y)$	DIC	$\hat{D}_{avg}(y)$
1	-1105	-1121	-833	-850
2	-1215	-1241	-962	-989
3	-1186	-1224	-932	-969
4	-1182	-1230	-905	-955
5	-1147	-1209	-882	-944
6	-1112	-1189	-870	-947

TABLE 6 Values of the Parameter's Posteriors of the VAR Model

Model

$$\begin{pmatrix} \Delta y_t \\ \Delta e_t \\ \Delta k_t \\ \Delta g_t \end{pmatrix} = \psi D_t + \sum_{j=1}^{k-1} \Pi_j \begin{pmatrix} \Delta y_{t-j} \\ \Delta e_{t-j} \\ \Delta k_{t-j} \\ \Delta g_{t-j} \end{pmatrix} + \begin{pmatrix} \pi_{yy} & \pi_{ye} & \pi_{yk} & \pi_{yg} \\ \pi_{ey} & \pi_{ee} & \pi_{ek} & \pi_{eg} \\ \pi_{ky} & \pi_{ke} & \pi_{kk} & \pi_{kg} \\ \pi_{gy} & \pi_{ge} & \pi_{gk} & \pi_{gg} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ e_{t-1} \\ k_{t-1} \\ g_{t-1} \end{pmatrix} + u_t$$

Panel (A) ψ includes a constant and trend if data support the use of it. The results of the regressions with shorter and longer lag lengths are similar.

Sample Period	Variable (*g*)	π_{yg}
1860-2003	Infrastructure Capital	0.144 (0.061)
1900-2003	Infrastructure Capital	0.196 (0.077)
1948-2003	Infrastructure Capital	0.158 (0.067)

Sample Period	Variable (*g*)	π_{kg}
1860-2003	Infrastructure Capital	0.013 (0.014)
1900-2003	Infrastructure Capital	0.034 (0.019)

Model

$$\begin{pmatrix} \Delta y_t \\ \Delta e_t \\ \Delta k_t \\ \Delta \ln(1-\tau_t) \\ \Delta g_t \end{pmatrix} = + \begin{pmatrix} \pi_{yy} & \pi_{ye} & \pi_{yk} & \pi_{y\tau} & \pi_{yg} \\ \pi_{ey} & \pi_{ee} & \pi_{ek} & \pi_{e\tau} & \pi_{eg} \\ \pi_{ky} & \pi_{ke} & \pi_{kk} & \pi_{k\tau} & \pi_{kg} \\ \pi_{\tau y} & \pi_{\tau e} & \pi_{\tau k} & \pi_{\tau\tau} & \pi_{\tau g} \\ \pi_{gy} & \pi_{ge} & \pi_{gk} & \pi_{g\tau} & \pi_{gg} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ e_{t-1} \\ k_{t-1} \\ \ln(1-\tau_{t-1}) \\ g_{t-1} \end{pmatrix}$$

$$+ \psi D_t + \sum_{j=1}^{k-1} \Pi_j \left(\Delta y_{t-i} \quad \Delta e_{t-i} \quad \Delta k_{t-i} \quad \Delta \ln(1-\tau_{t-i}) \quad \Delta g_{t-i} \right)' + u_t$$

Panel (B) in all post World-War II regressions ψ includes a constant. The results of the regressions with shorter and longer lag lengths are similar.

Sample Period	Variable (*g*)	π_{yg}
1948-2003	Infrastructure Capital	0.158 (0.067)
1948-2003	Kilometres of Paved Roads	0.017 (0.010)

Sample Period	Variable (*g*)	π_{kg}
1948-2003	Infrastructure Capital	0.012 (0.013)
1948-2003	Kilometres of Paved Roads	0.001 (0.002)

SUMMARY

Bayesilaisia sovelluksia dynaamisissa ekonometrisissa malleissa

Tämä väitöskirja koostuu viidestä esseestä, joissa tarjotaan muutamia uusia ideoita Bayesilaisen ekonometrian kenttään. Väitöskirja esittelee uuden ekonometrisen mallin ja näyttää miten subjektivismia (priori-informaatiota) voidaan hyödyntää talousteoreettisten rajoitteiden validiuden testaamisessa. Lisäksi se tarjoaa uuden tehokkaan tavan parantaa suosittujen uusien Keynesiläisten makromallien ennustetarkkuutta. Väitöskirja keskittyy dynaamisiin ekonometrisiin malleihin joista useimmat sisältävät epälineaarisia komponentteja. Tämä tuo oman haasteensa analyysiin. Esseissä Bayesilaisia metodeja käytetään ratkaisemaan ongelmia joita ei pystytä helposti ratkaisemaan traditionaalisen tilastotieteen keinoin.

Priorijakaumien keskeisestä roolista johtuen niiden kehittäminen, esittäminen ja raportointi saa esityksessä kohtalaisen paljon tilaa. Käytetyt priorijakaumat perustuvat eri talousteorioihin (kollektiivisesti subjektiiviseen informaatioon) ja muista alan empiirisistä tutkimuksista saatuun informaatioon. Joissakin malleissa käytetään myös standardeja epäinformatiivisia prioreita. Esityksessä lähes kaikkien mallien posteriori-jakaumat ovat ei-standardeja, joten niiden evaluoiminen vaatii numeeristen Monte Carlo -metodien käyttöä. Empiiristen tulosten luotettavuus riippuu voimakkaasti näiden metodien oikeanlaisesta soveltamisesta. Tästä syystä väitöskirjassa annetaan yksityiskohtaiset kuvaukset käytetyistä numeerisista algoritmeista ja simulaatiorutiineista. Lisäksi Markov-ketjujen konvergoituminen on varmistettu käyttämällä formaalia testidiagnostiikkaa.

Väitöskirjan keskeinen tutkimustulos on tämä, kun valtavirtaekonometria epäonnistuu tehtävässään tuottaa luotettavia empiirisiä tuloksia, Bayesilaiset ekonometriset menetelmät suoriutuvat monissa tapauksissa kyseisestä tehtävästä kunnialla. Esitetyt tulokset antavat olettaa, että Bayesilaiset menetelmät ovat erityisen hyödyllisiä varsinkin silloin kun vertailtavat taloustieteelliset mallit ovat ei-sisäkkäisiä tai mallin uskottavuusfunktio on monihuippuinen tai se saavuttaa maksimiarvonsa taloudellisessa mielessä järjettömillä parametriarvoilla. Monissa tapauksissa taloustieteellinen teoria tarjoaa myös tärkeää tai välttämätöntä informaatiota mallinnettavasta ilmiöstä, jota käsillä oleva aineisto ei pysty tarjoamaan. Tulokset ovat luonnollisesti hyvin tunnettuja Bayesilaisten ekonometrikkojen keskuudessa. Näin ollen väitöskirja kontribuoi alan kirjallisuuteen laajentamalla käytettävissä olevien tutkimusvälineistöä.

Esseet tuovat myös oman lisäarvonsa taloustieteelliseen tutkimukseen. Ensimmäisessä esseessä esitellään helposti sovellettavissa oleva estimointitapa uuden Keynesiläisen dynaamisen stokastisen yleisen tasapainon (DSGE) mallin implementointiin. Esseessä tutkitaan kuinka hyvin kyseinen malli kykenee ennustamaan talouden avainmuuttujia inflaatiota, korkoa ja tuotantoa. Tässä tut-

kimuksessa käytetään reaaliaikaista aineistoa, aineistoa jota Yhdysvaltain tilastoviranomaiset eivät ole jälkikäteen 'revisoineet', lisäämään tulosten painoarvoa. Ensimmäisen esseen tutkimustulokset osoittavat, että kyseinen yksinkertainen makromalli suoriutuu tehtävästään hyvin suhteessa vaihtoehtoisiin yleisesti käytettyihin ennustemalleihin. Tulos on hyvin mielenkiintoinen, koska aikaisemmassa kirjallisuudessa samoihin tuloksiin on päästy vain kasvattamalla voimakkaasti käytettyjen makromallien kokoa ja siten heikentämällä analyysin käyttökelpoisuutta.

Toisessa esseessä tutkitaan markkinaportfolion tuoton ja riskin välisen relaation robustisuutta. Tutkimus perustuu malliin $E_{t\text{-}1}(r_t) = \mu_0 + \mu_1 Var_{t\text{-}1}(r_t)$, missä r_t on markkinaportfolion ylituotto. Tutkimusaihe on tärkeä, koska yleisesti tiedetään, että μ_0:n tarpeeton lisääminen yhtälöön vääristää tuloksia, mutta toisaalta rajoite $\mu_0 = 0$ pakottaa odotetut ylituotot nollaksi hypoteesin $\mu_1 = 0$ alla. Rajoite $\mu_0 = 0$ on talousteoreettisesti hyväksyttävissä, mutta konfliktissa empiirisen rahoituskirjallisuuden kanssa. Tutkimusongelmaa lähestytään käyttämällä subjektiivista oletusta, että μ_0 on normaalisti jakautunut odotusarvolla nolla. Oletuksen robustisuutta testataan vaihtelemalla priori-jakauman varianssia. Intuitiivisesti pieni priori-varianssi on konsistentissa talousteorian kanssa, kun taas suuri priori-varianssi on konsistentissa empiirisen rahoituskirjallisuuden kanssa. Esitetyt tulokset indikoivat, että μ_0:n todellisen arvon ollessa lähellä nollaa kyseinen teoreettinen rajoite on hyväksyttävissä, mutta muussa tapauksessa rajoite on kyseenalainen.

Kolmas essee esittelee uuden ekonometrisen mallin, joka sallii Bayesilaisessa instrumenttimuuttujamallissa ei-vakioisen mallivirheen varianssin. Tutkimus on ensimmäinen laatuaan, mikä on suhteellisen yllättävää, sillä heteroskedastisuuden mallintaminen Bayesilaisessa ympäristössä parantaa estimaattien tarkkuutta ja ennustepäättelyn laatua. Käytännön sovelluksena tutkimuksessa estimoidaan Cobb-Douglas tuotantofunktio käyttämällä maakohtaista poikkileikkausaineistoa. Sovellus on relevantti, sillä alan kirjallisuus on huolellisesti raportoinut endogeenisuus- ja heteroskedastisuusongelmien olemassaolosta tämän tyyppisissä aineistoissa.

Neljännessä esseessä esitellään uusi kotitalouksien inflaatio-odotuksia mallintava malli. Mallissa luovutaan rationaalisten kotitalouksien oletuksesta, ja oletetaan, että taloudenpitäjät muodostavat oletuksensa mediassa esitettyjen inflaatiouutisten perusteella. Malli olettaa, että taloudenpitäjät havaitsevat inflaatioluvun mediasta tietyllä todennäköisyydellä. Estimoitujen mallien posteriori-todennäköisyyksien perusteella Yhdysvaltain aineisto tukee voimakkaasti esittämäämme mallia, suhteessa malliin jossa taloudenpitäjät ovat osittain rationaalisia. Neljännessä esseessä esitetty malli selittää hyvin myös kuluttajien odotusten heterogeenisyyttä. Tämä tulos perustuu mallin kuluttajatason versioon, joka on estimoitu mikrotason aineistoa hyväksi käyttäen.

Viidennessä esseessä tutkitaan infrastruktuuri-investointien pysyviä kasvuvaikutuksia käyttäen hyväksi hyvin pitkän aikavälin (1860-2003), aggregaattitason infrastruktuuri-investointi vuosiaineistoa. Tutkimuksen tilastollisessa päättelyssä käytetään posteriorianalyysiä, koska se tekee eksaktin päättelyn helpoksi myös yksikköjuuriaineistoilla. Tutkimustulokset indikoivat, että myös hyvin pitkällä aikavälillä infrastruktuurishokit aiheuttavat pysyviä vaikutuksia bruttokansantuotteeseen.

REFERENCES

Adolfson, M., Laséen, J., Lindé, J. and Villani, M., 2008. Evaluating an Estimated New Keynesian Small Open Economy Model. *Journal of Economic Dynamics and Control* forthcoming.

Adolfson, M., Lindé, J. and Villani, M., 2007a. Forecasting performance of an open economy DSGE model. *Econometric Reviews* 26, 289–328

Adolfson, M., Lindé, J. and Villani, M., 2007b. Bayesian Inference in DSGE Models - Some Comments. *Econometric Reviews* 26, 173–185.

Adolfson, M., Laséen, J., Lindé, J. and Villani, M., 2005. The Role of Sticky Prices in An Open Economy DSGE Model: A Bayesian Investigation. *Journal of the European Economic Association Papers and Proceedings* 3, 444–457.

Adolfson, M., Laseen, S., Linde, J. and Villani, M., 2007. Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics* 72, 481-511

Altug, S., 1989. Time-to-Build and Aggregate Fluctuations: Some New Evidence. *International Economic Review* 30, 889-920.

Atkeson, A. and Ohanian, L.E., 2001. Are Phillips curves useful for forecasting inflation?. *Federal Reserve Bank of Minneapolis Quarterly Review* 25, 2–11.

Barro, R.J., 1990. Government Spending in a Simple Model of Endogenous Growth. *Journal of Political Economy* 98, 103–125.

Bauwens, L., Lubrano, M. and Richard, J.-F., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University Press.

Bawa, V.S., Brown, S.J. and Klein, R.W., eds, 1979. Estimation Risk and Optimal Portfolio Choice, Amsterdam: North-Holland Publishing Co.

Benhabib, J. and Spiegel, M.M., 1994. The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data. *Journal of Monetary Economics* 34, 143–73.

Berry, D.A., Chaloner, K.M. and Geweke, J.K., 1996. Bayesian Analysis in Statistics and Econometrics: Essay in Honor of Arnold Zellner, New York: Wiley Series in Probability and Statistics.

Beyer, A. and Farmer, R., 2004. On the Indeterminacy of New Keynesian Economics. ECB Working Paper, 323.

Boarnet, M.G., 1995. Highways and Economic Productivity: Interpreting Recent Evidence. University of California School of Social Ecology Working Paper No 291.

Branch, W.A., 2007. Sticky information and model uncertainty in survey data on inflation expectations. *Journal of Economic Dynamics and Control* 31, 245–276.

Brave, S., Fisher, J.D., 2004. In search of a robust inflation forecast. *Economic Perspectives* 28, 12–31.

Canova, F., 1994. Statistical Inference in Calibrated Models. *Journal of Applied Econometrics* 9, S123-S144.

Canova, F., 2006. Monetary Policy and the Evolution of the US Economy. CEPR Discussion Papers No. 5467.

Canova, F. and Sala, L., 2005. Back to Square One: Identification Issues in DSGE Models. Manuscript, IGIER and Bocconi University.

Carroll, C.D., 2003. Macroeconomic Expectations of Households and Professional Forecasters. *Quarterly Journal of Economics* 118, 269-298.

Carroll, C.D., 2006. The Epidemiology of Macroeconomic Expectations, in: Blume, L., Durlauf, S. (Eds.), The Economy as an Evolving Complex System, III, Oxford University Press.

Chang, Y. and Schorfheide, F., 2003. Labor-Supply Shifts and Economic Fluctuations. *Journal of Monetary Economics* 50, 1751-1768.

Christiano, L.J., Eichenbaum, M. and Evans, C.L., 2005. Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* 113, 1–46.

Clarida, R., Galí, J. and Gertler, M., 1999. The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature* 37, 1661–1707.

Cogley, T., Morozov, S. and Sargent, T.J., 2005. Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control* 29, 1893-1925.

Cohen, D. and Soto, M. (2001). Growth and Human Capital: Good Data, Good Results. OECD Development Centre Technical Papers No.179.

DeGroot, M., 1970. Optimal Statistical Decisions. New York: McGraw-Hill.

DeJong, D. and Ingram, B., 2001. The Cyclical Behavior of Skill Acquisition. *Review of Economic Dynamics* 4, 536-561.

DeJong, D., Ingram, B. and Whiteman, C., 1996. A Bayesian Approach to Calibration. *Journal of Business Economics and Statistics* 14, 1-9.

DeJong, D., Ingram, B. and Whiteman, C., 2000. A Bayesian Approach to Dynamic Macroeconomics. *Journal of Econometrics* 98, 201-223.

Del Negro, M., 2003. Fear of Floating: A Structural Estimation of Monetary Policy in Mexico. Manuscript, Federal Reserve Bank of Atlanta.

Del Negro, M. and Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs. *International Economic Review* 45, 643–673.

Del Negro, M., and Schorfheide, F., 2008. Forming Priors for DSGE Models. Manuscript.

De Walque, G. and Wouters, R., 2004. An Open Economy DSGE Model Linking the Euro Area and the U.S. Economy. Manuscript, National Bank of Belgium.

Doan, T., Litterman, R. and Sims, C., 1984. Forecasting and Conditional Projections Using Realistic Prior Distributions. *Econometric Reviews* 3, 1–100.

Drèze, J.H., 1962. The Bayesian Appriach to Simultaneous Equations Estimation. Research Memorandum No. 67, Technological Institute, Northwestern University.

Drèze, J.H., 1976. Bayesian Limited Information Analysis of the Simultaneous Equation Model. *Econometrica* 44, 1045–1075.

Drèze, J.H. and Richard, J-F., 1983. Bayesian Analysis of Simultaneous Equations Systems. in Griliches and M.D. Intrilligator, eds., Handbook of Econometrics 1, Ch. 9, Amsterdam: North-Holland Publishing Co.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W., 2005. Growth Econometrics. in *Handbook of Economic Growth* 1, 555–677.

Fernández-Villaverde, J. and Rubio-Ramírez, J.F., 2004. Comparing dynamic equilibrium economies to data: a Bayesian approach. *Journal of Econometrics* 123, 153–187.

Fisher, J.D., Liu, C., Zhou, R., 2002. When can we forecast inflation?. *Economic Perspectives, Federal Reserve Bank of Chicago*, First Quarter, 30–42.

Fisher, W.D., 1962. Estimation in the Linear Decision Model. *International Economic Review* 3, 1–29.

Friedman, M. and Savage, L.J., 1948. The Utility Analysis of Choices Involving Risks. *Journal of Political Economy* 56, 279–304.

Friedman, M. and Savage, L.J., 1952. The Expected-Utility Hypothesis and the Measurability of Utility. *Journal of Political Economy* 60, 463–474.

Galí, J. and Rabanal, P., 2004. Technology Shocks and Aggregate Fluctuations:

How Well Does the RBC Model Fit Postwar U.S. Data. in Mark Gertler and Kenneth Rogof (eds.) NBER Macroeconomics Annual 2004, MIT Press, Cambridge, MA.

Geisel, M., 1975. Bayesian Comparison of Simple Macroeconomic Models. in Fienberg, S.E. and Zellner, A., eds., Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, 227-256, Amsterdam: North-Holland Publishing Co.

Geisser S., 1965. Bayesian Estimation in Multivariate Analysis. *The Annals of Mathematical Statistics* 36, 150-159.

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.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2004. Bayesian Data Analysis 2nd edition. Chapman & Hall/CRC, Boca Raton.

Geweke, J., 1999a. Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication. *Econometric Reviews* 18, 1-126.

Geweke, J., 1999b. Computational Experiments and Reality. Manuscript, University of Iowa.

Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return tradeoff after all. *Journal of Financial Economics* 76, 509–548.

Hoeting, J.A., Madigan, D.,Raftery A.E. and Volinsky, C.T., 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science* 14, 382–417.

Ireland, P., 2004. A Method for Taking Models to the Data. *Journal of Economic Dynamics and Control* 28, 1205-1226.

Jeffreys, H., 1939. Theory of Probability. 1st ed. Oxford, UK: Oxford University Press.

Jeffreys, H., 1961. Theory of Probability, 3rd ed. Oxford, UK: Clarendon Press.

Justiniano, A. and Preston, B., 2004. Small Open Economy DSGE Models-Specification, Estimation, and Model Fit. Manuscript, International Monetary Fund and Department of Economics, Columbia University.

Kadiyala, K.R. and Karlsson, S., 1997. Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics* 12, 99–132.

Kleibergen, F. and Zivot, E., 2003. Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics* 114, 29–72.

Kloek, T. and van Dijk, H.K., 1978. Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica* 46, 1–20.

Kocherlakota, N.R. and Yi, K.M., 1997. Is There Endogenous Long-Run Growth? Evidence from the United States and the United Kingdom. *Journal of Money, Credit, and Banking* 29, 235–262.

Koop, G., 2003. Bayesian Econometrics. West Sussex, UK: Wiley.

Laforte, J-P., 2004. Comparing Monetary Policy Rules in an Estimated General Equilibrium Model of the U.S. Economy. Manuscript, Federal Reserve Board of Governors.

Lancaster, T., 2004. An Introduction to Modern Bayesian Econometrics. Malden, MA: Blackwell.

Landon-Lane, J., 1998. Bayesian Comparison of Dynamic Macroeconomic Models. Ph.D. Dissertation, University of Minnesota.

Lanne M., Saikkonen , P., 2006. Why is it so difficult to uncover the risk-return tradeoff in stock returns?. *Economic Letters* 92, 118–125.

Leeper, E., and Sims, C., 1994. Toward a Modern Macroeconomic Model Usable for Policy Analysis" in Stanley Fischer and Julio Rotemberg (eds.) NBER Macroeconomics Annual 1994, MIT Press, Cambridge, MA.

Levin, A., Onatski, A., Williams, J. and Williams, N., 2005. Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models" in Mark Gertler and Kenneth Rogof (eds.) NBER Macroeconomics Annual 2005, MIT Press, Cambridge, MA.

Lindé, J., 2005. Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach. *Journal of Monetary Economics* 52, 1135–1149.

Litterman, R. B., 1980. A Bayesian Procedure for Forecasting with Vector

Litterman, R. B., 1986. Forecasting with Bayesian vector autoregressions - Five years of experience. *Journal of Business and Economic Statistics* 5, 25–38.

Lubik, T. and Schorfheide, F., 2005. A Bayesian Look at New Open Economy

Macroeconomics. in Mark Gertler and Kenneth Rogof (eds.) NBER Macroeconomics Annual 2005, MIT Press, Cambridge, MA.

Lubik, T.A. and Schorfheide, F., 2007. Do central banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics* 54, 1069–1087.

Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145–161.

Merton, R.C., 1973. An intertemporal capital asset pricing model. *Econometrica* 41, 867–887.

Otrok, C., 2001. On Measuring the Welfare Cost of Business Cycles. *Journal of Monetary Economics* 47, 61-92.

Papageorgiou, C., 2003. Distinguishing Between the Effects of Primary and Post-Primary Education on Economic Growth. *Review of Development Economics* 7, 622–635.

Poirier, D.J. 1988. Frequentist and Subjectivist Perspectives on the Problem of Model Building in Econometrics. *Journal of Economic Perspectives* 2, 121-144.

Poirier, D.J., 1989. A Report from the Battle front. *Journal of Business and Economic Statistics* 7, 137-139.

Poirier, D.J., 2004. What Is in a Word or Two?. Research paper, Department of Economics, University of California.

Qin, D., 1993. The Formation of Econometrics: A Historical Perspective, Oxford: Oxford University Press.

Onatski, A. and Williams, N., 2004. Empirical and Policy Performance of a Forward-looking Monetary Model," Manuscript, Columbia University.

Quintana, J., Chopra, V. and Putnam, B., 1995. Global Asset Allocation: Stretching Returns by Shrinking Forecasts. American Statistical Association's Section on Bayesain Statistical Science's Proceedings Volume.

Rabanal, P. and Rubio-Ramirez, J., 2005. Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach. *Journal of Monetary Economics* 52, 1151–1166.

Rabanal, P. and Tuesta, V., 2005. Euro-Dollar Real Exchange Rate Dynamics in an Estimated Two-Country Model: What Is Important and What Is Not?. Manuscript, International Monetary Fund.

Raiffa, H. and Schlaifer, R., 1961. Applied Statistical Decision Theory, Boston: Graduate School of Buseiness Administration, Harvard University.

Rothenberg, T.J., 1963. A Bayesian Analysis of Simultaneous Equation Systems. Report 6315, Econometric Institute, Netherlands School of Economics, Rotterdam.

Sala-i-Martin, X., Doppelhoffer G., and Miller R., 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach, *American Economic Review* 94, 813-835.

Sargent, T., 1989. Two Models of Measurements and the Investment Accelerator. *Journal of Political Economy* 97, 251-287.

Schirm, A.L. and DiCarlo J.V. (1998). Using Bayesian Shrinkage Methods to Derive State Estimates of Poverty, Food Stamp Program Eligibility, and

Food Stamp Program Participation. Report submitted to: U.S. Department of Agriculture.

Schorfheide, F., 2000. Loss Function-Based Evaluation of DSGE Models. *Journal of Applied Econometrics* 15, 645-670.

Schorfheide, Frank 2005. Learning and Monetary Policy Shifts. *Review of Economic Dynamics* 8, 392-419.

Shiller, R.J., 1997. Why Do People Dislike Inflation. in: Romer, C.D., Romer, D.H. (Eds.), Reducing Inflation: Motivation and Strategy, Chicago: University of Chicago Press.

Sims, C., 2002a. The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity* 2, 1–62.

Sims, C., 2002b. Solving Linear Rational Expectations Models. *Computational Economics* 20, 1-20.

Sims, C.A. and Uhlig, H., 1991. Understanding Unit Rooters: A Helicopter Tour. *Econometrica* 59, 1591–1599.

Sims, C.A. and Zha, T., 1999. Error Bands for Impulse Responses. *Econometrica* 67, 1113–1156.

Smets, F., and Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association* 1, 1123–1175.

Smets, F., and Wouters, R., 2005. Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach. *Journal of Applied Econometrics* 20, 161–183.

Smets, F. and Wouters R., 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review* 97, 586–606.

Stein, C., 1956. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. in Proceedings of the Third Berkley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkley: University of California Press, 197-206.

Stein, C., 1962. Confidence Sets for the Mean of a Multivariate Normal Distribution. *Journal of the Royal Statistical Society*, Series B, 24, 265-296.

Stock, J.H., Watson, M., 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J.H., Watson, M., 2007. Why Has U.S. Inflation Become Harder to Forecasts?. *Journal of Money, Credit and Banking* 39, 3–33.

Sungbae, A. and Schorfheide, F., 2007. Bayesian Analysis of DSGE Models. *Econometric Reviews* 26, 113–172.

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature* XXXVII, 112–156.

Tiao, G.C. and Zellner, A., 1964. On the Bayesian Estimation of Multivariate Regression. *Journal of Royal Statistical Society*, Series B, 26, 277-285

Zellner, A., 1965. Bayesian Inference and Simultaneous Equation Econometrics Models. paper presented to the First World Conference of the Econometric SOciety, Rome, Italy.

Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*. J. Wiley and Sons, Inc., New York.
Zellner, A., 1980. Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys, Amsterdam: North-HOlland Publishing Co.
Zellner, A., 2006. Past, Present and Future of Econometrics with Questions and Answers. *Medium for Econometric Applications* 14, 2–9.
Zellner, A., 2008. Bayesian Econometrics: Past, Present and Future. *Advances in Econometrics* 23 (Bayesian Econometrics) forthcoming.
Zellner, A. and Palm, F.C., 1974. Time Series Analysis and Simultaneous Equation Econometric Models. *Journal of Econometrics* 2, 17-54.
Zellner, A. and Palm, F.C., 1975. Time Series Analysis of Structural Models of the U.S. Economy. *Sankya*, Series C, 37, 12-56.
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.