THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Model Specification, Estimation and Inference in Studying Economic Output

Trinh Thi Thuy Dung
BA(Honours)

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2018
School of Economics

Abstract

The first aim of this thesis is to expand a range of econometric models available to investigate the behaviours and the drivers of economic output (e.g. real Gross Domestic Product (GDP)), labour productivity growth (real GDP per capita)). The second aim is to develop algorithms to estimate the proposed models. The third aim is to develop an inferential framework for each proposed model.

Chapter 2 generalises some of the existing methods to account for statistical noise in the estimation of a frontier, and to gain new insights on the drivers of labour productivity growth. We apply the model specification to the data set used in Kumar and Russell (*The American Economic Review*, 2002). We confirm that capital deepening is the main source of labour productivity growth, and also is the factor driving the transformation of the distribution of labour productivity from one single mode density to a density of two modes. When accounting for statistical noise in the estimation of the frontier, we find that capital deepening does not contribute to growth convergence, but instead efficiency change is a driver of growth convergence. The results indicate that improving technical efficiencies might reduce the gap between the poor and the rich.

Chapter 3 proposes a non-linear state space stochastic frontier model which provides a researcher a tool to jointly model the dynamic effects of environmental variables on both a production function and an inefficiency term. A test for time-variation in technical inefficiencies is also provided. We apply the proposed model to the economic study of 21 OECD countries. We wish to investigate under what channels foreign direct investment (FDI) affects a production process: through a production function, an inefficiency term, or both. The results show that FDI plays more important role in influencing the shift of the production frontier rather than the distribution of technical inefficiencies. This is to suggest that an output growth of an economy might further increase by investing FDI in innovation to improve technological change (shift of the frontier). We find statistical evidence in favour of time-variation in technical inefficiencies, and the temporal effect therefore should not be ignored in measuring technical inefficiencies.

Chapter 4 considers an econometric model where interdependent relationships between GDP and other key macroeconomic variables are allowed. We achieve impulse response functions of some key macroeconomic variables from a large VAR model with 119 variables. The precision of estimating a such large VAR model is obtained by using a dimension reduction approach, namely a reduced rank regression which has a specification independent to the order of the variables. The impulse responses of a selection of macroeconomic variables to a contractionary monetary policy have expected sign. For instance, a contractionary monetary policy is followed.

by a decrease in GDP, price level and an increase in unemployment rate. The results support the conventional channel of the effects of a contractionary monetary policy on a real economy. That is, a contractionary monetary policy tends to depress economic activity. An increase in interest rate often leads to an increase in the cost of capital, which then affects capital accumulation, and capital accumulation affects the labour demand, and the labour demand affects unemployment rate. Comparing the forecast performance of the proposed model to other popular approaches used in the literature of large VARs, we find that the proposed model provides a better forecast for GDP, consumer price index and producer price index in terms of a point forecast measure (e.g. mean squared forecast errors) and a density forecast measure (e.g. log predictive likelihood). In addition, we perform an extensive Monte Carlo simulation to investigate the performance of a range of econometric techniques (i.e. cross entropy, predictive likelihood and Laplace approximation) used in rank selection of a large VAR model. The results reveal that the approaches underestimate the rank. The lower rank models appear to provide better long horizon forecasts than the benchmark (the model with a correct rank) in terms of point forecast and density forecast measures.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

No publications.

Publications included in this thesis

No publications.

Contributions by others to the thesis

This thesis is supervised by Professor Rodney Strachan and Associate Professor Valentin Zelenyuk.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I would like to thank my supervisors, Associate Professor Valentin Zelenyuk and Professor Rodney Strachan for their guidance, patience, and of course for putting up with the stubborn student like me. I would like to thank my supervisor, Rodney Strachan for introducing and providing me guidance to the area of dimension reduction in high dimensional models. I would like to thank my supervisor, Valentin Zelenyuk, for building up the fundamental steps and skills for my early research career. I will never forget all the meetings you exchanged the ideas, taught me how to do research, and never give up on me even I have made many mistakes. Without the two, I would never be a better me like today.

Along the way, I have received many helpful comments from Associate Professor Alicia Rambaldi, Professor Chris Odonell, Professor Prasada Rao, Professor Knox Lovell, and Dr Antonio Peyrache. I would like to thank them all.

I would like to thank my friends, Linh Nguyen and Taylor Payne for being my "mental counsellors throughout the challenging years. Thank you for always being my side. Also, special thank to Thom Cuddhy for teaching me how to write scripts for distributed cloud computation.

I would like to thank professional staffs Alys Hohnen, Louise West, Marie Taylor, Bev Forrest, Tanya Mitchell, Magaret Cowan, Kathy Hung for their prompt assistance relating to administrative paper work.

Last but not least, to Mum, Dad, and my younger brother, I hope that I will soon no longer be a student (you probably hope for the day a long time ago), but I will still be studying for the rest of my life though. Thank you for your unconditional love and support. I love you all!

Keywords

Output growth, Stochastic frontier model, Local likelihood estimator, State space model, Vector autoregressive model, Reduced rank regression.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 140202, Economic Development and Growth 20% ;

ANZSRC code: 140302, Econometric and Statistical Methods, 80%.

Fields of Research (FoR) Classification

FoR code: 1402, Applied Economics, 20%

FoR code: 1403, Econometrics, 80%

Contents

List of Figures

List of Tables

List of Acronyms

AIC	Akaike Information Criteria
ARMH	Accept-reject Metropolis Hasting
ALPL	Average Log Predictive Likelihood
BRRR	Bayesian Reduced Rank Regression
CDF	Cumulative Distribution Function
CE	Cross Entropy
CV	Cross Validation
CSWSFE	Cumulative Sum of Weighted Squared Forecast Errors
CSSFE	Cumulative Sum of Squared Forecast Errors
CSLPL	Cumulative Sum of Log Predictive Likelihood
DEA	Data Envelopment Analysis
DFM	Dynamic Factor Model
DMUs	Decision Making Units
DGP	Data Generating Process
FDI	Foreign Direct Investment
GDP	Gross Domestic Product
HQ	Hannan-Quinn Information Criteria
KR	Kumar and Russell (2002)
LLE	Local Likelihood Estimator
LLS	Local Least Squares
MAFE	Median Absolute Forecast Errors
MSFE	Mean Squared Forecast Errors
MCMC	Markov Chain Monte Carlo
MLCV	Maximum Likelihood Cross-Validation
MH	Metropolis-Hasting

MVALPL	Multivariate Average Log Predictive Likelihood
MAI	Multivariate Autoregressive Index
OECD	Organisation for Economic Cooperation and Development
TE	Technical Efficiency
RRR	Reduced Rank Regression
RMSFE	Root Mean Squared Forecast Errors
Pred.Like	Predictive Likelihood
PDF	Probability Density Function
SVD	Singular Value Decomposition
SDEA	Stochastic Data Envelopment Analysis
SC	Schwarz Information Criteria
SD	Savage Dickey
STD.ERR	Standard error
VAR	Vector Autoregression
WMSFE	Weighted Mean Squared Forecast Errors

Prologue

"What you seek is seeking you"-Rumi

To me, the PhD experience is a journey in search of what I really want to pursue for my research career, and understand my ability. The journey brought me into joy, happiness and stress at the same time. The joy of learning new knowledge every day and searching solutions for research questions; the happiness of debugging the codes; and the stress of not knowing anything. Overall, I am so glad that I took this journey. Thanks to the experience, I have learnt to become an independent and a self-motivated person. The end of this journey is the beginning of a new journey that I am looking forward to venturing out.

Again, I will not be a better me like today without the guidance and support of my supervisors Rodney Strachan and Valentin Zelenyuk.

CHAPTER 1

Introduction and Motivation

Economic growth is important for economic development. It is associated with an improvement in quality of life and living standards of a nation. Higher economic growth is also associated with achieving other macroeconomic goals such as a low unemployment rate and government budget balance. Economists are often interested in answering important questions such as: What are the components and drivers of economic growth? How do the components change over time? How does economic growth respond to structural shocks such as monetary or fiscal shocks? The answers to these questions assist policy makers to design informed policies with the aim of boosting the growth of an economy.

This thesis makes a number of contributions in exploring a range of econometric models available to address the above questions. The first contribution is to expand the set of models available to investigate the behaviours and the drivers of economic output (e.g. Gross Domestic Product (GDP) growth, labour productivity growth (GDP per capita). The second contribution is the development of algorithms to estimate the proposed models. The third contribution is the development of an inferential framework for each proposed model to gain evidence addressing the important questions relating to the behaviours and drivers of economic output.

The second and third chapters of this thesis explain output growth using the concept of an efficient frontier (the maximum technically feasible output achievable from given inputs) from the literature on productivity and efficiency. The fourth chapter focuses on forecasting output growth as well as its response to monetary shocks. The model specification in the second chapter is in a nonparametric framework, whereas the model specifications in the third and fourth chapters are developed in parametric frameworks with flexible forms. The model specification in the third chapter explicitly accounts for time variation. The model in the fourth chapter permits joint modelling of a large set of potentially endogenous variables.

The study in **Chapter 2** is motivated by research using the concept of an efficiency frontier to explain economic growth (see Färe *et al.* (1994), Ray and Desli (1997), Henderson and Russell (2005), Mastromarco and Simar (2015), and among many others).[1] In this framework, the labour productivity growth (GDP per capita) is decomposed into three components: efficiency change (the movement towards the frontier), technological change (the shift of the frontier) and capital deepening (the movement along the frontier). The efficient frontier is the key element for the labour productivity decomposition, and is often estimated by Data Envelopment Analysis (DEA). DEA is a data-driven method that does not require *a priori* assumptions relating to the functional form of the frontier (e.g. Cobb Douglas, translog), or assumptions about market structure and technological change. However, DEA assumes no statistical noise (e.g. measurement error) in the estimation of the frontier. We aim to account for such a statistical noise in **Chapter 2**.

We generalise the local likelihood estimator considered in Kumbhakar *et al.* (2007) and Park *et al.* (2015) to account for statistical noise in the estimation of the efficient frontier, to measure technological change in a nonparametric framework, and to gain new insights about the drivers of output changes. To the best of our knowledge, our study is the first application of local likelihood estimator in the context of growth and convergence. We apply our model specification to the data set used in Kumar and Russell (2002). We find capital deepening is not only the main source of labour productivity growth, but also the factor driving the transformation of the distribution of labour productivity from a single mode density to a density of two modes. This result implies that the world is divided into two clubs: the rich and the poor. However, we do not find capital deepening contributes to growth convergence. We instead find statistical evidence suggesting technical efficiency change is a driver of growth convergence. This indicates that an improvement in technical efficiencies might reduce the gap between the poor and the rich.

Chapter 3 focuses on explaining the dynamic variations of technical efficiencies in terms of observable variables.[2] Such a study provides answers to a number of questions such as: What are the factors driving changes in technical inefficiencies? How do the effects vary over time? When explaining variations in technical inefficiency, researchers often face an issue as to where the factors influencing technical inefficiencies (sometimes known as "exogenous" or "environmental" variables) should be incorporated into a production process: through a technical ineffi-

[1] Compared to the conventional growth accounting approach, the efficient frontier approach avoids a number of caveats relating to the assumptions about functional forms and returns to scale of a production function (see Caselli (2005) for further discussion).

[2] Technical inefficiency measures how effective a decision making unit (e.g. a firm, an organisation, a country) is at using inputs to produce output relative to the technological frontier, which is reflected by a deviation from the observation toward the frontier.

ciency term or a production frontier, or both.

We are motivated by studies that incorporate the effects of the environmental variables on both a production function and an inefficiency term. One of the approaches is a conditional nonparametric frontier (Daraio and Simar (2005), Mastromarco and Simar (2015)). The approach does not require a specific assumption about the form of the production function. However, it ignores the presence of statistical noise in the estimation of the production frontier. Simar *et al.* (2017) overcome the issue of ignoring statistical noise by using local linear least squares. Obtaining statistical inferences (e.g. standard errors, hypothesis testing, etc.) for such a nonparametric stochastic frontier model is not a trivial task. A proper bootstrap algorithm is required. An alternative is the stochastic parametric frontier model with a scaling property (Simar *et al.* (1994), Wang and Schmidt (2002), Alvarez *et al.* (2006)). The scaling property feature implies that the distribution of inefficiency is multiplicatively decomposed into a function of the environmental variables and a random variable (which is known as a *base inefficiency*). The distribution of the base inefficiency is independent of the environmental variables. The statistical inferences for technical inefficiencies in a stochastic parametric frontier model are directly obtained, however, an assumption about a production function is required.

In **Chapter 3**, building upon the model specification of Wang and Schmidt (2002) and Alvarez *et al.* (2006), we extend their static model specification to a dynamic case. We allow the base technical inefficiencies to vary over time and across decision making units (DMUs) (e.g. countries, firms, organizations, and etc.). In practice, it might be the case that the base inefficiencies are not identically distributed across DMUs. DMUs can learn from their mistakes and have incentives to improve their own efficiencies over time. The adjustments in technical efficiencies can vary across DMUs due to heterogeneity factors such as the regulation system and the average quality of labours. To accommodate these, we allow the coefficients associated with the environmental variables in technical inefficiencies to vary across DMUs and over time. We also provide a formal test for time-varying effects of technical inefficiencies.

Our time-varying specification of technical inefficiencies is closely related to that of Tsionas (2006) who allows the logarithm of the base inefficiencies to follow a first order autoregressive process.[3] Deviating from Tsionas (2006), we do not restrict the dynamic pattern of the base technical inefficiencies to be identical across DMUs. Nor do we restrict the effects of environmental variables on technical inefficiencies to be time-invariant and homogeneous across DMUs. Both Tsionas's (2006) model specification and ours contain a non-linear feature in

[3] Ahn and Sickles (2000), Emvalomatis *et al.* (2011) also consider the autoregressive process of inefficiencies in their model specifications. However, their studies do not consider the effects of environmental variables on inefficiencies.

technical inefficiencies. The algorithm used to estimate our proposed model is more efficient. Tsionas (2006) uses a Metropolis Hasting with random walk algorithm to address the non-linearity issue, and estimates the time-varying parameters at each time period. In our work, we adapt the idea of using a Gaussian approximation from Durbin and Koopman (1997) and Chan and Strachan (2014) to provide an efficient way of sampling the non-linear parameters in technical inefficiencies. We further use the precision-based algorithm (Chan and Jeliazkov (2009)) which utilises the sparse structure of the matrices to improve efficiency and accelerate the computation time.

We provide a formal test for time-variation in technical inefficiencies. To perform such a hypothesis test, we first re-parametrize our model specification in the form of a non-centred parametrisation developed in Frühwirth-Schnatter and Wagner (2010). We then adopt the idea of using the Savage-Dickey ratio in Chan (2016) to perform the test for time variation in the parameters. To the best of our knowledge, this is the first study providing a formal test for time variation in technical inefficiencies.

We also allow the coefficients associated with a production function to be time-varying. The time-varying parameters with a random walk in a production/cost function have recently been considered in a number of studies such as Jin and Jorgenson (2010) and Peyrache and Rambaldi (2012). These studies focus on the measurement of technological change, and do not consider the effects of environmental variables upon the production frontier. Peyrache and Rambaldi (2012) show the time-varying parameters in a production function provide a flexible way of capturing technological change compared to a time-dummy variable approach. Unlike the previous studies, we allow the coefficients associated with a production to follow a random walk with an intercept, and show that such a time-varying specification encompasses the commonly used production functions (e.g. Cobb-Douglas, translog) with a linear trend.

We apply the proposed model to the study of economic growth of 21 OECD countries over the period 1970-2011. We wish to investigate whether foreign direct investment (FDI) plays a role in influencing the frontier or the technical inefficiency component. We find that FDI plays an important role in influencing the production frontier rather than technical inefficiencies. Capital deepening is still the main factor driving output growth. We also find evidence suggesting inefficiencies of the countries are time-varying, and therefore the temporal effects should be included in capturing the behaviours of technical inefficiencies.

Chapter 4 studies the behaviours of output (real Gross Domestic Product (GDP)) growth in an econometric system which allows for interdependent relationships among output growth and other macroeconomic variables such as interest rates, the inflation rate, the unemployment

rate, and many others. We are interested in impulse response functions of some important macroeconomic variables to structural shocks (e.g. monetary shocks), and the forecast of the variables. A vector autoregressive (VAR) model is a natural tool to address these issues (see Blanchard and Quah (1989), Sims (1992), Cochrane (1998), and many others). Many studies find that large VAR models tend to provide better forecasts (see Banbura *et al.* (2010), Carriero *et al.* (2011), Koop (2013)). Such large VAR models often risk over-parametrisation, which in turn can worsen the forecast performance and impulse responses. Researchers in the literature of large VARs have made significant contributions in proposing various approaches to handling the over-parametrisation issue, and developing algorithms which are computationally tractable and efficient in the estimation of such large VAR models.

In **Chapter 4** we use the reduced rank regression (RRR) approach to tackle the issue of over-parametrisation and contribute to this research line on large VARs in the following aspects. To achieve a unique identification in RRR in Bayesian analysis, a linear normalization condition is often used. A limitation of this specification is the lack of invariance in the variable ordering. That is, empirical results can change if we change the ordering. The model specification in this paper is invariant to the ordering of the variables. The invariant specification is built upon a singular value decomposition (SVD) of the matrix that is potentially reduced rank. Such an invariant specification using SVD is also considered in other research areas such as factor models (Chan *et al.* (2017) and Kaufmann and Schumacher (2013)), co-integrating vector error correction models (Strachan and Inder (2004), Koop *et al.* (2010)).

Given an SVD specification, sampling some of the identified parameters from their conditional posteriors is not computable as the parameters belong to the Stiefel manifold, i.e. they are subject to semi-orthogonal constraints. To simplify computation we make use of the parameter expansion approach (see Ghosh and Dunson (2009), Koop *et al.* (2010), Chan *et al.* (2017)) to map the parameters from the Stiefel manifold to the real space. More importantly, the conditional posterior of the parameters in the expanded model are simple to draw from, i.e. they belong to the class of normal distributions. Sampling the parameters from their conditional posterior is therefore efficient. Further, using the conventional method to sampling the parameters from its high-dimensional posterior density can be demanding when the size of the VAR system increases. To improve the computation time we adopt the computation algorithm exploiting a particular Kronecker structure of covariance matrices as suggested Carriero *et al.* (2016a) to our context.

Regarding rank selection of the VAR coefficient matrix, we first estimate a set of models with all possible values of the rank, and then choose a model yielding the highest marginal likeli-

hood. In our context, the marginal likelihood computation is not trivial as it does not have a closed form. We approximate the marginal likelihood by using a number of methods such as cross entropy, predictive likelihood, and Laplace approximation, which are often used in a low dimensional set-up in a number of studies such as Geweke (1996), Geweke (2001), Strachan and Inder (2004), Chan and Eisenstat (2015). The performance of these methods in a high dimensional set-up is still not quite clear. In **Chapter 4**, we conduct an extensive Monte Carlo simulation to investigate the performance of the methods in a high dimensional set-up. The Monte Carlo results suggest a downward bias in estimating the rank which potentially leads to a selection of misspecified models. The downward bias is more evident when the singular values of VAR coefficients are small (close to zero). This result is not a surprise, but the implication for forecasting is important and therefore we explore this area of application.

We carry out the recursive forecasting exercise to explore the forecast performance of misspecified rank models against that of a benchmark model (a model with a correct rank). Using the point forecast measures (e.g. mean squared errors, mean absolute errors, and the weighted mean squared errors) we find that RRR with lower rank models perform worse than the benchmark for shorter forecast horizons. However, for longer forecast horizons they appear to perform as well as or even better than the benchmark. With regard to density forecasts, the lower rank models appear to provide better forecasts than the benchmark.

We construct forecasts and impulse responses for a set of macroeconomic variables to a monetary policy shock using a large VAR model with 119 dependent variables. Within these 119 variables, we focus on forecasting real gross domestic product, consumer price index, federal funds rate, civilian unemployment rate, industrial production growth, money stock M2, real personal consumption, producer price index, and personal consumption expenditure. We compare the forecasting performance of the macroeconomic variables using our proposed approach to various popular alternatives. Our results suggest no single model dominates in forecasting the variables. Our proposed model provides better forecasts for GDP, consumer price index, and producer price index. The impulse response of the macroeconomics of interest have expected sign and support economic theories. For example, a contractionary monetary policy is followed by a decrease in GDP, price level and an increase in unemployment rate.

Chapter 5 of this thesis concludes and provides future research areas.

CHAPTER 2

Productivity Growth and Convergence: Local Linear Maximum Likelihood

2.1 Introduction

Explaining economic growth is one of the central problems that economists have grappled with for decades. A comprehensive survey on the voluminous literature of economic growth is referred to Barro (1999), Sickles *et al.* (2015), Badunenko *et al.* (2017) and references therein. In this chapter, we only discuss the studies most closely related to our research. A growth accounting approach was often considered as a main tool-kit to explain economic growth (see Solow (1956), Maddison (1987)). The approach explains output changes by input changes and the unexplained residual which is often referred to as "technological progress" (see Barro (1999)). The accounting approach neglects the fact that economies utilise their productive capacities differently. A number of studies propose an alternative approach which accounts for the potential inefficiency of production, and gives us a structural interpretation of the unexplained residual (see Färe *et al.* (1994), Ray and Desli (1997), Kumar and Russell (2002), Henderson and Russell (2005), Mastromarco and Simar (2015) and among others). This chapter is motivated by this line of research on economic growth.

The basic idea underpinning the studies that account for the potential inefficiency of production in explaining economic growth is to use the concept of efficient frontier from the literature on productivity and efficiency. The efficient frontier approach avoids the assumptions about a production function form (Cobb-Douglas) and returns to scale (constant), which are often made in the growth accounting approach (see Caselli (2005) for further discussion). In the efficient frontier framework labour productivity growth is decomposed into the tripartite components:

efficiency change (the movement towards the frontier), technological change (the shift of the frontier) and capital deepening (the movement along the frontier due to an increase of capital per worker).[1] Such a tripartite decomposition of labour productivity growth provides a framework to address a number of questions such as: Which countries make the most efficient use of their own inputs? Is labour productivity growth of the countries driven by improving their efficiencies? Or is it due to capital deepening or technological change? The answers to these questions assist policy makers to design informed policies. For example, if efficiency changes play an important role in economic growth, then policies should focus on the need for improvement in technical efficiency such as improving the legal system, providing workers with more training to improve their skills. If economic growth is driven by input changes, policies should aim to change the input mix, e.g. more investment in capital stock.

The construction of the efficient frontier is the key element for the labour productivity decomposition. Data Envelopment Analysis (DEA), a data-driven approach implemented with standard mathematical programming, is often used to estimate the efficient frontier (Färe *et al.* (1994), Henderson and Russell (2005), Badunenko *et al.* (2008)). The conventional DEA does not require *a priori* assumptions about the functional form of the frontier (e.g. Cobb-Douglas, translog function, etc.) or the structure of markets or technological change. However, a drawback of this approach is to assume no statistical noise (e.g. measurement error, outliers) and so could potentially suffer from extreme observations (Simar (2003), Cazals *et al.* (2002)). The estimated efficiency scores calculated from datasets that include outliers can be misleading. In this chapter, we generalise some of the existing models (i.e. Local Likelihood Estimator (LLE) and Stochastic DEA (SDEA)) to account for such statistical noise and to estimate technological change nonparametrically in the context of two-period panel data.

One of the well-known approaches that account for statistical noise is a parametric stochastic frontier, pioneered by Aigner *et al.* (1977) and Meeusen and van den Broeck (1977). This approach is implemented in some studies of economic growth and convergence such as Koop *et al.* (1999), Hultberg *et al.* (1999), Koop *et al.* (2000), and Kumbhakar and Wang (2005). A requirement of a parametric stochastic frontier approach is an assumption about the functional form of the frontier. Kumbhakar *et al.* (2007) relax the assumption by using the local likelihood estimator (LLE).[2] Their proposed method is first applied to continuous explanatory variables. Park *et al.* (2015) later generalise it to encompass categorical variables. Both methods are originally applied to cross-sectional data.

[1] Labour productivity measures how much a nation produces per unit of labour given its resources (e.g. capital) and available technology.

[2] Further discussion about nonparametric estimation in stochastic frontier model is referred to Parmeter and Kumbhakar (2014).

Another estimator that accounts for such a statistical noise we adapt here is stochastic DEA (Simar (2007), Simar and Zelenyuk (2011)), hereafter referred to as SDEA. SDEA can be viewed as a combination of the estimation methods of DEA and LLE. While DEA uses "original" input-output data to construct the frontier, SDEA uses the data filtered from noise via LLE.

A contribution of this chapter is to generalise the methods of Kumbhakar *et al.* (2007), Park *et al.* (2015) and Simar and Zelenyuk (2011), to estimate technological change nonparametrically in the context of two-period panel data using LLE and SDEA. To the best of our knowledge this study is the first application of LLE and SDEA on economic growth and convergence, and also is the first paper that models technology change using LLE and SDEA. Some new insights about the drivers of labour productivity growth are found by using the new methods.

The rest of this chapter is structured as follows. Section 2.2 presents the nonparametric deterministic and stochastic methods used in the chapter. Section 2.3 discusses some computational issues relating to the estimation of the efficient frontier using stochastic nonparametric approaches. Section 2.4 discusses the empirical results while Section 2.5 concludes the chapter.

2.2 Methodology

We first present the estimation of the world frontier using DEA, LLE, and SDEA, and then discuss the similarities and differences amongst the approaches in the context of a single output.

Let $X = (K, L) \in \mathbb{R}^2_+$ denote a vector of two inputs, where K is physical capital and L is labour, and let $Y \in \mathbb{R}_+$ denote a single output (GDP). The technology describing a conversion of the inputs to GDP in period t for all countries, can be characterised via a set

$$\mathcal{T}_t = \{(X, Y) : X \text{ can produce } Y \text{ in period } t\}. \tag{2.2.1}$$

The boundary of this technology set is referred to as the (production) frontier.

2.2.1 Data Envelopment Analysis (DEA)

Under the regularity conditions, the technology set (2.2.1) can be characterised by the Shephard (1970) output distance function. The output distance function of a country i in period t relative

to the frontier in period t is defined as

$$D_t(X_{it}, Y_{it}) \equiv \inf\{\delta > 0 : (X_{it}, Y_{it}/\delta) \in \mathcal{T}_t\}. \tag{2.2.2}$$

The distance function is also viewed as a criterion of efficiency of a country i, denoted as $TE_{it} \equiv D(X_{it}, Y_{it})$. A country is considered as fully efficient when $TE_{it} = 1$, and technically inefficient if $TE_{it} < 1$.

The true technology set $\mathcal{T}_t$ is not known, but can be estimated by DEA. The constant return to scale technology set at period t, $\hat{\mathcal{T}}_t^{DEA}$, estimated using DEA is defined as:[3]

$$\hat{\mathcal{T}}_t^{DEA} = \left\{(X, Y) \in \mathbb{R}_+^2 \times \mathbb{R}_+^1 : \sum_{i=1}^{n} \alpha_{it} Y_{it} \geq Y; \sum_{i=1}^{n} \alpha_{it} X_{it} \leq X; \alpha_{it} \geq 0, i = 1, \ldots, n\right\}, \tag{2.2.3}$$

where α_{it} are intensity variables. The efficiency score for a country i in period t is then obtained via the following optimisation problem

$$\widehat{TE}_{it}^{DEA} = \min_{\delta, \alpha_{1t}, \ldots, \alpha_{nt}} \{\delta \geq 0 : (X_{it}, Y_{it}/\delta) \in \hat{\mathcal{T}}_t^{DEA}\}. \tag{2.2.4}$$

2.2.2 Local Likelihood Estimator (LLE)

Our nonparametric stochastic frontier here is a variant of the models proposed by Kumbhakar *et al.* (2007) and Park *et al.* (2015). We generalise their approaches to estimate technological change (shift of the frontier) nonparametrically for the two time-period panel data.

Let y_{it} denote the logarithm of output $(\log(Y_{it}))$ for a country i in period t, and x_{it} denote the logarithm of inputs $(\log(X_{it}))$. Given the sample of n countries observed in periods $t = b$ and $t = c$, the nonparametric stochastic frontier model is:

$$y_{it} = m(x_{it}, d_{it}) - u_{it} + v_{it}, \tag{2.2.5}$$

where $m(.,.)$ is the unknown frontier function; a two-sided error v_{it} representing statistical noise normally distributed with zero mean and variance $\sigma_v^2(x, d)$, $(v_{it}|x_{it} = x, d_{it} = d) \sim N(0, \sigma_v^2(x, d))$; a one-sided error u_{it} representing technical inefficiency following a half-normal distribution, $(u_{it}|x_{it} = x, d_{it} = d) \sim |N(0, \sigma_u^2(x, d))|$; a dummy variable d_{it} takes the value of 1 if an observation is in the period c, and 0 otherwise.

[3]The variable return to scale implies that $\sum_{i=1}^{n} \alpha_{it} = 1$ and $\alpha_{it} \geq 0$; the non-increasing return to scale $\sum_{i=1}^{n} \alpha_{it} \leq 1$ and $\alpha_{it} > 0$.

Given the model (2.2.5), $\theta(x,d) = [m(x,d), \sigma_u^2(x,d), \sigma_v^2(x,d)]'$ is the unknown function to be estimated. The conditional log-likelihood function of θ is given by

$$\mathcal{L}(\theta|x_{it}, d_{it}, y_{it}) = \sum_{t=1}^{2}\sum_{i=1}^{n} \log\Big[g(y_{it}, \theta(x_{it}, d_{it}))\Big], \tag{2.2.6}$$

where $g(.,.)$ is the known conditional probability density function of y_{it} given $x_{it} = x$ and $d_{it} = d$. The functional form of $g(.)$ is presented in Section 2.3.

Direct maximisation of the log-likelihood function (2.2.6) over θ is intractable and suffers from over-fitting (Tibshirani and Hastie (1987)), and therefore the local version of (2.2.6) is considered instead, i.e.

$$\begin{aligned}\mathcal{L}_n(\theta_0, \Theta_1; x, d) &= \sum_{t=1}^{2}\sum_{i=1}^{n} \Big(\log\, g(y_{it}, \theta_0(x,d) + \Theta_1'(x,d)(x_{it} - x))\Big) \\ &\quad \times \quad \mathcal{K}_H(x_{it} - x) W_\gamma(d_{it}, d),\end{aligned} \tag{2.2.7}$$

where $\Theta_1(x,d)$ is the first derivative of $m(x,d)$ at given point (x,d) , and $\Theta_1'(x,d)$ is the transposed vector of $\Theta_1(x,d)$.

The basic idea of LLE is to replace $\theta(x_{it}, d_{it})$ in (2.2.6) by polynomial approximation in a neighbourhood of a fixed interior point (x,d) in the direction of continuous variables x, and maximise (2.2.7) instead. A maximiser of the local likelihood (2.2.7) is $(\hat{\theta}_0(x,d), \hat{\Theta}_1(x,d))$. The proposed estimator of $\theta(x,d)$ is then $\hat{\theta}_0(x,d)$, while $\hat{\Theta}_1(x,d)$ is an estimator of the first partial derivative of $\theta(x,d)$ with respect to x. The asymptotic properties and the bias in a finite sample of LLE estimator are referred to in Theorem 3.1 in Park *et al.* (2015). The form of the bias term shows that the bias increases with the curvature of the LLE estimator, $\theta(.)$.

In (2.2.7), the neighbouring observations around (x,d) are placed more weight than those further from (x,d). The weight is chosen by using a kernel approach. For the discrete variable d_{it}, we use the kernel function proposed by Aitchison and Aitken (1976), i.e. $W_\gamma(d_{it}, d) = (1-\gamma)^{\mathbf{I}(d_{it} \neq d)} \gamma^{\mathbf{I}(d_{it}=d)}$, with a bandwidth $\gamma \in [0.5, 1]$, and an indicator function $\mathbf{I}(B)$ taking a value of 1 if B holds, and 0 otherwise. Other kernel functions for categorical variables are proposed in Racine and Li (2004) and Park *et al.* (2015). Their kernel densities are used for categorical variables with two or more than two values whilst the kernel proposed by Aitchison and Aitken (1976) is designed for categorical variables with two values only. Park *et al.* (2015) show that their proposed kernel is equivalent to that of Aitchison and Aitken (1976) in the context of binary categorical variables. For a vector of $p \times 1$ of continuous variables x_{it}, we choose the most common kernel used in empirical studies, the Gaussian product, i.e.

$\mathcal{K}_H(x_{it}, x) = |H|^{-1} \prod_{s=1}^{p} k\left(\frac{x_{it}^s - x^s}{h^s}\right)$, where $H = \text{diag}(h^1, \ldots, h^p)$, with h^s being a bandwidth of continuous variables x^s ($s = 1, \ldots, p$), $|H|$ being the determinant of H, and k being an univariate Gaussian density. Other kernel choices for continuous variables such as biweight kernel, triweight kernel, Epanechnikov kernel can be used (see Pagan and Ullah (1999, pp. 54-55) for further details).

The bandwidths γ and H in (2.2.7) are chosen using the leave-one-out maximum likelihood cross-validation estimator, which is a data-driven approach. They are jointly estimated by maximising

$$MLCV(H, \gamma) = \frac{1}{2n} \sum_{t=1}^{2} \sum_{i=1}^{n} \log[g(y_{it}, \hat{\theta}_{H,\gamma}^{-i}(x_{it}, d_{it}))], \tag{2.2.8}$$

where $\hat{\theta}_{H,\gamma}^{-i}(x_{it}, d_{it})$ is the 'leave-the ith observation-out' version of LLE discussed above. Further discussion about the estimation of the frontier and the selection of the bandwidths is given in Section 2.3.

To obtain the predictor of an individual efficiency score for an observation i in period t, we follow Park *et al.* (2015) and adapt the procedure in Jondrow *et al.* (1982, p.235) to the context of LLE with a 2-period panel

$$\begin{aligned} \widehat{TE}^{LLE}(x_{it}, d_{it}) &\approx \exp(E(-u_{it}|\hat{\varepsilon}_{it})) \\ &= \hat{\mu}^*(x_{it}, d_{it}) + \hat{\sigma}^*(x_{it}, d_{it}) \frac{\phi(-\frac{\hat{\mu}^*(x_{it}, d_{it})}{\hat{\sigma}^*(x_{it}, d_{it})})}{1 - \Phi(-\frac{\hat{\mu}^*(x_{it}, d_{it})}{\hat{\sigma}^*(x_{it}, d_{it})})}, \end{aligned} \tag{2.2.9}$$

where $\phi(.), \Phi(.)$ are the probability density function (pdf) and cumulative distribution function (cdf) of a standard normal variable, respectively; $\hat{\varepsilon}_{it} = y_{it} - \hat{m}^{(0)}(x_{it}, d_{it})$, $\hat{\sigma}^2(x_{it}, d_{it}) = \hat{\sigma}_u^2(x_{it}, d_{it}) + \hat{\sigma}_v^2(x_{it}, d_{it})$, $\hat{\mu}^*(x_{it}, d_{it}) = -\frac{\hat{\sigma}_u^2(x_{it}, d_{it})\hat{\epsilon}_{it}}{\hat{\sigma}^2(x_{it}, d_{it})}$, $\hat{\sigma}^*(x_{it}, d_{it}) = \frac{\hat{\sigma}_u^2(x_{it}, d_{it})\hat{\sigma}_v^2(x_{it}, d_{it})}{\hat{\sigma}^2(x_{it}, d_{it})}$.

2.2.3 Stochastic DEA (SDEA)

SDEA is another approach that accounts for statistical noise in the estimation of the production frontier. Simar (2007) and Simar and Zelenyuk (2011) propose SDEA in the context of continuous variables. We extend their approaches to handle both continuous and categorical variables.

The construction of the constant returns to scale, period-t technology using SDEA, $\hat{\mathcal{T}}_t^{SDEA}$ is

defined as

$$\hat{\mathcal{T}}_t^{SDEA} = \Big\{ \sum_{i=1}^{n} \alpha_{it} \hat{Y}_{it}^* \geq Y; \sum_{i=1}^{n} \alpha_{it} X_{it} \leq X, \alpha_{it} \geq 0, i = 1, \ldots, n \Big\}, \tag{2.2.10}$$

where $\hat{Y}_{it}^* = \exp(\hat{y}_{it}^*)$ with $\hat{y}_{it}^*$ being the fitted value of the output estimated by LLE ((2.2.7)). The associated individual efficiencies obtained from SDEA follow a similar procedure to that obtained from DEA (equation (2.2.4)), i.e.

$$\widehat{TE}_{it}^{SDEA} = \min_{\delta, \alpha_{1t}, \ldots, \alpha_{nt}} \{\delta \geq 0 : (X_{it}, Y_{it}/\delta) \in \hat{\mathcal{T}}_t^{SDEA}\}. \tag{2.2.11}$$

SDEA can be viewed as a symbiosis of DEA and LLE estimations. LLE is used to filter the noise; and DEA is used to super-impose the monotonicity and convexity upon the LLE-estimated frontier.

The asymptotic properties of the SDEA estimator have not, to date, been discussed in the literature and this is beyond the scope of this thesis. However, the Monte Carlo simulations in Simar and Zelenyuk (2011) reveal that such an approach performs well in constructing the frontier in the presence of statistical noise.

2.2.4 Comparison among DEA, SDEA and LLE

This section discusses the similarities and differences among DEA, SDEA, and LLE approaches in the context of a single output. This allows researchers to better understand about the strength and limitation of each approach and then they can choose the most suitable approach for their research. For a single output case, both deterministic and stochastic frontier models can be presented as follows

$$Y_{it} = f_t(K_{it}, L_{it}) e^{-u_{it}} e^{v_{it}}, \tag{2.2.12}$$

where $f_t(K_{it}, L_{it})$ is a production function in period t.

The key differences across the approaches lie in the following aspects: the presence of statistical noise, the regularity conditions imposed on the technology set, and the estimation of the efficiency scores. DEA assumes statistical noise v_{it} equal to 0, and thus the stochastic frontier model (2.2.12) becomes $Y_{it} = f_t(K_{it}, L_{it}) e^{-u_{it}}$. The technical efficiency therefore measures the distance from observed data to the frontier, which is simply the ratio of observed data to maximal output given inputs, denoted as $TE_{it}^{DEA} = Y_{it}/f_t(K_{it}, L_{it})$. The estimation of frontier $f_t(K_{it}, L_{it})$ and the associated efficiency scores are obtained through mathematical program-

ming algorithms as described in (2.2.4).

Unlike DEA, LLE accounts for statistical noise v_{it}, i.e. v_{it} is different from 0. The technical efficiency in this context is the ratio of observed output to the corresponding stochastic frontier, denoted as $TE_{it}^{LLE} = Y_{it}/(f_t(K_{it}, L_{it})e^{v_{it}})$. The approximation of the technical inefficiency is obtained by using the procedure described in (2.2.9). The frontier $f_t(K_{it}, L_{it})$ is estimated by the maximum likelihood-based method.

Both DEA and SDEA require mathematical programming algorithms to estimate the individual technical efficiency scores. DEA uses "original" data $\{(K_{it}, L_{it}, Y_{it})\}_{i=1}^{n}$ to construct the frontier $f_t(K_{it}, L_{it})$, while SDEA uses "filtered" data $\{(K_{it}, L_{it}, \hat{Y}_{it}^{*})\}_{i=1}^{n}$, where $\hat{Y}_{it}^{*}$ is filtered from possible noise using LLE. SDEA contains two stages of estimation. The first stage is to obtain the fitted data $\hat{Y}_{it}^{*}$ using a stochastic nonparametric model (e.g. LLE), and the second stage is to estimate the frontier and the associated technical inefficiency using DEA.

A number of regularity conditions (such as convexity, monotonicity) are implicitly imposed on the frontier estimated in DEA and SDEA while they are not in LLE. If one wish to impose the regularity conditions in LLE, one can adapt the methods discussed in Henderson and Parmeter (2009), Parmeter *et al.* (2014), Du *et al.* (2013). In our context, the constant returns to scale can be directly imposed in LLE by estimating the normalized version of the frontier where output and all inputs are normalized by labour (i.e. $\log(Y_{it}/L_{it}) = m_t(K_{it}/L_{it}, 1) - u_{it} + v_{it}$). The frontier estimate, $\hat{Y}_{it}^{*}$, is then recovered as $\hat{Y}_{it}^{*} = (\exp(\widehat{\log(Y_{it}/L_{it})}))L_{it}$.

2.3 Computational Details

In this section we discuss a number of computational issues arising in the estimation of the frontier using LLE and the bandwidth selection using the local likelihood cross-validation estimator.

2.3.1 The Frontier Estimation

Theoretically, the estimator of the function of interest $\theta(x,d) = [m(x,d),\, \sigma_u^2(x,d), \sigma_v^2(x,d)]'$ is obtained by maximising the log-local likelihood (2.3.1), i.e.

$$\begin{aligned}(\hat{\theta}_0(x,d), \hat{\Theta}_1(x,d)) &= \underset{\theta_0,\Theta_1}{\operatorname{argmax}}\, \mathcal{L}_n(\theta_0,\Theta_1;x,d), \\ &= \underset{\theta_0,\Theta_1}{\operatorname{argmax}} \sum_{t=1}^{2}\sum_{i=1}^{n} \Big(\log\, g(y_{it}, \theta_0(x,d) + \Theta_1'(x,d)(x_{it}-x))\Big) \\ &\quad \times \quad \mathcal{K}_H(x_{it}-x) W_\gamma(d_{it},d). \end{aligned} \tag{2.3.1}$$

$\hat{\theta}_0(x,d)$ is the proposed estimator of $\theta(x,d)$, while $\hat{\Theta}_1(x,d)$ is the estimator of the first partial derivative of θ with respect to x. The functional form of $g(y_{it}, \theta_0(x,d) + \Theta_1'(x,d)(x_{it}-x))$ for a stochastic frontier model (2.2.5) with a normal distribution for an error term v_{it}, and a half-normal distribution for an inefficiency term u_{it} is given by Park *et al.* (2015).[4]

$$\begin{aligned} g(.) &= \frac{2}{\sqrt{\Psi_u(\sigma_{0u}^2,\sigma_{1u}^2;x_{it}-x) + \Psi_v(\sigma_{0v}^2,\sigma_{1v}^2;x_{it}-x)}} \\ &\times \quad \phi\left(\frac{y - \Psi_m(m_0,m_1;x_{it}-x)}{\sqrt{\Psi_u(\sigma_{0u}^2,\sigma_{1u}^2;x_{it}-x) + \Psi_v(\sigma_{0v}^2,\sigma_{1v}^2;x_{it}-x))}}\right) \\ &\times \quad \Phi\Bigg(-(y - \Psi_m(m_0,m_1;x_{it}-x)) \times \\ &\qquad \frac{1}{2}\sqrt{\frac{\Psi_u(\sigma_{0u}^2,\sigma_{1u}^2;x_{it}-x)}{\Psi_v(\sigma_{0v}^2,\sigma_{1v}^2;x_{it}-x)(\Psi_u(\sigma_{0u}^2,\sigma_{1u}^2;x_{it}-x) + \Psi_v(\sigma_{0v}^2,\sigma_{1v}^2;x_{it}-x))}}\Bigg), \end{aligned}$$

where $\phi(.), \Phi(.)$ are pdf and cdf of a standard normal variable, respectively, and

$$\begin{aligned} \Psi_m(m_0,m_1;x_{it}-x) &= m_0(x,d) + m_1'(x,d)(x_{it}-x), \\ \Psi_v(\sigma_{0v}^2,\sigma_{1v}^2;x_{it}-x) &= \sigma_{0v}^2(x,d) + \sigma_{1v}'^2(x,d)(x_{it}-x), \\ \Psi_u(\sigma_{0u}^2,\sigma_{1u}^2 x_{it}-x) &= \sigma_{0u}^2(x,d) + \sigma_{1u}'^2(x,d)(x_{it}-x). \end{aligned}$$

The function $\theta(x,d) = [m(x,d),\, \sigma_u^2(x,d), \sigma_v^2(x,d)]'$ in our context is subject to a number of constraints. First, the variances $\sigma_u^2(x,d)$ and $\sigma_v^2(x,d)$ must be non-negative. Second, the variances of an inefficiency term u_{it} and a random error term v_{it} are smaller than that of the composited error $\varepsilon_{it}(= -u_{it} + v_{it} = y_{it} - m(x_{it},d_{it}))$, i.e. $\sigma_u^2(x_{it},d_{it}) \leq \sigma_\varepsilon^2(x_{it},d_{it})$,

[4] Alternatively, one can adopt the parametrisation of the function $g(.)$ for continuous variables in Kumbhakar *et al.* (2007) to the context of continuous and discrete variables.

$\sigma_v^2(x_{it}, d_{it}) \leq \sigma_\varepsilon^2(x_{it}, d_{it})$. The non-negativity restrictions on the functions of $\sigma_u^2(x, d)$ and $\sigma_v^2(x, d)$ can be avoided in the local linear estimator by working with the logarithm of the functions (e.g. Kumbhakar *et al.* (2007), Park *et al.* (2015)), i.e. $\tilde{\sigma}_u^2(x, d) = \log(\sigma_u^2(x, d))$ and $\tilde{\sigma}_v^2(x, d) = \log(\sigma_v^2(x, d))$. The vector of optimisation variables is therefore given by $\tilde{\theta}(x, d) = [m(x, d),\ \tilde{\sigma}_u^2(x, d), \tilde{\sigma}_v^2(x, d)]'$. To ensure the second constraints hold, a penalty is introduced to the local log likelihood function. Such a penalty assigns a large negative value on the conditional local log likelihood function (i.e. the local likelihood approaches 0) when the estimators do not belong to their space.

Some initial values for $\theta_0(x, d)$ and $\Theta_1(x, d)$ are required. Here we choose the local least squares estimators (Simar *et al.* (2017)) as starting values for $\hat{\theta}_0(x, d)$ and $\hat{\Theta}_1(x, d)$. That is $\hat{\theta}_0(x, d) = [\hat{m}_0^{CLLS}(x, d), \log(\widehat{\sigma}_{0u}^{2LLS}(x, d)),\ \log(\widehat{\sigma}_{0v}^{2LLS}(x, d))]$, where $\hat{m}_0^{CLLS}(x, d)$ is the corrected estimator for the local least squares estimator, i.e. $\hat{m}_0^{CLLS}(x, d) = \hat{m}_0^{LLS}(x, d) + \sqrt{\frac{2\hat{\sigma}_{0u}^{2LLS}(x,d)}{\pi}}$, $\widehat{\sigma}_{0u}^{2LLS}(x, d)$ and $\widehat{\sigma}_{0v}^{2LLS}(x, d))$ are the local least squares (LLS) estimators and $\hat{\Theta}_1(x, d) = [\hat{m}_1^{LLS}(x, d), 0, 0]$. We also consider the parametric maximum likelihood and parametric least squares estimators from a parametric stochastic frontier model (i.e. $y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 d_{it} - u_{it} + v_{it}$) as the initial values of θ_0 and Θ_1. We find that the choice of the local least squares estimators as starting values appear to be stable in the optimization algorithm applied for the data in our empirical study.

2.3.2 Bandwidth Selection

The asymptotic theory behind the estimator requires fairly mild assumptions on the bandwidths, $h^s \propto n^{-1/(p+4)}$ with $s = 1, \ldots, p$, and $\lambda \propto n^{-2/(p+4)}$ so that the bandwidths are not influenced when they are scaled by a constant (Park *et al.* (2017)). In practice, choosing bandwidths is a crucial task in nonparametric estimation as the performance of the estimation depends on the choice of the bandwidths. Many studies have proposed and investigated the performance of various bandwidth selection approaches in the context of a kernel density estimator (see Marron (1988), Park and Marron (1990), Hall and Marron (1991), Henderson and Parmeter (2015) for a summary). The performance of the approaches in the context of nonparametric regression estimator is still limited. A brief discussion about the performance of the approaches in a kernel density estimator is useful for the study of their performance in a nonparametric regression estimator.

Two common bandwidth selection approaches for a kernel density estimator are an optimal selection (e.g. a rule-of-thumb approach (Silverman (1986)) and a data-driven approach (e.g. LLS

cross-validation (Bowman (1984), Hall (1983), Stone (1984), local likelihood cross-validation (Duin (1976), van Es (1991)). Each of the approaches has its own advantages and disadvantages. The optimal selection approach is computationally simple. A LLS cross-validation is one of the most widely used for data-driven bandwidth selection approaches. Its popularity stems from the asymptotic theory regarding the bandwidths obtained from a LLS cross-validation (see Stone (1984)). The optimisation of LLS cross-validation can be challenging in practice. Multiple local minima is often observed in LLS cross-validation (see Marron (1988), Hall and Marron (1991), Hall *et al.* (2004)). To select a bandwidth when multiple local minima occurs, some studies suggest to choose the largest local minimiser (Marron (1988), Park and Marron (1990)), while Hall *et al.* (2004) suggest to choose the second smallest local minima. To the best of our knowledge, there is still no theoretical explanation explaining why the largest or the second is the best for a bandwidth selection.

In this chapter, we use a data-driven approach, and carefully examine whether multiple local minima occurs here. We follow Park *et al.* (2015) and choose a leave-one-out maximum likelihood cross-validation (MLCV) estimator to choose the bandwidths γ and H in (2.2.7). The MLCV has a strong intuitive appealing in the sense that it chooses a bandwidth which maximises the likelihood of observing a given data point:

$$\text{MLCV}(H,\gamma) = \frac{1}{2n}\sum_{t=1}^{2}\sum_{i=1}^{n}\log[g(y_{it},\hat{\theta}^{-i}_{H,\gamma}(x_{it},d_{it}))], \tag{2.3.2}$$

where $\hat{\theta}^{-i}_{h,\gamma}(x_{it},d_{it})$ is the "leave-the ith observation-out" version of LLE discussed above. The boundary of the bandwidth H is $(0,\infty)$, and that of γ is $[0.5,1]$ with $\gamma > 0.5$ indicating the frontiers in two time periods are differentiated. The estimators of bandwidths are obtained by minimising the negative MLCV over H and γ.

Some initial values are required for the optimisation of (2.3.2). The benchmark starting value for the bandwidth H is Silverman (1986) rule of thumb, i.e. $\text{std}(x)n^{-1/5}$, where $\text{std}(x)$ is a standard deviation of x. We then scale the rule of thumb bandwidth by a constant a, where a is from $\{1.2,1.5,1.8,2\}$. The starting values for the bandwidth γ are from the set $\{0.5,0.6,0.7,0.8,0.9,0.95\}$. The set of initial values for bandwidth selection we consider is the combination between the values of H and γ. Given the set of the starting values, our results show that resulting estimator for the bandwidth γ varies within a small interval (0.8-0.99). On the other hand, the resulting bandwidth H for the continuous variable given the starting values vary within a larger range (0.6,1.3). To examine if multiple local minima occurs here, we plot the estimated bandwidths H against the negative MLCV (see Figure 2.3.1). Figure 2.3.1 reveals that the negative MLCVs associated with the bandwidths in the intervals of (0.95,1.09) and (1.14, 1.31) are

marginally different. We also observe that a smaller bandwidth (less than 1) causes the efficient frontier to have substantially less smooth or rugged shape, particularly at the low capital-labour ratio income. Here we follow Marron (1988) and choose the largest bandwidth of 1.307.

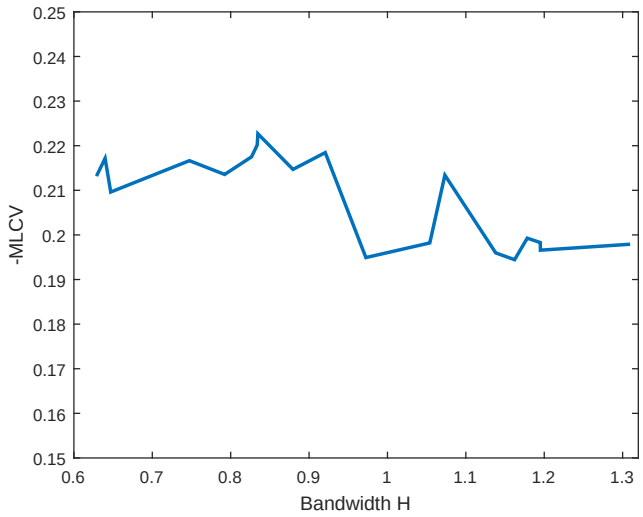


Figure 2.3.1: The bandwidth H versus the negative of MLCV values

It is noted that we also try various bandwidth values in the range of (1.14, 1.31) to check the robustness of our conclusions. We find that the main conclusions are robust to the choices. The bandwidth γ associated with our chosen H is 0.97, which indicates technological progress in our context.

2.4 Empirical Study

In this section we employ DEA, SDEA, and LLE approaches to KR's data, and investigate the robustness of KR's results when statistical noise is included in the estimation of the frontier. The data used in this chapter comprise 57 countries for two periods 1965 and 1990.

We investigate the following issues: What are drivers of labour productivity (output per worker) growth? What are the components of labour productivity growth driving the world convergence? What are the components causing the bimodal transformation of the distribution of labour productivity growth over the period 1965-1990? To address these we follow KR and decompose labour productivity based on Fisher decomposition (Caves *et al.* (1982), Färe *et al.*

(1994)). The labour productivity is decomposed into three factors: efficiency change (the movement toward to the frontier), technological change (the shift of the frontier) and capital deepening (the movement along the frontier). The basic idea underlying the labour productivity decomposition is that the countries are assumed to operate under a common efficient frontier. Accordingly, the countries can be thought of as performing on or within the frontier, and the distance from a country to the frontier reflects the technical efficiency. Over time, a country can perform more efficiently and move toward the frontier, or the frontier itself can shift outward over time, indicating technological progress. A country can also move along the frontier by changing inputs.

Mathematically, the decomposition of labour productivity growth between a base period b and a current period c is defined as

$$\frac{Y_c/L_c}{Y_b/L_b} = \frac{TE_c}{TE_b} \times \left(\frac{f_c(K_c/L_c,1)}{f_b(K_c/L_c,1)} \times \frac{f_c(K_b/L_b,1)}{f_b(K_b/L_b,1)}\right)^{1/2} \times \left(\frac{f_c(K_c/L_c,1)}{f_b(K_b/L_b,1)} \times \frac{f_b(K_c/L_c,1)}{f_b(K_b/L_b,1)}\right)^{1/2}$$

$$= \text{EFF} \quad \times \quad \text{TECH} \quad \times \quad \text{KACCUM}, \qquad (2.4.1)$$

where $f_t(.,.)$ is the potential output per unit of labour at period t. The first component on the right-hand side EFF measures the effect of efficiency change. The second component TECH reflects technological change. The measure is based on Fisher decomposition which is a geometric average of the effects of technological change which is based on the current period capital/efficiency-labour ratio $\left(\frac{f_c(K_c/L_c,1)}{f_b(K_c/L_c,1)}\right)$, and based on the base-period capital/labour ratio $\left(\frac{f_c(K_b/L_b,1)}{f_b(K_b/L_b,1)}\right)$. Similar to TECH, the last component measuring capital deepening (KACCUM) is constructed by using a geometric average of the effects of KACCUM which is based on the current and the base-period capital labour ratio.

2.4.1 Analysis of Tripartite Decomposition of Labour Productivity Growth

Table 2.4.1 presents the percentage change of labour productivity and its tripartite decomposition for some selected countries in the sample used by KR. The results for all countries are presented in Appendix 2.A. Column (i) of Table 2.4.1 provides the results estimated via DEA, column (ii) reports the values computed via SDEA, and column (iii) presents the estimates obtained via LLE.[5]

[5] Focusing on the estimates obtained from DEA, we note that the estimates of the tripartite decomposition for each individual country are marginally different from those presented in Table 2 of KR. The difference is minimal on average. These minor differences could be due to numerical errors resulting from different optimization methods

Table 2.4.1: Percentage change of labour productivity and its tripartite decomposition from 1965 to 1990 using DEA, SDEA and LLE

Methods		**DEA(i)**			**SDEA(ii)**			**LLE(iii)**		
Countries	**Labour Productivity Change**	ΔEFF	ΔTECH	ΔKACCUM	ΔEFF	ΔTECH	ΔKACCUM	ΔEFF	ΔTECH	ΔKACCUM
Australia	42.67	8.06	14.03	15.78	-14.62	8.33	54.26	-9.34	8.35	55.15
Austria	95.15	-14.57	15.49	97.79	-36.34	7.69	184.64	-21.03	7.94	185.67
Belgium	78.36	22.04	12.94	29.41	-11.46	8.58	85.52	-6.65	8.59	86.22
France	78.29	3.59	16.72	47.45	-23.34	8.65	114.04	-12.55	8.56	114.64
Germany	70.75	13.24	14.49	31.70	-29.86	17.76	106.72	-22.07	7.05	130.41
Hong Kong	251.08	120.17	2.36	55.79	83.51	14.23	67.48	46.64	8.74	59.59
Iceland	66.42	-9.13	1.92	79.69	-19.24	2.65	100.73	-10.66	9.01	113.59
Japan	208.52	3.27	15.16	159.42	-28.19	6.49	303.45	-21.80	6.58	302.67
New Zealand	7.42	-15.82	9.49	16.54	-35.27	9.02	52.22	-21.24	8.95	52.69
South Korea	424.45	41.47	3.04	259.76	10.80	6.78	343.30	9.73	6.34	341.60
Peru	-16.11	-32.15	1.51	21.80	-38.12	8.57	24.86	-30.47	8.05	24.59
United Kingdom	60.74	-3.70	1.39	64.63	-11.50	2.00	78.05	-7.32	9.33	91.31
United States	31.29	0.00	10.02	19.33	-12.55	0.34	49.62	-8.49	8.80	62.86
Zimbabwe	11.38	37.24	2.47	-20.80	36.77	5.64	-22.92	30.15	5.39	-22.77
Mean (over 57 countries)	**75.06**	**5.24**	**6.17**	**58.47**	**-8.50**	**5.31**	**86.04**	**-6.33**	**7.36**	**86.56**
std err of the mean	**(10.48)**	**(3.50)**	**(1.86)**	**(7.19)**	**(3.07)**	**(1.48)**	**(10.05)**	**(2.29)**	**(0.88)**	**(10.17)**

ΔEFF=(EFF-1) $\times$ 100 represents percentage change in efficiency; ΔTECH=(TECH-1) $\times$ 100 represents percentage change in technology; ΔKACCUM=(KACCUM-1) $\times$ 100 represents percentage change in capital-labour ratio (capital deepening).

For LLE, the bandwidth selection is based on the leave-one-out likelihood cross-validation, where the bandwidth for the continuous variable (capital per worker) is 1.307 and the bandwidth for the categorical variable (time dummy) is 0.97. Here we use the kernel proposed by Aitchison and Aitken (1976) for the discrete variable and the Gaussian kernel for the continuous variable.

When comparing the three sets of estimates from DEA, SDEA and LLE, we note that the mean contribution of technological change to labour productivity growth appear to be robust across the methods; however, there are discrepancies in the contribution of technical efficiency change and capital deepening.[6] In particular, both stochastic approaches (LLE and SDEA) suggest a negative change in technical efficiencies (-8.5% in the case of SDEA and approximately -5.4% in the case of LLE) while deterministic DEA indicates a positive average change (5.2%) over the period of 1965-1990. The standard errors of the mean decomposition suggest that the contribution of efficiency change to labour productivity growth is significant in SDEA and LLE while it is insignificant in DEA.

The technical explanation for the difference in the result across the methods is due to the presence of statistical noise in the construction of the world production frontier in SDEA and LLE. Any deviation of observations from the frontier is explained by both inefficiency and statistical noise in SDEA and LLE whereas it is due to inefficiency in DEA. Another possible explanation for the decrease in average efficiency is that capital was under utilised during the period e.g. capital was idle or not efficiently used, resulting in inefficiency in capital utilisation. The idleness of capital can be explained as a result of rational *ex ante* investment plans (see Winston (1974) for further discussion), or as a consequence of unwanted accidents that occur after a firm is built. For instance, firms are shut down or temporarily closed due to an inadequate demand for a product or an unexpected high cost incurred. The machines are therefore not efficiently used. After the oil crisis 1973, a number of countries experienced economic stagflation (e.g. high increase in raw material price and unemployment rate) for a long time (see Grubb *et al.* (1983), Bruno and Sachs (1985)). In addition, the newer version of Penn World Table 8.1 revealed a substantial reduction in the average annual hours worked by persons engaged during the period 1965-1990, e.g. the reduction was 19% in Belgium and greater than 22% in France and Germany. It is noted that we do not advocate the argument that a reduction in working hours implies lower productivity or high inefficiency. Our justification is that a reduction in working hours implies the number of hours capital used in a production process likely decreased, and therefore capital was idle, and was not efficiently used. Another explanation of the decrease in average efficiency is due to the diffusion of general purpose technologies (GPT), i.e. it often

employed to compute individual technical efficiencies in KR's and our study. In this study, we used a quadratic programming sub-problem and implemented it in MATLAB version 2013.

[6]The monotonicity and convexity assumptions on the technology set are implicitly imposed on DEA and SDEA, while they are not in LLE. One could impose such economic restrictions on the frontier in LLE by adapting some of the approaches discussed in Henderson and Parmeter (2009), Parmeter *et al.* (2014) and references therein. The estimates of the first order derivatives are all non-negative, which indicate that monotonicity holds for LLE in our application. Since we use a local linear approximation for the LLE estimator, the concavity assumption on the production function is not be able to checked in our application as the estimation of the second derivatives is required.

takes time before the appearance of new GPT (for example, computers and information processing equipment) can be utilized 100% efficiently (Helpman (1998), Helpman and Rangel (1999) and references therein).

Capital deepening is confirmed as the primary source of worldwide productivity growth over the 25 year period. Yet the magnitude of its mean contribution to labour productivity growth seems to be underestimated when statistical noise is not modelled in the estimation of the world frontier. DEA suggests that its magnitude is 59% over 25 years, which implies 1.9% per annum. On the other hand, the new methods suggest the larger contribution from capital deepening, above 85% (approximately 2.5% per annum).[7] A possible explanation for the difference in the magnitude of capital deepening in DEA, SDEA and LLE is due to the presence of statistical noise in the construction of the world production frontier by SDEA and LLE. For instance, both DEA and SDEA implicitly impose regularity conditions on the technology set, and require mathematical programming algorithms to determine the frontier and to compute technical efficiencies. However, the observations used to construct the frontier in SDEA are filtered from the noise while they are not in DEA. This is the key difference between SDEA and DEA. Compared to SDEA and DEA, LLE is more flexible in the sense that it does not implicitly imposed regularity conditions on the technology set. Some of the conditions can be imposed by adapting some approaches discussed in Henderson and Parmeter (2009), Parmeter *et al.* (2014) and references therein.

Another possible explanation for the large contribution of capital deepening to labour productivity growth is due to remarkably high growth rate of capital stock over the 25 year period: the average growth for all countries was 286.33% and the growth rate was even higher for some countries such as Austria (380.42%), Japan (794%) and South Korea (1459.6%). The remarkably high increase in capital stock can be explained by the attempts of many countries to rebuild capital stock damaged during World War II (Smolny (2000)) and the rise of Asia (Young (1995). Also, when other factors such as human capital and foreign direct investment are considered in the decomposition of labour productivity growth, a number of studies found the contribution of capital deepening to labour productivity growth is smaller, but it is still one of the major attributes (e.g. Henderson and Russell (2005), Badunenko *et al.* (2013), and an unpublished paper of Gu and Russell (2010)). For the purpose of the comparison with KR, we focus only on the tripartite decomposition considered in KR.

[7]The compound annual growth rate is used to compute the growth per annum.

2.4.2 Convergence Hypothesis

A question we wish to address here is: what factors contribute to the world convergence (the poor catching up to the rich). Convergence studies often focus on two types of convergence: β-convergence which implies that output growth per capita is negatively dependent on an initial economic level (poor countries grow faster than rich countries), and σ-convergence which is defined as lowering of variance of output per capita among economies in time (Young *et al.* (2008)). β-convergence is necessary but not sufficient condition for the existence of σ-convergence (Sala-i Martin (1996), Young *et al.* (2008)). In the analysis of β-convergence, researchers often run a cross-country regression the growth rate of output per capita (between the base period and the current period) on the output per capital (at the base period) (see Barro and Sala-i Martin (1992), Sala-i Martin (1996)). Quah (1996) introduces an alternative model relating to the distribution of labour productivity to study whether the poor catches up with the rich. Quah (1996) argues that the conventional approach (Generalised Least Squares regression) tends to provide information relating the performance of a single country towards to its steady state while investigating the distribution of labour productivity provides the performance of an economy relative to the others. We investigate both of the approaches- Generalised Least Squares regression and the labour productivity distributions in this section to better understand about the convergence.

Convergence hypothesis-Analysis of labour productivity distributions

Panel (a) of Figure 2.A.1 plots the distributions of output per worker of the 57 countries in 1965 and 1990. Over this 25-year period, the distribution of labour productivity appeared to transform from a single peak into a twin peak distribution with a higher mean. We use Silverman's (1981) test to formally test for the twin-peak transformation. The null hypothesis is that the distribution has s modes and the alternative is that it has more than s modes. Using the test, we find that the 1965 distribution contained a single mode (p-value = 0.35) whereas the 1990 distribution has more than one mode (p-value = 0.00) but no more than two (p-value=0.21).

Following KR, we investigate what components of the tripartite decomposition cause the bimodal transformation of the labour productivity distribution. Our analysis of the change in the distribution of labour productivity is based on Li's (1996, 1999) test which is a nonparametric test for the comparison of two unknown distributions (e.g. $r_1(z)$ and $r_2(z)$). A null hypothesis is $r_1(z) = r_2(z)$ for all z, and the alternative is $r_1(z) \neq r_2(z)$ for some z. In our context, we want to test if the distribution of labour productivity based on 1990 capital-labour ratio is statistically different from the counterfactual labour productivity which is constructed by consequently in-

troducing each of the decomposition.

Re-write the tripartite decomposition of labour productivity changes in (2.4.1) as follows:

$$y_{Lc} = (\text{EFF} \times \text{TECH} \times \text{KACCUM}) y_{Lb}, \tag{2.4.2}$$

where $y_{Lc} = Y_c/L_c$, $y_{Lb} = Y_b/L_b$. The labour productivity distribution in the current year can be constructed by successively multiplying labour productivity in the base year by each of the three factors. This in turn allows us to construct counterfactual distributions by sequential introduction of each of these factors (where $b = 1965$ and $c = 1990$). For instance, the counterfactual 1990 labour-productivity distribution of the variable,

$$y^K_{L1990} = \text{KACCUM} y_{L1965},$$

isolates the effect on the distribution of changes in capital accumulation only, no technology, and no technical inefficiency.

Table 2.4.2 contains the results of Li (1996, 1999) for testing the equality between the 1990 labour productivity distribution and its counter-factual distributions based on efficiency change, technological change and capital deepening.[8]

Table 2.4.2: Testing for the significant contributions of various sources to the bimodal transformation of output per worker

Methods	**DEA (i)**		**SDEA (ii)**		**LLE (iii)**	
H_0 : Distributions are equal H_1 : Distributions are not equal	**Test statistic (ia)**	**Bootstrap p-value (ib)**	**Test statistic (iia)**	**Bootstrap p-value (iib)**	**Test statistic (iiia)**	**Bootstrap p-value (iiib)**
$pdf(y_{L1990})$ vs. $pdf(y_{L1965})$	3.52	0.00	3.52	0.00	3.52	0.00
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{EFF})$	2.78	0.00	4.69	0.00	4.26	0.00
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{TECH})$	2.06	0.01	2.27	0.01	2.79	0.01
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{KACCUM})$	0.77	0.19	0.80	0.16	1.26	0.07
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{EFF} \times \text{TECH})$	1.61	0.04	3.76	0.00	3.71	0.00
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{EFF} \times \text{KACCUM})$	0.68	0.27	0.37	0.59	0.10	0.90
$pdf(y_{L1990})$ vs. $pdf(y_{L1965} \times \text{TECH} \times \text{KACCUM})$	-0.06	0.93	1.83	0.03	2.13	0.01

$pdf(.)$ stands for probability density function.

We use the bootstrapped Li (1996) tests with 2000 bootstrap replications and Silverman's (1986) rule-of-thumb bandwidth.

[8] Appendix 2.A provides the figures depicting the distribution of labour productivity based on the actual 1990 data and its counter-factual distributions following the order KACCUM-TECH-EFF for all methods. For other orders, the figures are available upon request.

We confirm the results reported in Table 3 of KR for DEA. We find that the conclusions of most hypotheses are robust across the methods (see rows 1-6 of columns (ib), (iib), (iiib) in Table 2.4.2). That is, when the effects of efficiency change and technological change are first introduced, the null hypothesis of equality between the actual and the counterfactual distributions are rejected. However, this is not the case for capital accumulation, which suggests capital accumulation primarily accounts for the bimodal distribution. (Figures 2.A.1-2.A.3 are consistent with the findings.)

In row 7, the conclusion is different across the methods. SDEA and LLE results suggest the rejection of the hypothesis regarding the equivalence of the 1990 labour productivity distribution and the counter-factual distribution incorporating technology change and capital deepening. It is not the case in DEA. Also, SDEA and LLE produce p-values smaller than 0.05 regarding hypothesis tests where the counter-factual distributions are based on technological change or a combination of the other components (see row 3, 5 and 7 of columns (iib), (iiib) in Table 2.4.2). This result suggests that technological change dampens the twin-peak divergence process, and so it is possibly a source impeding the division of the world into two clubs: the rich and the poor.

Convergence hypothesis-Generalised least squares approach

Another approach to identify the determinants of world convergence is a Generalised Least Squares regression (GLS) (see Barro and Sala-i Martin (1992), Sala-i Martin (1996)). The approach provides an equation with labour productivity growth on the left-hand side, which is explained by other factors on the right-hand side. We regress each of the components of the labour productivity decomposition on the output per worker in the base period (1965). The signs and significance of the resulting coefficients from these regressions shed light on catch-up convergence (if they are negative and significant), or divergence (if they are positive and significant).

Figure 2.4.1 contains plots of the tripartite decomposition against output per worker in 1965, along with the fitted regression lines. Table 2.4.3 summarises the estimated coefficients and their significance (based on p-values). Figure 2.4.1 (a) reveals the relationship between the contribution of efficiency to productivity growth and the output per worker in 1965. The KR's results show a no clear pattern of the contribution of efficiency to productivity growth, with many negative as well as positive changes. The regression slope coefficient is not statistically significant. On the other hand, we find the coefficient is significantly negative at 5% level of significance in the case of SDEA, and at 10% in the case of LLE.

Figure 2.4.1 (b) indicates that technological change makes a greater contribution to productivity growth in most countries. We also find technological regress for low-income countries and larger-than-average contributions to growth for most high-income countries. The coefficients of the least-squares regressions are positive and statistically significant in the cases of DEA and SDEA, but not in the case of LLE. The positive regression slope coefficients obtained from DEA and SDEA suggest that relatively wealthy countries have benefited more from technological progress than have less developed countries.

Figure 2.4.1 (c), on the other hand, suggests that the pattern of productivity growth attributable to capital accumulation is remarkably similar to the overall pattern of changes in labour productivity. For capital deepening, our results do not suggest that it contributes to convergence of income per worker across the sample as found in the KR's study. We find instead that the coefficients are insignificantly negative (positive) in SDEA (LLE). The insignificant results should not be interpreted as capital deepening not contributing to labour productivity growth. In fact, the results in Section 2.4.1 indicate that capital deepening, on average, is the primary source in the mean growth of labour productivity.

Table 2.4.3: Growth regressions of the percentage change in output per worker 1965 and the three decompositions

Methods	**DEA (i)**	**SDEA (ii)**	**LLE (iii)**
Dependent Variables	**Coefficient (p-value)**	**Coefficient (p-value)**	**Coefficient (p-value)**
$(\text{EFF} - 1) \times 100$	0.09	-0.87	-0.49
	(0.81)	(0.01)	(0.06)
$(\text{TECH} - 1) \times 100$	0.77	0.60	0.05
	(0.05)	(0.01)	(0.65)
$(\text{KACCUM} - 1) \times 100$	-2.37	-0.83	0.26
	(0.00)	(0.46)	(0.83)

The independent variable is output per worker in 1965.

p-values are in parentheses.

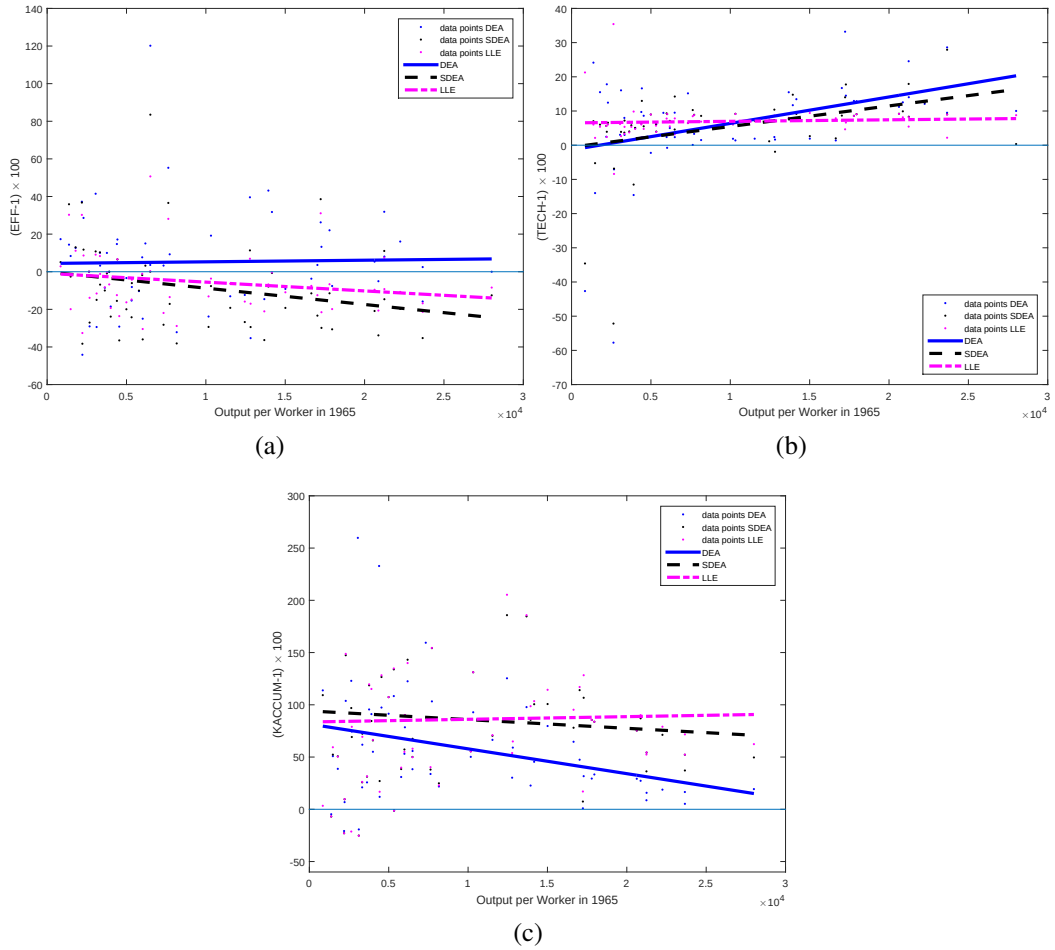


Figure 2.4.1: The tripartite decomposition plotted against output per worker in 1965

2.5 Conclusions

This chapter first generalised the models of Kumbhakar *et al.* (2007) and Park *et al.* (2015) to measure technological change nonparametrically. The methods (SDEA and LLE) used in this paper might be appealing to some economic growth researchers. The nonparametric stochastic methods provide an alternative tool to construct the worldwide production frontier in a set-up where no assumption about the functional form of the frontier is needed and statistical noise is modelled in the frontier model. To the best of our knowledge, this is the first application of LLE in the context of economic growth and convergence.

We apply the nonparametric stochastic frontier model to the data used in KR to investigate

the robustness of the KR's conclusions. We find that the nonparametric stochastic methods (which account for noise and do not assume a parametric form for the frontier) confirm some of the findings as well as providing additional insights to understanding the evolution of labour productivity of countries. In particular, we confirm that capital deepening made the largest contribution to labour productivity growth since 1965 to 1990 and was also the main factor driving the transformation of the distribution of labour productivity from a unimodal to a bi-modal distribution during this period. A message can taken from this finding is that to improve labour productivity growth, most of the countries were better off in investing in capital such as equipment, machines, building good infrastructure. However, the new methods suggest that the magnitude of capital deepening is underestimated when statistical noise is not accounted (i.e. in standard DEA applied in KR). The methods do not confirm the hypothesis that capital deepening is a significant factor for unconditional beta-convergence, which was found in the KR study.

2.A Appendix A: Figures and Tables

Table 2.A.1: Percentage change of labour productivity and its tripartite decomposition from 1965 to 1990 using DEA, SDEA and LLE

Methods		**DEA(i)**			**SDEA(ii)**			**LLE(iii)**		
Countries	**Labour Productivity Change**	ΔEFF	ΔTECH	ΔKACCUM	ΔEFF	ΔTECH	ΔKACCUM	ΔEFF	ΔTECH	ΔKACCUM
Argentina	4.59	-35.30	1.64	59.04	-29.41	-1.91	51.05	-18.33	8.05	66.61
Australia	42.67	8.06	14.03	15.78	-14.62	8.33	54.26	-9.34	8.35	55.15
Austria	95.15	-14.57	15.49	97.79	-36.34	7.69	184.64	-21.03	7.94	185.67
Belgium	78.36	22.04	12.94	29.41	-11.46	8.58	85.52	-6.65	8.59	86.22
Bolivia	32.71	-18.50	5.02	55.05	-23.86	5.06	65.91	-19.59	4.74	65.81
Canada	54.55	16.06	12.09	18.80	-19.76	12.46	71.27	-11.04	7.87	80.15
Chile	16.57	-23.82	1.88	50.19	-29.35	6.13	55.46	-15.18	7.50	57.58
Colombia	68.78	7.66	2.38	53.12	-1.80	9.31	57.23	-0.09	8.77	56.64
Denmark	39.08	-7.58	12.86	33.33	-30.61	9.08	83.76	-19.72	8.95	84.20
Dominican Rep.	51.81	-29.18	8.61	97.38	-36.51	5.53	126.59	-23.83	5.01	127.37
Ecuador	80.89	-3.35	-2.24	91.46	-19.96	8.98	107.37	-15.54	8.50	106.33
Finland	96.23	43.15	11.70	22.71	-7.87	14.76	85.60	-7.07	7.55	100.15
France	78.29	3.59	16.72	47.45	-23.34	8.65	114.04	-12.55	8.56	114.64
Germany	70.75	13.24	14.49	31.70	-29.86	17.76	106.72	-22.07	7.05	130.41
Greece	129.46	9.26	3.31	103.28	-17.11	8.89	154.23	-12.91	8.46	153.32
Guatemala	28.54	-10.30	9.46	30.91	-10.03	3.14	38.52	-3.91	3.37	39.67
Honduras	22.87	-8.52	6.81	25.75	-9.98	4.02	31.21	-8.23	3.86	31.35
Hong Kong	251.08	120.17	2.36	55.79	83.51	14.23	67.48	46.64	8.74	59.59
Iceland	66.42	-9.13	1.92	79.69	-19.24	2.65	100.73	-10.66	9.01	113.59
India	80.52	12.66	15.51	38.72	12.99	6.02	50.70	9.77	6.28	51.38
Ireland	133.08	19.18	1.41	92.85	-7.63	9.18	131.12	-3.24	8.73	130.22
Israel	86.13	39.58	2.39	30.24	11.34	10.40	51.42	7.19	9.87	51.36
Italy	117.45	31.76	13.45	45.46	-0.68	9.16	100.57	-0.34	8.97	100.80
Ivory Coast	15.00	-29.05	-7.07	74.42	-27.04	-6.83	69.18	-14.73	-6.37	80.52
Jamaica	-3.56	-8.13	6.12	-1.08	-6.06	3.99	-1.28	-4.26	3.93	-1.28
Japan	208.52	3.27	15.16	159.42	-28.19	6.49	303.45	-21.80	6.58	302.67
Kenya	35.29	14.40	24.14	-4.73	35.88	7.02	-6.96	29.29	7.12	-7.01
Luxembourg	78.47	31.88	24.55	8.65	11.04	17.92	36.29	6.27	6.21	53.85
Madagascar	-29.68	-44.14	17.81	6.85	-38.28	3.93	9.62	-33.07	3.35	9.77
Malawi	43.85	17.32	-42.65	113.81	5.09	-34.59	109.25	-55.09	48.91	33.81

Table 2.A.1 Continued:

Mauritius	56.99	3.60	9.51	38.37	0.00	4.60	50.09	0.27	3.52	49.60
Mexico	47.47	-13.06	1.91	66.44	-19.21	7.16	70.34	-8.49	8.47	72.59
Morocco	52.89	17.14	16.62	11.92	6.55	12.95	27.03	6.47	3.15	16.85
Netherlands	51.45	5.38	11.21	29.24	-20.88	9.20	75.29	-9.52	9.05	75.71
New Zealand	7.42	-15.82	9.49	16.54	-35.27	9.02	52.22	-21.24	8.95	52.69
Nigeria	40.58	8.36	-14.00	50.85	-2.59	-5.25	52.31	-5.96	-5.02	60.06
Norway	69.72	26.25	33.20	0.92	38.58	14.00	7.43	29.55	5.60	17.54
Panama	32.87	-24.98	-0.77	78.48	-35.97	9.10	90.21	-29.83	8.61	89.30
Paraguay	63.25	0.00	-14.57	91.08	0.00	-11.52	84.49	5.75	1.31	100.90
Peru	-16.11	-32.15	1.51	21.80	-38.12	8.57	24.86	-30.47	8.05	24.59
Philippines	43.84	10.08	7.98	21.01	10.29	3.67	25.79	8.23	3.51	26.03
Portugal	168.82	15.04	5.02	122.51	3.22	7.11	143.15	2.13	6.75	141.73
Sierra Leone	-5.80	0.00	-57.74	122.92	0.00	-52.16	96.91	0.00	20.16	-5.38
South Korea	424.45	41.47	3.04	259.76	10.80	6.78	343.30	9.73	6.34	341.60
Spain	111.74	-12.35	7.19	125.37	-26.75	1.15	185.81	-15.65	7.85	204.00
Sri Lanka	72.07	3.23	3.00	61.82	-6.85	7.20	72.30	-4.20	6.82	71.58
Sweden	36.03	-5.04	12.54	27.29	-33.84	9.91	87.06	-20.79	8.29	90.97
Switzerland	38.68	2.48	28.61	5.23	-20.95	27.92	37.14	-18.86	3.57	74.92
Syria	107.90	55.26	0.08	33.80	36.58	10.28	38.03	28.64	9.70	37.73
Taiwan	318.96	14.73	9.71	232.84	-15.50	6.35	366.20	-11.91	6.18	363.73
Thailand	194.68	28.61	12.45	103.76	11.75	6.57	147.43	8.21	6.44	148.96
Turkey	129.27	9.99	6.59	95.56	-0.92	5.83	118.66	0.17	5.37	118.42
United Kingdom	60.74	-3.70	1.39	64.63	-11.50	2.00	78.05	-7.32	9.33	91.31
United States	31.29	0.00	10.02	19.33	-12.55	0.34	49.62	-8.49	8.80	62.86
Yugoslavia	88.10	-15.32	6.62	108.35	-24.22	6.13	133.88	-14.40	5.68	133.38
Zambia	-33.86	-29.40	16.04	-19.27	-14.96	3.90	-25.14	-12.04	3.45	-25.42
Zimbabwe	11.38	37.24	2.47	-20.80	36.77	5.64	-22.92	30.15	5.39	-22.77
Mean (over 57 cpuntries)	**75.06**	**5.24**	**6.17**	**58.47**	**-8.50**	**5.31**	**86.04**	**-6.33**	**7.36**	**86.56**
std.err of the mean	**(10.48)**	**(3.50)**	**(1.86)**	**(7.19)**	**(3.07)**	**(1.48)**	**(10.05)**	**(2.29)**	**(0.88)**	**(10.17)**

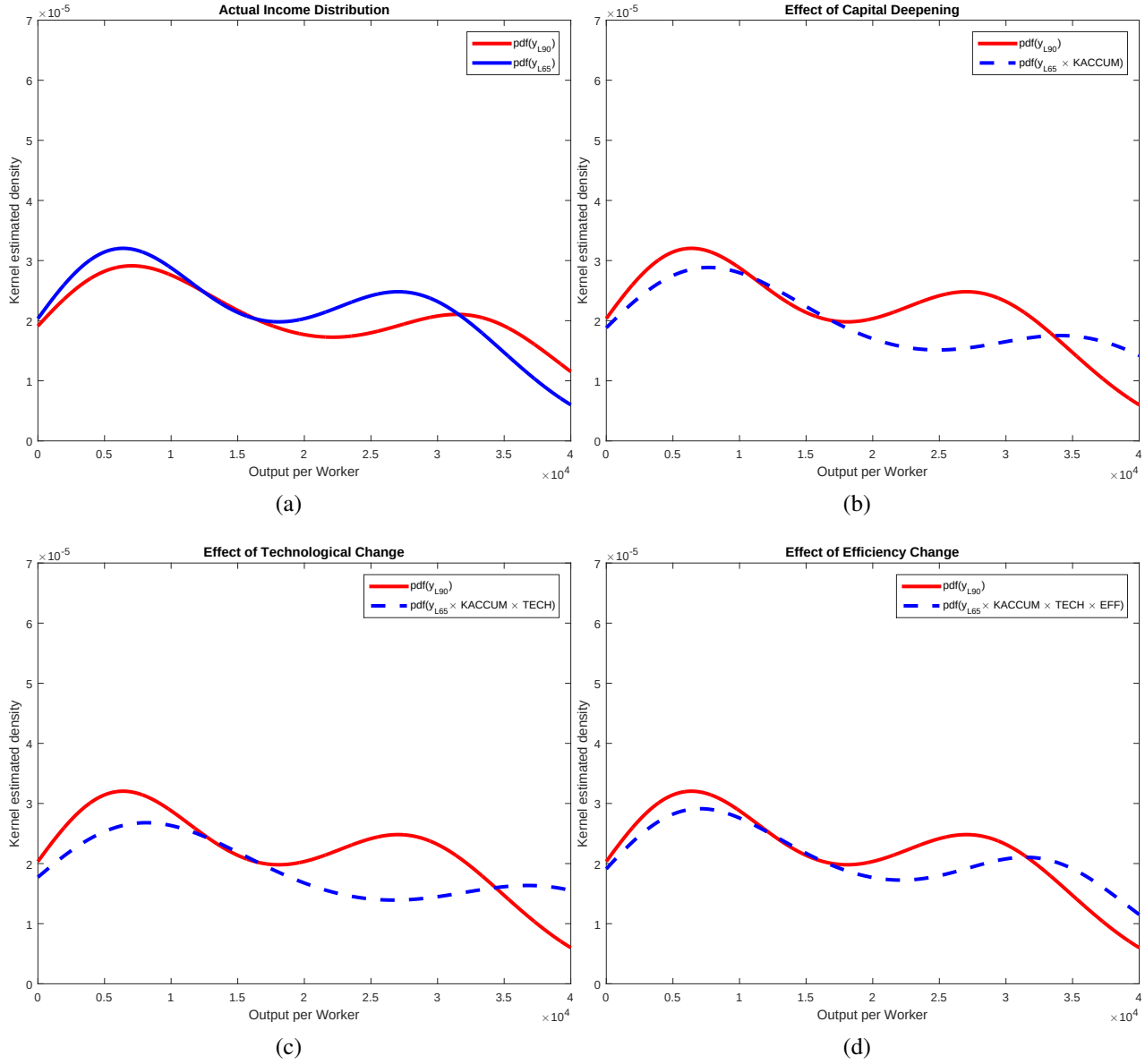


Figure 2.A.3: The counter-factual distributions of output per worker are constructed following the order KACCUM-TECH-EFF using LLE

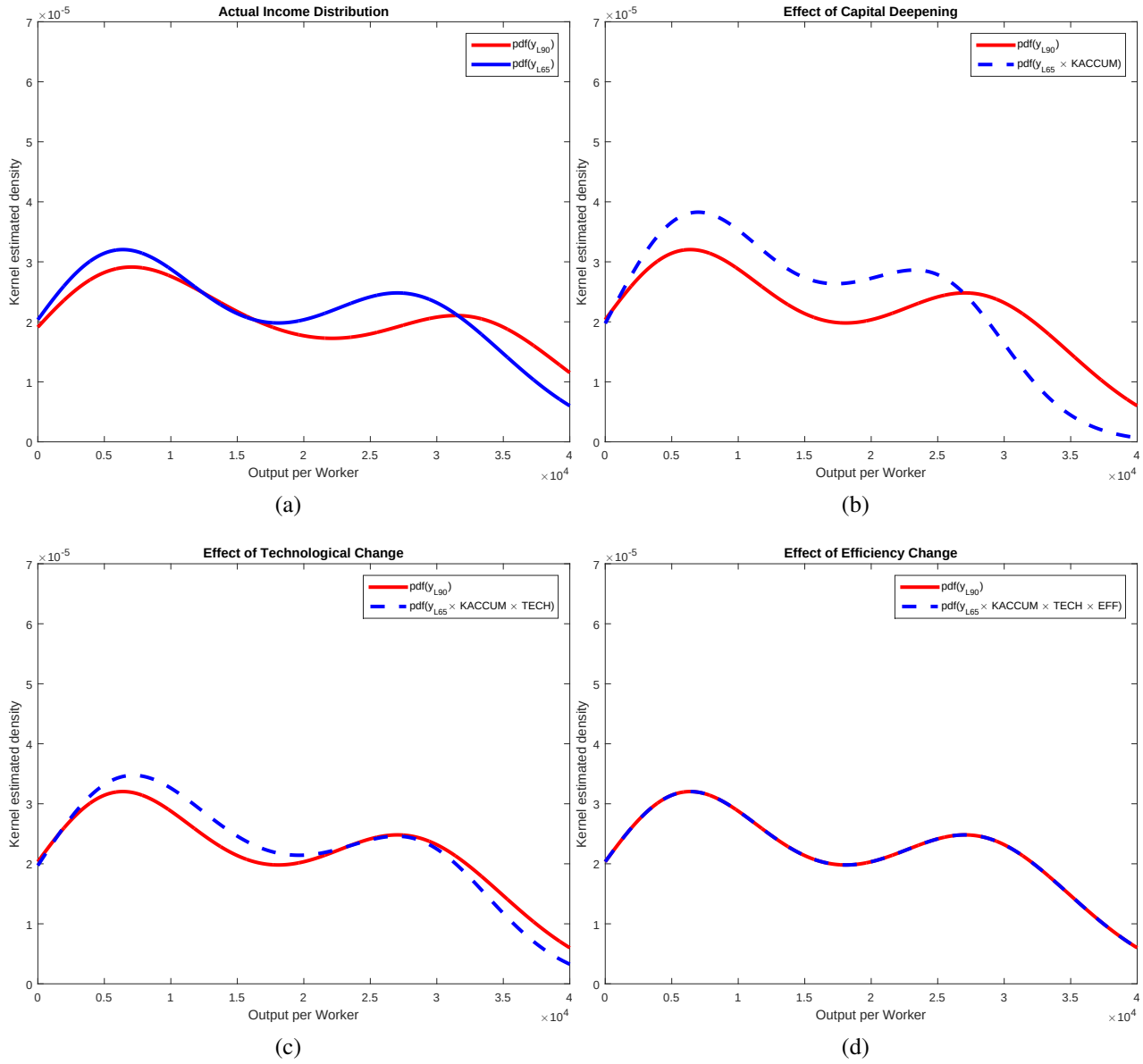


Figure 2.A.1: The counter-factual distributions of output per worker are constructed following the order KACCUM-TECH-EFF using DEA

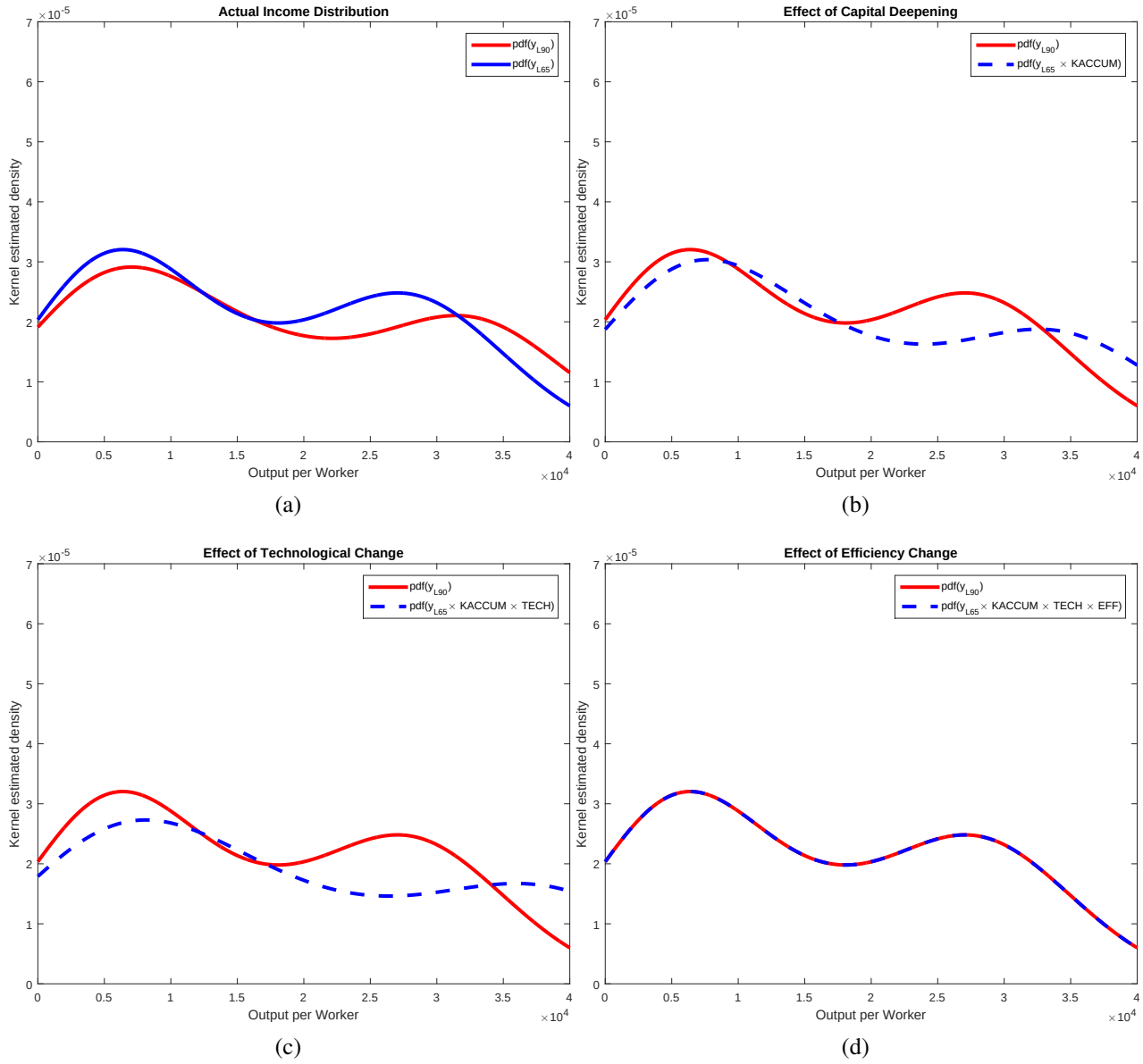


Figure 2.A.2: The counter-factual distributions of output per worker are constructed following the order KACCUM-TECH-EFF using SDEA

CHAPTER 3

Non-linear Time-varying Stochastic Frontier Model: Assessing Determinants of Output Growth

3.1 Introduction

In the previous chapter labour productivity growth was decomposed into three components: capital deepening, technological change and (technical) efficiency change.[1] In this chapter, we are interested in explaining dynamic variations of efficiencies in terms of observable variables. Chapter 3 aims to address a number of questions such as: What are the factors driving an inefficiency of a country? Are the effects constant or varying over time? Do the factors only affect an inefficiency component or a production function or both? Such a study helps us to better understand about the drivers of inefficiencies, and under what channels the factors affect the production process: an efficiency component and/or a production function. In this chapter, we extend the static model specification of Wang and Schmidt (2002), Alvarez *et al.* (2006) to a dynamic case. We aim to capture the dynamic pattern of technical efficiencies and the effect of the factors on production process over time. We also provide a formal hypothesis test to examine if the dynamic specification is needed.

When explaining variations of efficiencies, researchers often face an issue of where factors influencing efficiencies (often known as "environmental variables" in this literature) should be incorporated into a production process: through an inefficiency term, or through a production frontier, or through both. The majority of studies in the early literature on productivity and efficiency often considered the effects of environmental variables on an inefficiency component of

[1] In this chapter, a technical (in)efficiency is referred to as a (in)efficiency, unless stated otherwise.

a frontier model.[2] Such frontier models are estimated by either a two-stage DEA or a one-stage stochastic frontier approach. A two-stage DEA involves two stages of estimation. The first is to estimate the frontier and the resulting efficiencies using DEA. The second is to regress the estimated efficiencies on the environmental variables using an appropriate limited dependent variable model (e.g. a truncated normal regression). To ensure the valid inferences for inefficiencies, a two stage DEA requires a "separability" assumption which implies that the support of the output variables does not depend on the environmental variables (Simar and Wilson (2007), Simar and Wilson (2011)). A one-stage stochastic frontier approach is often preferred in this literature as it allows for potential relationships between the environmental variables and the inputs. Ignoring the potential relationships result in unbiased and inconsistent estimators (see Kumbhakar *et al.* (1991), Reifschneider and Stevenson (1991), Battese and Coelli (1995), Huang and Liu (1994), and many others).

O'Donnell (2015) recently develops a theoretical framework suggesting that incorporating the environmental variables into a production function implies the environmental variables can affect both a production function and an inefficiency component. His model specification, however, does not provide an economic explanation regarding variations of inefficiencies, e.g. how much inefficiencies change for a given change in an environmental variable. Also the estimators are not consistent unless valid instruments are used (O'Donnell (2015)). This is due to an endogeneity issue, i.e. a one-side random error term is implicitly a function of inputs, outputs, and environment variables in his model specification.

Recent studies incorporate the effects of the environmental variables to both a production frontier and an inefficiency component of a production process. A conditional nonparametric frontier is one of the approaches to estimate such frontier models (see Daraio and Simar (2005), Mastromarco and Simar (2015)). The approach does not require a particular assumption about a production function. However, it ignores the presence of statistical noise in the estimation of the production frontier. Simar *et al.* (2017) overcome this problem using a local least squares estimator. Obtaining statistical inferences (e.g. standard errors, confidence intervals) for such a nonparametric stochastic model is not a trivial task. A proper bootstrap algorithm is required to obtain statistical inference in their framework. Modelling and testing time varying in inefficiency are still challenging in their framework. An alternative approach is the stochastic parametric frontier model with the scaling property (Simar *et al.* (1994), Wang and Schmidt (2002), Alvarez *et al.* (2006)). This scaling feature implies that the distribution of inefficiency is multiplicatively decomposed into a function of the determinants of inefficiency and a random

[2]For a comprehensive survey about modelling the effects of environmental variables on inefficiencies, the reader is referred to Kumbhakar and Lovell (2000).

variable (known as the *base inefficiency*) whose distribution is independent of the environmental variables. Statistical inferences for such a stochastic frontier model are directly obtained. However, a particular functional form of a production function is required.

We contribute to the line research of using the stochastic parametric frontier model with the scaling property to capture the effects of environmental variables on a production process in the following aspects. Building upon the model specification considered in Wang and Schmidt (2002) and Alvarez *et al.* (2006), we extend their static models to a dynamic case. We believe that in practice it is sensible to assume that the base inefficiencies are not identically distributed across decision making units (DMUs) (e.g. firms, organisations, countries, etc.), and that DMUs have incentives to improve their inefficiencies over time. For example, to stay in a competitive industry firms often have incentives to improve their own inefficiencies. The adjustment of inefficiencies can vary across DMUs due to heterogeneity factors such as the average quality of a firm's management and workers. To accommodate these features, we allow the coefficients associated with the base inefficiencies and the environmental variable to vary over time and across individuals. We also provide a formal test for time variation in technical inefficiency.

Our time-varying specification of inefficiency is closely related that of Tsionas (2006) who allows the logarithm of base inefficiencies to follow an autoregressive process of first order.[3] Deviating from Tsionas (2006), we do not restrict the dynamic pattern of the base inefficiencies to be identical across individuals, and the effects of environmental variables on inefficiency to be time-invariant and identical across individuals.

Both Tsionas (2006) model and ours have the non-linear feature in technical inefficiency. To handle the non-linearity issue Tsionas (2006) uses a Metropolis-Hasting random walk algorithm for each observation at each time period. Sampling the time-varying parameters at each time period is well-known for slow convergence to a posterior distribution and far less efficient (e.g. Carter and Kohn (1994)). Here we adapt the idea of using Gaussian approximation proposed by Chan and Strachan (2014) to obtain the location (e.g. the mean and the variance) of the proposed density, and then use accept reject Metropolis-Hasting to decide whether the draws are accepted. We sample all the time-varying parameters at once using the precision-based algorithm Chan and Jeliazkov (2009) to improve the efficiency of estimating the non-linear parameters. The algorithm utilises the sparse structure of the matrices and therefore the computation time is improved.

[3] Ahn and Sickles (2000), Emvalomatis *et al.* (2011) also consider autoregressive process of inefficiencies in their model specifications. However, their studies do not consider the effects of environmental variables on technical inefficiency.

To test for time-varying pattern in the base efficiencies, we first re-parametrise our model specification in the form of non-centred parametrisation following Frühwirth-Schnatter and Wagner (2010). We then adapt the idea of using Savage-Dickey ratio in Chan (2016) to perform the test for time variation in parameters. To the best of our knowledge, this is the first study providing a formal test for time variation in inefficiencies.

We also allow the coefficients associated with a production function to be time-varying, and show that such a specification encompassing the commonly used production functions (e.g. Cobb-Douglas, translog) with linear trend. The time-varying parameters in a production/cost function have recently considered in a number of studies such as Jin and Jorgenson (2010) and Peyrache and Rambaldi (2012). Peyrache and Rambaldi (2012) show the time-varying parameters in production function provide a flexible way capturing technological change, in comparison to a time-dummy variable approach. However, these studies only focus on measuring technological change, and do not consider the effects of environmental variables to production process.

For the empirical illustration, we apply the proposed model in a study of economic growth of 21 OECD countries over the period 1970-2011 to investigate whether foreign direct investment (FDI) plays a role as influencing the frontier or inefficiency. We also investigate the factors driving output growth using stochastic frontier framework. We find that FDI plays a role as influencing the production frontier rather than inefficiency. We find statistical evidence suggesting inefficiency are time-varying. Capital deepening is the key factor driving output growth.

The rest of this chapter is structured as follows. Section 3.2 presents the proposed model, discusses the underlying assumptions of the model. Section 3.3 discusses Bayesian estimation. Section 3.4 presents Savage-Dickey density ratio test for time-variation and model specification of the stochastic frontier model. Section 3.5 illustrates the empirical study, and Section 3.6 concludes.

3.2 The Model

In this section, we first present the static stochastic frontier model with the scaling property considered in studies of Simar *et al.* (1994), Wang and Schmidt (2002) and Alvarez *et al.* (2006), and then provide our justifications for an extension to a dynamic case.

Consider a panel data where observations are indexed by i ($i = 1, \ldots, N$), and time is indexed

by t ($t = 1, \ldots, T$). Let y_{it} be an output for observation i at period t, x_{it} be a $1 \times k$ vector of an intercept and variables determining the production frontier, and z^*_{it} be a $1 \times r$ vector of variables influencing inefficiencies. The stochastic frontier model with the scaling property in Simar *et al.* (1994), Wang and Schmidt (2002), and Alvarez *et al.* (2006) has the following specification:

$$y_{it} = x_{it}\beta - u_{it} + v_{it}, \tag{3.2.1}$$

$$u_{it} = \eta_i \exp(z^*_{it}\gamma^*), \tag{3.2.2}$$

where the random noise v_{it} is normally distributed with mean zero and constant variance, $v_{it} \sim N(0, \sigma^2)$, and is independent of x_{it} and z^*_{it}. The random error u_{it}, representing for inefficiency, has a scaling feature as described in (3.2.2), i.e. it is multiplicatively decomposed into a function of the determinants of inefficiency and a random variable η_i.[4] The variable η_i is known as the base inefficiency level which can be thought of measuring things such as a manager's natural skills. How well the natural skills are exploited to improve the firm's inefficiency depends on other factors (z^*_{it}) such as the manager's education, experience or an environment in which a firm operates. A change in z^*_{it} changes only the scale, but not the shape of the base inefficiency level η_i. An appealing feature of the scaling property specification is that it provides a simple interpretation relating to an effect of z^*_{it} on a firm's inefficiency level (see Alvarez *et al.* (2006)). For instance, suppose the environmental variables z^*_{it} are in logarithm, the coefficient $\gamma^{(j)}$, where $j = 1, \ldots, r$, measures the percentage change in inefficiency given 1% change in $z^{*(j)}_{it}$, i.e. $\gamma^{(j)} = \partial \ln(u_{it}) / \partial \ln z^{*(j)}_{it})$.

In the specification (3.2.2), the base inefficiency η_i is assumed to be identically distributed across firms, and the effects of the environmental variables are homogeneous across DMUs and constant over time. However, in practice the base inefficiencies might not be identically distributed across DMUs. This could be due to heterogeneity in the average quality of a firm's management. Also, to stay in a competitive industry a firm most likely learns from its own mistakes, and aims to improve its inefficiency over time. To accommodate these features, we extend the static model to the dynamic stochastic frontier. We allow the base inefficiency η_i and the coefficients associated with the environmental variables (γ^*) to vary across individuals and over time. We further allow the coefficients associated with the production function to be time-varying, and show that such a time-varying production function encompasses the commonly used production functions (e.g. Cobb-Douglas, translog) with a linear trend. The proposed

[4]The models of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill *et al.* (1995) can also be presented as a stochastic frontier model with the scaling property.

model is[5]

$$
\begin{aligned}
y_{it} &= x_{it}\beta_t - u_{it} + v_{it}, & v_{it} &\sim N(0, \sigma^2_{v_i}), && (3.2.3)\\
u_{it} &= \exp(z_{it}\gamma_{it}), & & && (3.2.4)\\
\beta_t &= \beta_{t-1} + b_{t-1} + \varepsilon_{\beta_t}, & \varepsilon_{\beta_t} &\sim N(0, \mathbf{w}^2_{\beta}), && (3.2.5)\\
b_t &= b_{t-1} + \varepsilon_{b_t}, & \varepsilon_{b_t} &\sim N(0, \mathbf{w}^2_{b}), && (3.2.6)\\
\gamma_{it} &= \gamma_{it-1} + \varepsilon_{\gamma_{it}}, & \varepsilon_{\gamma_{it}} &\sim N(0, \mathbf{w}^2_{\gamma_i}), && (3.2.7)
\end{aligned}
$$

where $z_{it} = [1\ z^*_{it}]$, $\mathbf{w}^2_{\beta} = \text{diag}(\omega^2_{\beta}) = \text{diag}(\omega^2_{\beta(1)}, \ldots, \omega^2_{\beta(k)})$, $\mathbf{w}^2_{b} = \text{diag}(\omega^2_{b}) = \text{diag}(\omega^2_{b(1)}, \ldots, \omega^2_{b(k)})$, $\mathbf{w}^2_{\gamma_i} = \text{diag}(\omega^2_{\gamma_i}) = \text{diag}(\omega^2_{\gamma_i(1)}, \ldots, \omega^2_{\gamma_i(r+1)})$. A statistical noise v_{it} is normally distributed with mean zero and variance $\sigma^2_{v_i}$ varying across individuals, $v_{it} \sim N(0, \sigma^2_{v_i})$, and is independent of inputs x_{it} and z_{it}. The state equations (3.2.5)-(3.2.7) are initialized with $\beta_0 \sim N(\underline{\beta}, \underline{\mathbf{P}^{-1}_{\beta}})$, $b_0 \sim N(\underline{b}, \underline{\mathbf{P}^{-1}_{b}})$, $\gamma_{i0} \sim N(\underline{\gamma_i}, \underline{\mathbf{P}^{-1}_{\gamma_i}})$, where $\underline{\mathbf{P}_{\beta}}$, $\underline{\mathbf{P}_{b}}$, $\underline{\mathbf{P}_{\gamma_i}}$ are precision matrices which are the inverse of the covariance-variance matrix.

Our specification of inefficiency (3.2.4) is equivalent to that of Wang and Schmidt (2002), and Alvarez *et al.* (2006) if further restrictions are imposed on γ_{it}. Equation (3.2.4) can be written as

$$u_{it} = \exp(z_{it}\gamma_{it}) = \exp(\gamma^{(1)}_{it} + z^*_{it}\gamma^*_{it}) = \exp(\gamma^{(1)}_{it})\exp(z^*_{it}\gamma^*_{it}) = \eta_{it}\exp(z^*_{it}\gamma^*_{it}), \qquad (3.2.8)$$

where $\gamma_{it} = [\gamma^{(1)}_{it}\ \ \gamma^*_{it}]'$. If one is willing to assume $\gamma^{(1)}_{it}$ to be time invariant, and γ^*_{it} to be time invariant and homogeneous across individuals, then (3.2.4) has the specification of Simar *et al.* (1994), Wang and Schmidt (2002). In our specification the basic technical efficiency η_{it} is assumed to be a log normal distribution i.e. $\ln(\eta_{it}) \sim N(\underline{\eta_i}, \underline{\sigma^2_{\eta_i}})$ while the basic efficiency is often assumed to be a truncated normal distribution in the previous studies.

We now turn to the discussion about the motivations underpinning the time-varying specifications (3.2.5)-(3.2.7). We first discuss the specification in technical inefficiency and then the specification in the production function. For an inefficiency component we allow the coefficient γ_{it} following a random walk process to capture the high persistence in the movement of the inefficiencies (equation (3.2.7)). The high persistence in the adjustment of inefficiencies could be due to the potentially large adjustment costs associated with the inputs, union work rules, or persistent change in management skills. Therefore, unless there are changes affecting the operating environment (e.g. change in government regulation) or change in a firm ownership, the technical inefficiencies are likely persistent. The unexpected changes are captured by a random

[5] We use bold notations to refer to a matrix, and non-bold notations to a vector in Chapter 3.

error $\varepsilon_{\gamma_{it}}$ in our model. Further, given the random walk specification we are able to test for time-variation in inefficiencies.[6] To the best of our knowledge, this is the first study providing a test for persistence in inefficiencies. The details of the time-varying test is discussed in Section 4.

To capture the dynamic patterns of the base inefficiencies, there are a number of specifications proposed in the literature of productivity and efficiency. A common approach is to include a deterministic trend (i.e. linear or quadratic trends) (see Cornwell *et al.* (1990), Kumbhakar (1990), Battese and Coelli (1992) and many others). Although the studies might provide approximation for the dynamic of inefficiencies, the specifications are arbitrary approximations with little theoretical justifications. Many researchers aim to provide some economic interpretations regarding the dynamic patterns using an AR(1) process (see Ahn and Sickles (2000), Emvalomatis *et al.* (2011) and Tsionas (2006)). Ahn and Sickles (2000) propose an AR(1) of an technical inefficiency as follows $u_{it} = \rho u_{it-1} + \zeta_{it}$, where $0 < \rho < 1$ measures a firm's ability to adjust its past period inefficiency level. Ahn and Sickles (2000) and Emvalomatis *et al.* (2011) models are not designed for incorporating the variables affecting inefficiency, while Tsionas (2006) does. Our model specification differs from Tsionas's (2006) in the sense that the base inefficiencies are not identical across DMUs, and the effects of environmental variables are not time-invariant and the same across DMUs.

We also allow for time variation in the coefficients associated with the production function. Time-varying specification for a production function has been recently considered in some studies such as Jin and Jorgenson (2010) and Peyrache and Rambaldi (2012). Peyrache and Rambaldi (2012) allow the coefficients associated with a production function to follow a random walk, and focus on measuring technological change rather than explaining variations of inefficiencies, which is our interest of this chapter.

Deviating from their specification, we allow the coefficients associated with variables x_{it} to follow a random walk with a drift as described in (3.2.5)-(3.2.6). Such a specification (3.2.5)-(3.2.6) provides a flexible form for a production function. It encompasses a Cobb-Douglas and a translog function with a linear trend. To illustrate this, we re-parametrise the model (3.2.3)-(3.2.7) to a non-centred reparametrisation specification proposed by Frühwirth-Schnatter and Wagner (2010). We define $\tilde{\beta}_t^{(j)} = (\beta_t^{(j)} - \beta_0^{(j)})/\omega_{\beta^{(j)}}$, $\tilde{b}_t^{(j)} = (b_t^{(j)} - b_0^{(j)})/\omega_{b^{(j)}}$, $\tilde{\gamma}_{it}^{(l)} = (\gamma_{it}^{(l)} - \gamma_{i0}^{l})/\omega_{\gamma_i^{(l)}}$, where $j = 1, \ldots, k$; $l = 1, \ldots, m$, then the non-centred specification of the model

[6] If one allows inefficiencies following an autoregressive process, then the framework discussed in this Chapter cannot be used to test for time variation in inefficiencies.

(3.2.3)-(3.2.7) is

$$
\begin{aligned}
y_{it} &= x_{it}\beta_0 + x_{it}tb_0 + x_{it}\mathbf{w}_\beta\tilde{\beta}_t + x_{it}\mathbf{w}_b\tilde{B}_t - \exp(z_{it}\gamma_{i0} + z_{it}\mathbf{w}_{\gamma_i}\tilde{\gamma}_{it}) + v_{it}, && (3.2.9)\\
\tilde{\beta}_t &= \tilde{\beta}_{t-1} + \xi_{\beta_t} \qquad \xi_{\beta_t} \sim N(0, \mathbf{I}_k), && (3.2.10)\\
\tilde{b}_t &= \tilde{b}_{t-1} + \xi_{b_t} \qquad \xi_{b_t} \sim N(0, \mathbf{I}_k), && (3.2.11)\\
\tilde{B}_t &= \tilde{B}_{t-1} + \tilde{b}_{t-1}, && (3.2.12)\\
\tilde{\gamma}_{it} &= \tilde{\gamma}_{it-1} + \xi_{\gamma_{it}} \qquad \xi_{\gamma_{it}} \sim N(0, \mathbf{I}_m). && (3.2.13)
\end{aligned}
$$

Without loss of generality, we consider a simple example where the vector x_{it} includes an intercept, two inputs, the cross terms of the inputs, and the squared terms of inputs, i.e. $x^*_{it} = [1\ x^{(1)}_{it}\ x^{(2)}_{it}\ x^{(1)}_{it}x^{(2)}_{it}\ \ x^{2(1)}_{it}\ x^{2(2)}_{it}]$. If

- there is no time-variation in β_t, implying $\mathbf{w}_\beta = \mathbf{w}_b = \mathbf{0}$ in (3.2.9), the production function (3.2.9) has the form $f(x_{it}) = x_{it}\beta_0 + x_{it}tb_0$, which is expanded as $f(x_{it}) = \beta_0 + b_0 t + \beta^{(1)}_0 x^{(1)}_{it} + b^{(1)}_0 t x^{(1)}_{it} + \beta^{(2)}_0 x^{(2)}_{it} + b^{(2)}_0 t x^{(2)}_{it} + \beta^{(3)}_0 x^{(1)}_{it}x^{(2)}_{it} + b^{(3)}_0 t x^{(1)}_{it}x^{(2)}_{it} + \beta^{(4)}_0 x^{2(1)}_{it} + b^{(4)}_0 t x^{2(1)}_{it} + \beta^{(5)}_0 x^{2(2)}_{it} + b^{(5)}_0 t x^{2(2)}_{it}$.
- the coefficients $b^{(1)}_0, b^{(2)}_0, b^{(3)}_0, b^{(4)}_0$ and $b^{(5)}_0$ are further restricted to zero, the model becomes a translog function, i.e. $f(x_{it}) = \beta_0 + b_0 t + \beta^{(1)}_0 x^{(1)}_{it} + \beta^{(2)}_0 x^{(2)}_{it} + \beta^{(3)}_0 x^{(1)}_{it}x^{(2)}_{it} + \beta^{(4)}_0 x^{2(1)}_{it} + \beta^{(5)}_0 x^{2(2)}_{it}$;
- the coefficients $\beta^{(3)}_0, \beta^{(4)}_0$ and $\beta^{(5)}_0$ are further restricted to 0, the production function is Cobb-Douglas, i.e. $f(x_{it}) = \beta_0 + b_0 t + \beta^{(1)}_0 x^{(1)}_{it} + \beta^{(2)}_0 x^{(2)}_{it}$.

Ignoring the effects of heteroskedasticity in inefficiency and/or a symmetric error component can affect inferences about the parameters of the stochastic frontier as well as inefficiencies (see Kumbhakar and Lovell (2000, pp.116-130)). To capture the possibility of heteroskedasticity in the evolution of individual inefficiencies, we allow for the variance of innovations ε_{γ_i} to differ across observations. A similar feature is also applied for v_{it} to capture heterokedasticity in the stochastic frontier model.

An advantage of using Bayesian in the estimation of the model is that restrictions on returns to scale, marginal labour productivity, marginal capital productivity are straightforward to be imposed.

3.3 Bayesian Estimation

In this section we first elicit the priors for the parameters in the model (3.2.9)-(3.2.13), and then provide their conditional posterior densities.

We choose an inverse Gamma prior for the variance $\omega_{v_i}^2$, i.e. $p(\omega_{v_i}^2) \sim IG(\tau_i, S_i)$, $i = 1, \ldots, n$, IG denotes inverse gamma. For the standard deviation ω_β, ω_b and ω_γ, we choose normal priors, i.e. $p(\omega_\beta) \sim N(0, \underline{\mathbf{P}_{\omega_\beta}^{-1}})$, $p(\omega_b) \sim N(0, \underline{\mathbf{P}_{\omega_b}^{-1}})$, $p(\omega_\gamma) \sim N(0, \underline{\mathbf{P}_{\omega_\gamma}^{-1}})$. The choice of normal priors for the standard deviations are further discussed in Section 3.4. For the coefficients β_0, b_0 and γ_{i0} we also choose normal priors, i.e., $\beta_0 \sim N(\underline{\beta}, \underline{\mathbf{P}_\beta^{-1}})$, $b_0 \sim N(\underline{b}, \underline{\mathbf{P}_b^{-1}})$, $\gamma_{i0} \sim N(\underline{\gamma}_i, \underline{\mathbf{P}_{\gamma_i}^{-1}})$.

To facilitate further discussion we group the time invariant parameters (β_0, b_0), and the time-varying parameters $(\tilde{\beta}_t, \tilde{b}_t, \tilde{B}_t)$ into two vectors, δ_0 and $\tilde{\delta}_t$ respectively, i.e. $\delta_0 = [\beta_0', b_0']'$, $\tilde{\delta}_t = [\tilde{\beta}'_t, \tilde{b}'_t, \tilde{B}'_t]'$. We also stack y_{it}, x_{it}, z_{it}, v_{it}, $\tilde{\delta}_t$, $\tilde{\gamma}_{it}$ for all observations i over t, and rewrite the measurement equation (3.2.9) as

$$Y = \mathbf{X}_{\delta_0}\delta_0 + \mathbf{X}_{\tilde{\delta}}\tilde{\delta} - \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma}) + v, \tag{3.3.1}$$

where Y is a $nT \times 1$ vector of an output, i.e. $Y = [y_{11}, \ldots, y_{n1}, \ldots, y_{1T}, \ldots y_{nT}]'$. $\mathbf{X}_{\delta_0} = [\mathbf{X}\ (\mathbf{X} \odot t)]$ is a $nT \times 2k$ matrix, where $\odot$ represents Hadamard multiplication, and $\mathbf{X}$ is a $nT \times k$ matrix of factors influencing the frontier, i.e. $\mathbf{X} = [x'_{11}, \ldots, x'_{n1}, \ldots, x'_{1T}, \ldots x'_{nT}]'$. The matrix $\mathbf{X}_{\tilde{\delta}} = \begin{bmatrix} [\mathbf{X}_{1\beta}, \mathbf{0}, \mathbf{X}_{1b}] & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & [\mathbf{X}_{2\beta}, \mathbf{0}, \mathbf{X}_{2b}] & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & [\mathbf{X}_{T\beta}, \mathbf{0}, \mathbf{X}_{Tb}] \end{bmatrix}$ has the dimension of $nT \times 3kT$, where $\mathbf{X}_{t\beta} = \begin{bmatrix} x_{1t}\mathbf{w}_\beta \\ x_{2t}\mathbf{w}_\beta \\ \vdots \\ x_{nt}\mathbf{w}_\beta \end{bmatrix}$ with $t = 1, \ldots, T$ and $\mathbf{w}_\beta = \text{diag}(\omega_\beta) = \text{diag}(\omega_{\beta^{(1)}}, \ldots, \omega_{\beta^{(k)}})$. A similar procedure is applied to $\mathbf{X}_{tb}$. $\tilde{\delta} = [\tilde{\delta}'_1, \ldots, \tilde{\delta}'_T]'$ is a $3kT \times 1$ vector of the coefficients associated with $\mathbf{X}_{\tilde{\delta}}$. $\mathbf{Z}$ is a $nT \times nm$ ($m = r + 1$) matrix of factors influencing inefficiency, i.e. $\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_T \end{bmatrix}$ with $Z_t = \begin{bmatrix} z_{1t} & 0 & \ldots & 0 \\ 0 & z_{2t} & \ldots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & z_{nt} \end{bmatrix}$ and $t = 1, \ldots, T$. γ_0 is a $nm \times 1$ vector of the

coefficients associated with $\mathbf{Z}$. The matrix $\mathbf{Z}_{\tilde{\gamma}} = \begin{bmatrix} z_{11}\mathbf{w}_{\gamma_1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & z_{21}\mathbf{w}_{\gamma_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & z_{nT}\mathbf{w}_{\gamma_n} \end{bmatrix}$ has the dimension of $nT \times nmT$. The $nmT \times 1$ vector of $\tilde{\gamma}$ are the coefficients associated with $\mathbf{Z}_{\tilde{\gamma}}$.

The log likelihood function is obtained from the measurement equation (3.3.1):

$$\log p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2) \propto -\frac{1}{2}(Y - \hat{Y})'\mathbf{P}_\mathbf{v}(Y - \hat{Y}), \tag{3.3.2}$$

where $\hat{Y} = \mathbf{X}_{\delta_0}\delta_0 + \mathbf{X}_{\tilde{\delta}}\tilde{\delta} - \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma})$, and $\mathbf{P}_\mathbf{v} = \text{diag}(1/\sigma_{v_1}^2, \dots, 1/\sigma_{v_n}^2)$ is a precision matrix, which is an inverse of a covariance matrix.

Following Bayes' Theorem, the joint posterior density is:

$$\begin{aligned} p(\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2|Y) \quad &\propto p(\delta_0)p(\gamma_0)p(\tilde{\delta})p(\tilde{\gamma})p(\omega_\beta)p(\omega_b)p(\omega_\gamma)p(\omega_v^2) \\ &\quad p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2) \end{aligned} \tag{3.3.3}$$

To sample the parameters, $\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2$ from the joint posterior density we use the following MCMC algorithm

1. Sample δ_0 from $p(\delta_0|Y, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2)$;
2. Sample $\tilde{\delta}$ from $p(\tilde{\delta}|Y, \delta_0, \gamma_0, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2)$;
3. Sample $\tilde{\gamma}$ from $p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2)$;
4. Sample γ_0 from $p(\gamma_0|Y, \delta_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2)$;
5. Sample ω_β from $p(\omega_\beta|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_b, \omega_\gamma, \omega_v^2)$;
6. Sample ω_b from $p(\omega_b|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_\gamma, \omega_v^2)$;
7. Sample ω_γ from $p(\omega_\gamma|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_v^2)$
8. Sample ω_v^2 from $p(\omega_v^2|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \omega_\beta, \omega_b, \omega_\gamma)$.

Sampling the parameters $\delta_0, \tilde{\delta}, \omega_\beta, \omega_b, \omega_v^2$ from their respective conditional posteriors is straightforward and standard as they are linear. The procedure of sampling the linear parameters is the same as that of a conventional linear-Gaussian state space model. However, this is not the

case for $\gamma_0, \tilde{\gamma}, \omega_\gamma$ due to the non-linearity of the parameters. We elaborate the estimation of the parameters of $\gamma_0, \tilde{\gamma}, \omega_\gamma$ here and present the others in Appendix 3.A.

We first focus on time-varying parameters $\tilde{\gamma}$. The state equation (3.2.13) is re-written as [7]

$$\mathbf{K}_{\tilde{\gamma}}\tilde{\gamma} \;=\; \xi \qquad \xi \sim N(\mathbf{0}, \mathbf{I}_{mT}), \tag{3.3.4}$$

where $\mathbf{K}_{\tilde{\gamma}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ -\mathbf{I}_m & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_m & \mathbf{I}_m \end{bmatrix}$. Given the state equation (3.3.4), the log-density of prior is a log Normal prior

$$\log p(\tilde{\boldsymbol{\gamma}}) \propto -\frac{1}{2}\tilde{\boldsymbol{\gamma}}'\mathbf{K}_{\tilde{\gamma}}'\mathbf{K}_{\tilde{\gamma}}\tilde{\boldsymbol{\gamma}} + c_1, \tag{3.3.5}$$

where the constant c_1 is independent of $\tilde{\gamma}$.

The conditional posterior of $\tilde{\gamma}$ is

$$p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \theta) \propto p(\tilde{\gamma})p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta), \tag{3.3.6}$$

where $\theta = [\omega_\beta, \omega_b, \omega_\gamma, \omega_v^2]$, $p(\tilde{\gamma})$ is a prior of $\tilde{\gamma}$; $p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)$ is a likelihood function, and $p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \theta)$ is conditional posterior of interest, which is not a normal distribution.

One of the popular approaches handling the non-standard conditional posterior density for time-varying parameters is a particle filtering (Doucet *et al.* (2001), Doucet and Johansen (2011)). The approach involves sequential importance sampling and bootstrap resampling. Despite its recent advances in the techniques, a particle filter approach still requires an intensive computation, particularly when the dimensions of time-varying parameters are high and the time periods are long. An alternative is based on a fast approximation of the conditional posterior $p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \theta)$ (see Durbin and Koopman (1997), Chan and Strachan (2014)). Durbin and Koopman (1997) use a second order Taylor expansion to approximate the log of the target of the conditional posterior of the state vector. This approximation provides a Gaussian density, where the mean is the mode of the target density and its precision is equivalent to the negative Hessian evaluated at the mode. The Metropolis-Hastings (MH) algorithm is used to decide whether the draws from the proposal Gaussian approximation are accepted or rejected. The rate acceptance of MH is often quite low. Chan and Strachan (2014) build upon the approach and improve the efficiency of the algorithm using the precision-based algorithm rather than Kalman filter-based

[7] This is obtained by moving $\tilde{\gamma}_{it-1}$ to the left hand side of (3.2.13).

methods. The precision-based algorithm exploits the sparseness structure of the precision matrix for the conditional density of state vectors, and the computation therefore speeds up. To improve the acceptance rate of the MH algorithm, they use an accept reject MH (ARMH). Here we adapt Chan and Strachan (2014) approach to our context.

We first obtain the gradient $\mathbf{F}$ and the negative Hessian $\mathbf{G}$ of the log likelihood function $p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)$, and then use the second-order Taylor series approximation to approximate the log likelihood function at a given point $\tilde{\gamma}$, denoted as $\breve{\gamma}^*$. That is,

$$\begin{aligned} \log p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta) &\approx p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)|_{\tilde{\gamma}=\breve{\gamma}^*} + (\tilde{\gamma} - \breve{\gamma}^*)'\mathbf{F}_{\tilde{\gamma}} \\ &\quad -\frac{1}{2}(\tilde{\gamma} - \breve{\gamma}^*)'\mathbf{G}_{\tilde{\gamma}}(\tilde{\gamma} - \breve{\gamma}^*) + c_2, \\ &\approx -\frac{1}{2}[\tilde{\gamma}'\mathbf{G}_{\tilde{\gamma}}\tilde{\gamma} - 2\tilde{\gamma}'(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}\breve{\gamma}^*)] + c_2, \end{aligned} \tag{3.3.7}$$

where c_2 is independent of $\tilde{\gamma}$; $\mathbf{F}_{\tilde{\gamma}}$ and $\mathbf{G}_{\tilde{\gamma}}$ are the gradients (i.e. first derivative) and negative Hessian matrix at the point $\tilde{\gamma}$, respectively.

The approximation of the log conditional posterior density of the state vector (3.3.6) is

$$\begin{aligned} \log p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \boldsymbol{\theta}) &\propto \log p(Y|\delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta) + \log p(\tilde{\gamma}) \\ &\approx -\frac{1}{2}\Big[\tilde{\gamma}'(\mathbf{G}_{\tilde{\gamma}} + \mathbf{K}'_{\tilde{\gamma}}\mathbf{K}_{\tilde{\gamma}})\tilde{\gamma} - 2\tilde{\gamma}'(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma})\Big] + c_3, \end{aligned} \tag{3.3.8}$$

where the constant term c_3 is independent of $\tilde{\gamma}$. It can be seen that the approximation of the conditional posterior density (3.3.8) is Gaussian with the precision $\mathbf{P}_{\tilde{\gamma}} = \mathbf{G}_{\tilde{\gamma}} + \mathbf{K}'_{\tilde{\gamma}}\mathbf{K}_{\tilde{\gamma}}$ and the mean vector $\hat{\tilde{\gamma}} = \mathbf{P}_{\tilde{\gamma}}^{-1}(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma}^*)$.

To sample $\tilde{\gamma}$ from the Gaussian approximation $N(\hat{\tilde{\gamma}}, \mathbf{P}_{\tilde{\gamma}}^{-1})$, we first take Cholesky decomposition $\mathbf{C}_{\tilde{\gamma}}$ of the precision matrix $\mathbf{P}_{\tilde{\gamma}}$, i.e. $\mathbf{C}'_{\tilde{\gamma}}\mathbf{C}_{\tilde{\gamma}} = \mathbf{P}_{\tilde{\gamma}}$, and then use the backward and forward algorithm in Chan and Jeliazkov (2009) to solve $\hat{\tilde{\gamma}}$.[8] That is,

$$\begin{aligned} \hat{\tilde{\gamma}} &= \mathbf{P}_{\tilde{\gamma}}^{-1}(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma}^*), \\ \mathbf{C}'_{\tilde{\gamma}}\mathbf{C}_{\tilde{\gamma}}\hat{\tilde{\gamma}} &= (\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma}^*), \\ \mathbf{C}_{\tilde{\gamma}}\hat{\tilde{\gamma}} &= (\mathbf{C}'_{\tilde{\gamma}})^{-1}(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma}^*), \\ \hat{\tilde{\gamma}} &= \mathbf{C}_{\tilde{\gamma}}^{-1}(\mathbf{C}_{\tilde{\gamma}}^{-1})(\mathbf{F}_{\tilde{\gamma}} + \mathbf{G}_{\tilde{\gamma}}\breve{\gamma}^*). \end{aligned} \tag{3.3.9}$$

[8]For a $s \times s$ matrix, only $O(s)$ operations are needed for the Cholesky decomposition while $O(s^3)$ operations are required for taking an inverse of a full matrix. Additional draws from the Gaussian approximation can be obtained with low marginal cost.

As for the choice of $\breve{\gamma}^*$, the most natural choice is the posterior mode, which can be easily obtained via the Newton-Raphson method (see Kroese *et al.* (2011, pp.688-689)).

Given candidate draws of the Gaussian approximation $N(\widehat{\tilde{\gamma}}, \mathbf{P}_{\tilde{\gamma}}^{-1})$, we follow Chan and Strachan (2014) and use an ARMH algorithm (Chib and Greenberg (1995, pp.331-332)) to determine whether the candidate draws are accepted or rejected. To illustrate the ARMH algorithm, we define the set

$$\mathcal{D} = \{\tilde{\gamma} : p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta) \leq dq(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)\}, \tag{3.3.10}$$

where $p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)$ is the target density and $q(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)$ is the proposal density. The classic AR algorithm requires the existence of a constant d to ensure that the Gaussian proposal density dominates the target density. In practice, it is often difficult to chose such a d to ensure the condition (3.3.10) holds, especially if $\tilde{\gamma}$ is high dimensional vector. To relax the dominant condition, the MH algorithm is utilised to decide whether the candidate draws are accepted when the condition is not satisfied for some $\tilde{\gamma}$. The ARMH therefore can be summarised in the following steps:

1. AR step: Generate a draw $\tilde{\gamma}^*$ from the proposal density $q(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)$ with the acceptance-reject probability

$$\alpha_{AR}(\tilde{\gamma}^*) = \min\left\{1, \frac{p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}{dq(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}\right\}.$$

2. MH step: Given the current draw $\tilde{\gamma}$ and $\tilde{\gamma}^*$

   - if $\tilde{\gamma} \in \mathcal{D}$, $\alpha_{MH}(\tilde{\gamma}^*, \tilde{\gamma}) = 1$;
   - if $\tilde{\gamma} \in \mathcal{D}^c$ and $\tilde{\boldsymbol{\gamma}}^* \in \mathcal{D}$, then

$$\alpha_{MH}(\tilde{\gamma}^*, \tilde{\gamma}) = \frac{dq(\tilde{\gamma}^*|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}{p(\tilde{\gamma}^*|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)};$$

   - if $\tilde{\gamma} \in \mathcal{D}^c$ and $\tilde{\gamma}^* \in \mathcal{D}^c$, then

$$\alpha_{MH}(\tilde{\gamma}^*, \tilde{\gamma}) = \min\left\{1, \frac{p(\tilde{\gamma}^*|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta) q(\tilde{\gamma}^*|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}{p(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta) q(\tilde{\gamma}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}\right\}.$$

   Return to $\tilde{\gamma}^*$ with probability $\alpha_{MH}(\tilde{\gamma}^*, \tilde{\gamma})$; otherwise return $\tilde{\gamma}$.

To choose the constant d in (3.3.10), we follow Chan and Strachan (2014) to set $d = r\frac{p(\hat{\tilde{\gamma}}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}{q(\hat{\tilde{\gamma}}|Y, \delta_0, \gamma_0, \tilde{\delta}, \tilde{\gamma}, \theta)}$, where r is between 1 and 5, and $\hat{\tilde{\gamma}}$ is the mode of the conditional density of $\tilde{\gamma}$. Similar procedure

is applied to sample $\tilde{\gamma}, \omega_{\gamma}$.

3.4 Testing for Time-variation in Inefficiency

This section focuses on testing time-variation in base inefficiencies. Time-varying parameter models are widely used in many research areas, but specification tests for time-varying parameter models have been recently investigated (see Frühwirth-Schnatter and Wagner (2010), Chan (2016)). To test for time-varying parameters, Frühwirth-Schnatter and Wagner (2010) develop the stochastic model specification search approach. They introduce model indicators which yield 1 if the parameters are time-varying, and 0 otherwise, and then propose a Gibbs sampler to jointly sample the model indicators with other parameters in the model to improve the efficiency of the proposed algorithm. Their approach is developed for a linear Gaussian and a linear non-Gaussian state space model. Chan (2016) considers another approach which is based on a Bayes factor to conduct a formal test for time-varying parameters in a stochastic volatility model. Here we use a number of results from both Frühwirth-Schnatter and Wagner (2010) and Chan (2016) to conduct a test for time-varying in efficiencies. It is noted that the test can be used to for testing time-variation in parameters associated with a production function.

Given the model specification in (3.2.3)-(3.2.7), no time variation in the base technical inefficiencies implies the variances $\omega^2_{\gamma_i^1}$ are all equal to zeros. The null hypothesis is H_0: $\omega^2_{\gamma_i^1} = 0$ and the alternative hypothesis is H_1: $\omega^2_{\gamma_i^1} > 0$. As pointed out by Frühwirth-Schnatter and Wagner (2010), such a hypothesis test results in a non-regular testing problem because the null hypothesis lies on the boundary of parameter space. The boundary issue can be avoided by working with the standard deviation ω_{κ} instead. This is because the support of the standard deviation lies on the real space. The null hypothesis test for time-varying parameters is then H_0: $\omega_{\gamma_i^1} = 0$ and the alternative hypothesis is H_1: $\omega_{\gamma_i^1} \neq 0$. Bayes factor is the most widely used approach for such a model comparison in a Bayesian framework.

Let us denote the unrestricted model as $\mathcal{M}_o$ and the restricted model as $\mathcal{M}_j$. The Bayes factor in favour model $\mathcal{M}_o$ against $\mathcal{M}_j$ is defined as

$$BF_{oj} = \frac{p(Y|\mathcal{M}_o)}{p(Y|\mathcal{M}_j)}, \tag{3.4.1}$$

where $p(Y|\mathcal{M}_l), l = o, j$, is the marginal likelihood under model $\mathcal{M}_l$ evaluated at the observed

data Y. The Bayes factor is related to the posterior odds ratio between the two models such as

$$\frac{p(\mathcal{M}_o|Y)}{p(\mathcal{M}_j|Y)} = \frac{p(\mathcal{M}_o)}{p(\mathcal{M}_j)} BF_{oj}. \tag{3.4.2}$$

The first term of the right-hand side of (3.4.2) is referred to as the prior odds ratio. If one chooses the equal prior for each model, i.e. $p(\mathcal{M}_o) = p(\mathcal{M}_j)$, the posterior odds ratio between model $\mathcal{M}_o$ and model $\mathcal{M}_j$ is equivalent to the Bayes factor. The case of $BF_{oj} = 20$ suggests that Model $\mathcal{M}_o$ is 20 times more likely than Model $\mathcal{M}_j$ given the data.

The key element of computing the Bayes factor is the marginal likelihood which does not have a standard form in our context. A number of approaches for computing a non-standard marginal likelihood for a Gaussian and a non-Gaussian state space model are proposed such as importance sampling (Chan (2016)) or auxiliary mixture sampling (Frühwirth-Schnatter and Wagner (2008)). The computation can be demanding in the cases. In our context, the Savage Dickey (SD) ratio test proposed by Verdinelli and Wasserman (1995) can be used to calculate a Bayes factor for a nested model. An appealing feature of the SD ratio test is that it does not require the computation of marginal likelihood; it only requires the information from the unrestricted model. The SD ratio test in favour of unrestricted model is defined as

$$BF_{oj} = \frac{p(\omega_{\gamma_i^1} = 0)}{p(\omega_{\gamma_i^1} = 0|Y)}. \tag{3.4.3}$$

The numerator of (3.4.3) is computed by evaluating the normal prior at the value of 0, which is directly obtained. The normal priors of $\omega_{\gamma_i^1}$, $N(0, \sigma^2_{\omega_{\gamma_i^1}})$, implies that $\omega^2_{\gamma_i^1} \sim G(\frac{1}{2}, \frac{1}{2V_{\omega_{\gamma_i^1}}})$, where $G(.)$ is a Gamma distribution (see Kroese and Chan (2014)). The gamma prior has more mass concentrated around small values of $\sigma^2_{\gamma_i^1}$, and therefore the prior provides more shrinkage (see Frühwirth-Schnatter and Wagner (2010) and Chan (2016) for further discussion). This means that the Gamma prior of the variances favours the constant coefficient model. The denominator of (3.4.3) is computed by using the so-called Rao Blackwell algorithm. That is,

$$\widehat{p(\omega_{\gamma_i^1} = 0}|Y) = \frac{1}{R}\sum_{r=1}^{R} p(\omega_{\gamma_i^1} = 0|Y, {\delta_0}^{(r)}, {\gamma_0}^{(r)}, \tilde{\delta}^{(r)}, \tilde{\gamma}^{(r)}, \omega_\beta^{(r)}, \omega_b^{(r)}, \omega_{{\gamma_i^{-1}}^{(r)}} \omega_v^{(r)}), \tag{3.4.4}$$

where $\omega_{\gamma^{-1}}$ is a vector of ω_γ without an element of ω_{γ^1}. The ${\delta_0}^{(1)}, {\gamma_0}^{(1)}, \tilde{\delta}^{(1)}, \tilde{\gamma}^{(1)}, \omega_\beta^{(1)}, \omega_b^{(1)}, \omega_{{\gamma_i^{-1}}^{(1)}}$, $\omega_v^{(1)}, \ldots, {\delta_0}^{(r)}, {\gamma_0}^{(r)}, \tilde{\delta}^{(r)}, \tilde{\gamma}^{(r)}, \omega_\beta^{(r)}, \omega_b^{(r)}, \omega_{{\gamma_i^{-1}}^{(r)}}, \omega_v^{(r)}$ are posterior draws from the MCMC algorithm.

The SD approach can be used to test for the significance of a parameter. For example, if one

is interested in testing whether the coefficient δ_0 in (3.3.1) is significant, the Bayes factor is computed as

$$BF_{oj} = \frac{p(\delta_0 = 0)}{p(\delta_0 = 0|Y)}. \tag{3.4.5}$$

3.5 Empirical Study

A number of studies find that FDI has an impact on improving an economy's efficiency via transferring technology (see Iyer *et al.* (2008), Mastromarco and Ghosh (2009), and others). For example, a local labour force gains more skills through training received from foreign firms, or local firms adopt new technologies brought by foreign companies. FDI can also increase productivity by spurring competition between domestic and foreign companies (Glass and Saggi (1998)). An improvement in new foreign competition arrivals encourages domestic firms to use their resources more efficiently, which implies an improvement in their productivity. These studies often make a prior assumption about the effect of FDI on either the production function or the inefficiency term.

Deviating from the studies, we jointly model the effects of FDI on both the production function and efficiency, and allow the effects of FDI to the production process to vary over time. A study closely related to ours is Mastromarco and Simar (2015), which investigate the issue in the context of a deterministic nonparametric frontier, namely a conditional nonparametric estimator. The deterministic nonparametric frontier approach does not require an assumption about the production frontier, but assumes away statistical noise. A standard bootstrap algorithm cannot directly be applied to obtain statistical inferences for inefficiencies in the framework. Also, testing for time-variation in inefficiencies is not considered in their framework. Although our approach requires an assumption about the frontier, it takes into account of statistical noise in the estimation of the frontier. It is noted that our model is parametric but with flexible form; we do not restrict the production frontier to be a Cobb-Douglas or a translog function. The statistical inferences are straightforward to obtain in Bayesian framework.

In this section, we wish to investigate under what channels the environmental factor FDI affects the production process: through the inefficiency component, or the production What countries make the most efficient use of their inputs? Is economic growth driven by improving efficiency? Or is it due to change in inputs? Or is it driven by the shift of the frontier? Does FDI improve output growth? Another aim of this section is to test whether the inefficiencies of the economies are time-varying.

The data used in this section contain aggregate output real GDP (Y), labour (L), and capital stock (K) for 21 OECD countries over the period 1970-2011. The real GDP, labour and capital are obtained from Penn World Tables (version 8.0). All three variables are logged and normalised with respect to the mean to ensure homogeneity assumption in inputs and outputs. FDI is collected from the World Bank World Development Indicators and transformed as a ratio to GDP. The descriptive statistics for the data are presented in Table 3.B.1.

3.5.1 The Model

To facilitate further discussion we rewrite the stochastic frontier model here

$$
\begin{aligned}
y_{it} &= x_{it}\beta_t - u_{it} + v_{it} && v_{it} \sim N(0, \sigma^2_{v_i}), && (3.5.1)\\
u_{it} &= \exp(z_{it}\gamma_{it}) && && (3.5.2)\\
\beta_t &= \beta_{t-1} + b_{t-1} + \varepsilon_{\beta_t} && \varepsilon_{\beta_t} \sim N(0, \mathbf{w}^2_{\beta}), && (3.5.3)\\
b_t &= b_{t-1} + \varepsilon_{b_t} && \varepsilon_{b_t} \sim N(0, \mathbf{w}^2_{b}), && (3.5.4)\\
\gamma_{it} &= \gamma_{it-1} + \varepsilon_{\gamma_{it}} && \varepsilon_{\gamma_{it}} \sim N(0, \mathbf{w}^2_{\gamma_i}). && (3.5.5)
\end{aligned}
$$

The output y_{it} is log(GDP)$_{it}$, the vector z_{it} contains constant and FDI, i.e. $z_{it} = [1 \text{ FDI}_{it}]$. For the vector x_{it}, we consider a number of specifications. That is, x_{it} includes

(i) a constant, the inputs and FDI, i.e. $x_{it} = [1, \log(\text{L}_{it}), \log(\text{K}_{it}), \text{FDI}_{it}]$;

(ii) the variables in (i), the cross terms between the inputs, and the squared terms of the inputs, i.e. $x_{it} = [1, \log(\text{L}_{it}), \log(\text{K}_{it}), \log(\text{L}_{it})\log(\text{K}_{it}), \log^2(\text{L}_{it}), \log^2(\text{K}_{it}), \text{FDI}_{it}]$;

(iii) the variables in (ii), and the squared term of FDI, i.e. $x_{it} = [1, \log(\text{L}_{it}), \log(\text{K}_{it}), \log^2(\text{L}_{it}), \log^2(\text{K}_{it}), \log(\text{L}_{it})\log(\text{K}_{it}), \text{FDI}_{it}, \text{FDI}^2_{it}]$;

(iv) the variables in (iii), and the cross terms between FDI and the inputs FDI, i.e. $x_{it} = [1, \log(\text{L}_{it}), \log(\text{K}_{it}), \log^2(\text{L}_{it}), \log^2(\text{K}_{it}), \log(\text{L}_{it})\log(\text{K}_{it}), \text{FDI}_{it}, \text{FDI}^2_{it}, \text{FDI}_{it}\log(\text{L}_{it}), \text{FDI}_{it}\log(\text{K}_{it})]$;

It is noted that here we assume the exogeneity of FDI as many others do (see van Pottelsberghe de la Potterie and Lichtenberg (2001), Kneller and Stevens (2006), Wang and Wong (2012), and etc.). The issue of reverse causality between FDI and an output is still an open issue in the literature on the effects of FDI on output growth.

For the purpose of testing for time-varying parameters and model specifications, we rewrite the model (3.5.1)-(3.5.5) in the non-centred representation:

$$Y = \mathbf{X}_{\delta_0}\delta_0 + \mathbf{X}_{\tilde{\delta}}\tilde{\delta} - \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma}) + v, \tag{3.5.6}$$

The structure of the matrices $Y, \mathbf{X}_{\delta_0}, \mathbf{X}_{\tilde{\delta}}$, $\mathbf{Z}$, $\mathbf{Z}_{\tilde{\gamma}}$ are presented in Section 3.3. It is noted that the model representation of (3.5.1)-(3.5.5) and its non-centred representation (3.5.6) are equivalent. An advantage of the non-centred parametrisation is that it enables to test for time-varying parameters as the boundary support of the variances is not an issue (see Section (3.4) for the discussion).

The priors of the parameters $\delta_0, \gamma_0, \omega_\beta, \omega_b, \omega_\gamma, \omega_v^2$ in (3.5.6) are as follows. We choose fairly non-informative normal priors for the parameters δ_0, i.e. $p(\delta_0) = N(\mu_\delta, \mathbf{I}_{2k})$, where μ_δ is the maximum likelihood estimates of the conventional stochastic frontier model (see Aigner *et al.* (1977)).[9] For the parameters ω_β, ω_b, ω_γ, and ω_v^2, we also choose non-informative priors, e.g. $\omega_\beta \sim N(\mathbf{0}, \mathbf{I}_{2k}), \omega_b \sim N(\mathbf{0}, \mathbf{I}_{2k})$, $\omega_\gamma \sim N(\mathbf{0}, \mathbf{I}_{nm})$, and $\omega_v^2 \sim IG(10, 1)$. For the parameters $\gamma_0 = [\gamma_0^{'(1)}, \gamma_0^{'(2)}]$ where $\gamma_0^{(1)}$ is associated with the base efficiency of each country, and $\gamma_0^{(2)}$ associated with the effect of FDI on efficiency for each country. We chose a non-informative prior for $\gamma_0^{(2)}$, e.g. $\gamma_0^{(2)}) \sim N(\mathbf{0}, \mathbf{I}_r)$, while s we choose an informative prior for $\gamma_0^{(1)}$), e.g. $\gamma_0^{(1)} \sim N(\log(0.2)\mathbf{1}_n, 0.1\mathbf{I_n})$, indicating that the base inefficiency follows a log normal distribution $\text{LN}(0.204\mathbf{1}_n, 0.67\mathbf{I_n})$.

It is noted the priors for most of the parameters in our model are fairly non-informative whilst the prior for the base efficiency is informative (e.g. the variance is set at relatively small). To check the robustness of the results relative to the informative prior choice of the base efficiency, we have estimated the model with a range of mean values of $(\log(0.1), \log(0.9))$. We find that our main conclusions regarding the effects of FDI on the production process, and the contributions of capital deepening and technology change to the productivity growth are robust to the choice, although the magnitude of the estimates changes. Our results in the empirical study are based on 30,000 MCMC iterations with 5,000 burn-in.

[9]The conventional stochastic frontier model we consider here is $Y = \mathbf{X}_{\delta_0}\delta_0 + \mathbf{X}_{\tilde{\delta}}\tilde{\delta} - u + v$ where $u \sim N^+(0, \sigma_u^2 \boldsymbol{I}_{nT}), v \sim N(0, \sigma_v^2 \boldsymbol{I}_{nT})$.

3.5.2 Results

Efficiency levels and the role of FDI to a production process

To examine under what channels FDI influences the production process: we perform hypothesis tests for the significant effects of FDI on a production function and an inefficiency component. In our context, we are testing the significance of the coefficients δ_0 associated with the variables FDI_{it}, FDI_{it}^2, $\text{FDI}_{it}\log(\text{L}_{it})$, $\text{FDI}_{it}\log(\text{K}_{it})$, and $\gamma_0^{(2)}$. Tables 3.5.1 and 3.5.2 report the values of log Bayes factor ratio and their numerical standard errors for the 21 OECD countries. The numerical standard errors are computed using 10 parallel chains, each of which is of length 30000 with a burn-in of 5000. A positive log-Bayes factor suggests evidence supporting significant effects of FDI on the production process.

The results are robust across the model specifications. The first two lines of Table 3.5.1 provide the Bayes factor for testing the significant effect of FDI on the production function. This information reflects the effect of FDI on the shift of the frontier, which measures technological change. The last two lines of Table 3.5.1 capture the effect of the interaction relationship between FDI and the inputs on the production function. This reflects the effect of FDI on the shape of the frontier, which sometimes is referred to as technological bias. The log Bayes factors are positive and significant across all the models (see Table 3.5.1), which implies that FDI has significant impacts on the shift and the shape of the frontier. However, the log Bayes factors regarding the test for significant effects of FDI on the distribution of inefficiencies in Table 3.5.2 are mostly negative, which suggest that FDI appear to have less influence on technical efficiencies. Over all, FDI appears to play more important role in shifting the frontier (technological change), and has less influence on the distribution of technical efficiencies. This result supports the theoretical hypothesis that FDI improves productivity by spurring competition: foreign firms invest more in innovation to maintain their technological advantage (Glass and Saggi (1998)). FDI should be therefore invested in innovation to further improve technological change. A question arising here is whether FDI contributes to output growth of the countries. This question is soon discussed.

Table 3.5.1: The estimated log Bayes factors and their numerical standard errors for testing the significant effect of FDI on the production frontier

Variables	**Model (i)**	**Model (ii)**	**Model (iii)**	**Model (iv)**
FDI_{it}	1141.471	756.314	753.124	43.711
	(11.412)	(7.580)	(7.522)	(1.037)
FDI_{it}^2			1865.238	1301.764
			(18.830)	(13.199)
$\text{FDI}_{it} \times \log(\text{L}_{it})$				129.919
				(1.525)
$\text{FDI}_{it} \times \log(\text{K}_{it})$				42.245
				(0.468)

Table 3.5.2: The estimated log Bayes factors and their numerical standard errors (in parentheses) for testing the significant effect of FDI on the technical inefficies

	Model (i)	**Model (ii)**	**Model (iii)**	**Model (iv)**
Countries/Parameters	$\gamma_0^{(2)}$	$\gamma_0^{(2)}$	$\gamma_0^{(2)}$	$\gamma_0^{(2)}$
Australia	-0.569	-0.478	-0.485	-0.463
	(0.054)	(0.040)	(0.037)	(0.033)
Austria	-1.642	-1.872	-1.280	-1.523
	(0.042)	(0.032)	(0.032)	(0.027)
Belgium	-8.900	-8.724	-10.359	-9.225
	(0.074)	(0.062)	(0.028)	(0.026)
Canada	-0.133	-0.061	-0.185	-0.141
	(0.054)	(0.041)	(0.038)	(0.034)
Denmark	-1.514	-1.879	-1.490	-1.668
	(0.044)	(0.032)	(0.032)	(0.027)
Finland	-1.068	-1.409	-1.406	-1.135
	(0.047)	(0.033)	(0.031)	(0.029)
France	-0.187	-0.142	-0.155	-0.148
	(0.055)	(0.042)	(0.039)	(0.035)
Germany	-0.308	-0.331	-0.243	-0.164
	(0.054)	(0.040)	(0.038)	(0.035)
Greece	-0.186	-0.191	-0.167	-0.175
	(0.055)	(0.042)	(0.039)	(0.035)
Ireland	0.080	-1.282	-1.371	-1.796
	(0.052)	(0.034)	(0.031)	(0.031)
Italy	-0.031	-0.039	-0.038	-0.042
	(0.057)	(0.043)	(0.040)	(0.036)
Japan	-0.002	-0.003	-0.003	-0.003
	(0.058)	(0.043)	(0.041)	(0.036)
New Zealand	-1.776	-2.060	-1.905	-1.944
	(0.049)	(0.036)	(0.034)	(0.030)
Netherlands	4.551	4.971	8.324	33.756
	(0.046)	(0.029)	(0.037)	(0.035)
Norway	-0.069	-0.041	-0.043	-0.048
	(0.057)	(0.043)	(0.041)	(0.036)
Portugal	-2.591	-2.569	-2.457	-2.389
	(0.042)	(0.032)	(0.030)	(0.028)
Spain	-0.677	-0.646	-0.615	-0.618
	(0.052)	(0.039)	(0.036)	(0.033)
Sweden	-1.377	-1.354	-1.421	-1.312
	(0.044)	(0.034)	(0.031)	(0.029)
Turkey	-0.218	-0.168	-0.172	-0.152
	(0.055)	(0.042)	(0.039)	(0.035)
United Kingdom	-0.565	-0.548	-0.637	-0.592
	(0.052)	(0.039)	(0.036)	(0.033)
United States	0.006	-0.056	-0.051	-0.069
	(0.057)	(0.043)	(0.040)	(0.036)

We now turn to investigate the behaviours of technical inefficiencies of the economies, and examine whether the base inefficiencies of the countries are time-varying. The null hypothesis for such a hypothesis test is H_0: $\omega_{\gamma_i^{(1)}} = 0$. Table 3.5.3 reports the results of log Bayes factor ratio and their numerical standard error for the hypothesis test. The positive value of log Bayes factor indicates that the unrestricted model (time-variant inefficiency) is preferred to the restricted one (time-invariant inefficiency). Our results suggest that the efficiencies of most of the countries are time-varying (the log Bayes factor are positive and significant). For some countries such as Australia, Belgium and Italy, the log Bayes factors are negative, however, the stand errors indicate that the values are not significant.

Figure 3.5.1 describes the evolution of the base efficiencies of some selected OCED countries over the period 1970-2011. A country is considered as fully efficient if the efficiency score is 1. Some countries such as Canada, Norway, Ireland, Norway, France and US are relatively efficient in using their own resources, whereas Greece and Portugal have lower efficiency scores. Ireland and Norway had a high and increasing pattern of efficiencies during the period. Ireland experienced a slight decrease in the efficiencies in the early 1970, however, quickly improved its efficiency after 1975. An explanation for the improvement in the efficiency of Ireland might be due to economic reforms during the period 1980-2000. A number of economic reforms were implemented in Ireland for the purpose of transforming the country into a high-tech industry country (see Koop *et al.* (1999), Badunenko *et al.* (2008)). The efficiencies of US, France, and Canada seem to be stable over the periods while the efficiencies of Greece and Portugal appear to have a decreasing trend.

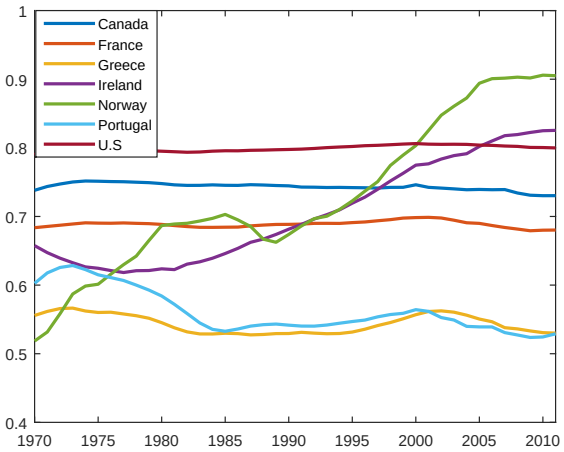


Figure 3.5.1: The evolution of the posterior means of base efficiencies of some selected OECD countries over 1970-2011 using the time-varying nonliear stochastic frontier model.

Table 3.5.3: The estimated log Bayes factors and the numerical standard errors for testing time-varying base efficiencies

	Model (i)	**Model (ii)**	**Model (iii)**	**Model (iv)**
Countries/Parameters	$\omega_{\gamma_i^{(1)}}$	$\omega_{\gamma_i^{(1)}}$	$\omega_{\gamma_i^{(1)}}$	$\omega_{\gamma_i^{(1)}}$
Australia	-1.375	-0.631	-0.891	-0.903
	(0.687)	(0.438)	(0.515)	(0.686)
Austria	16.067	9.975	10.370	11.623
	(8.033)	(3.770)	(3.666)	(3.675)
Belgium	-1.539	-1.171	-1.243	-1.101
	(0.769)	(0.443)	(0.439)	(0.348)
Canada	2.355	1.532	1.059	1.221
	(1.177)	(0.579)	(0.374)	(0.386)
Denmark	6.309	17.289	16.619	17.981
	(3.155)	(6.535)	(5.876)	(5.686)
Finland	25.493	18.494	17.935	23.026
	(12.746)	(6.990)	(6.341)	(7.281)
France	-1.558	-1.120	-1.252	-1.297
	(0.779)	(0.623)	(0.743)	(0.610)
Germany	3.665	10.747	8.376	9.285
	(1.832)	(4.062)	(2.961)	(2.936)
Greece	64.586	56.977	52.162	53.165
	(32.293)	(21.535)	(18.442)	(16.812)
Ireland	7.757	4.339	5.374	3.929
	(3.878)	(1.640)	(1.900)	(1.242)
Italy	-0.435	-0.849	-0.689	-1.057
	(0.218)	(0.521)	(0.444)	(0.534)
Japan	1.488	22.224	18.037	24.885
	(0.744)	(8.400)	(6.377)	(7.869)
New Zealand	38.469	89.228	75.431	80.708
	(19.234)	(33.725)	(26.669)	(25.522)
Netherlands	-1.519	-1.398	-0.993	-0.478
	(0.759)	(0.729)	(0.651)	(0.351)
Norway	20.384	11.600	11.529	11.540
	(10.192)	(4.384)	(4.076)	(3.649)
Portugal	269.229	237.574	218.008	219.653
	(134.614)	(89.794)	(77.078)	(69.460)
Spain	6.817	2.647	2.957	2.756
	(3.409)	(1.001)	(1.045)	(0.871)
Sweden	-1.531	-1.541	-1.593	-1.697
	(0.766)	(0.882)	(0.863)	(0.937)
Turkey	74.691	39.574	42.376	32.464
	(37.345)	(14.958)	(14.982)	(10.266)
United Kingdom	1.924	2.240	1.755	1.806
	(0.962)	(0.847)	(0.621)	(0.571)
United States	7.367	3.266	4.113	3.189
	(3.684)	(1.234)	(1.454)	(1.009)

Non dynamic model specification

For a model comparison, we go further to estimate the non-dynamic version of our model specifications with a linear trend in the production, which is

$$y_{it} = x_{it}^{*}\alpha - \eta_i \exp(z_{it}^{*}\gamma^{*}) + v_i, \tag{3.5.7}$$

where $x_{it}^{*} = [x_{it},\ t]$.

The non-dynamic models are estimated using a non-linear least squares and the results are presented in Table 3.5.4. Table 3.5.4 confirms our findings regarding the role of FDI on the production process. FDI has a significant impact on the production function (the 95% confidence intervals for the coefficients associated with FDI in the production function do not include zero). The effects of FDI on efficiencies appear to be insignificant (the 95% confidence intervals for the coefficients associated with FDI in the inefficiency term include zero).

We also plot the efficiencies of some selected OCED obtained from the non-dynamic model specification (Figure 3.5.2). The evolution of the efficiencies of US, France, Canada and Ireland obtained from our proposed model is smoother than that obtained from the non-dynamic model. For Portugal and Greece, the non-dynamic model suggests that the efficiencies of the countries greatly fluctuated during the time period. In particular, we observe a sharp increase in the efficiency of Portugal in the early 1970 and a sharp decline in its efficiency in 2001. On the other hand, our results suggest that Portugal experienced a decline in the efficiency from 1970-1985, and remained at the low efficiency level from 1985-2011. We observe a high persistence in the adjustment of Portugal's efficiency while it is not the case in the non-dynamic model. A high persistence in the adjustment of efficiencies is a feature that many researchers often wish to capture the dynamic pattern of efficiencies (Ahn and Sickles (2000), Tsionas (2006)). A rationale behind this is that efficiencies might take time to be adjusted due to the potentially large adjustment costs associated with the inputs, or persistent change in management skills. Such a high persistence is captured in our model specification. Sudden change in efficiencies might occur when there are changes affecting the operating environment (e.g. change in government regulation).

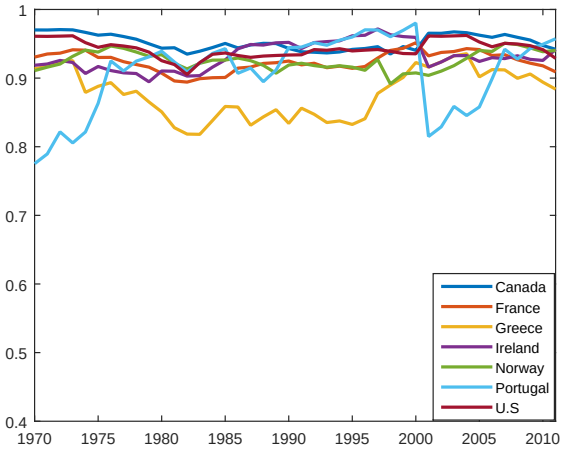


Figure 3.5.2: The evolution of the base efficiencies of some selected OECD countries over 1970-2011 using the non-dynamic specifications.

Table 3.5.4: The estimated coefficients and their 95% confidence intervals for the non-dynamic specification of Model (i), (ii) and (iii) with a linear trend in the production function

Variables	**Model Specification**		
Production Function	**Model (i)**	**Model (ii)**	**Model (iii)**
Intercept	5.110 (4.321; 6.969)	-6.604 (-16.219; 6.749)	-5.438 (-14.662; 6.949)
t	0.008 (0.004; 0.013)	0.009 (0.005; 0.014)	0.008 (0.005; 0.013)
$\log(L_{it})$	0.490 (0.422; 0.651)	-1.405 (-3.231; 1.696)	-1.180 (-2.991; 1.489)
$\log(K_{it})$	0.538 (0.374; 0.610)	2.543 (0.118; 4.243)	2.335 (0.146; 3.934)
$\log(L_{it})\log(K_{it})$		0.153 (-0.121; 0.321)	0.133 (-0.010; 0.303)
$\log^2(L_{it})$		-0.047 (-0.163; 0.125)	-0.035 (-0.152; 0.113)
$\log^2(K_{it})$		-0.085 (-0.161; 0.024)	-0.076 (-0.147; 0.022)
FDI_{it}	0.027 (0.017; 0.059)	0.027 (0.016; 0.057)	-0.021 (-0.043; 0.072)
$FDI^2{}_{it}$			1.21E-04 (-2.40E-04; 0.003)
Inefficiency Component			
FDI_{it}	0.014 (-0.009; 0.030)	0.014 (-0.010; 0.029)	-0.048 (-0.079; 0.057)

The 95% confidence intervals are computed by using the block bootstrap procedure. The bootstrap iterations are 1000.

Growth Decomposition

Here we decompose output growth into inefficiency change, capital deepening, technological change, and FDI change. We wish to investigate which component contributes to the output growth. To decompose output growth we follow the procedure similar to those of Koop *et al.* (2000) and Kumar and Russell (2002), which is summarised as follows. Given the specification (3.5.1)-(3.5.5), the expected increase in the log of GDP of a country i from the period t to $t+1$ is

$$(x_{t+1}\beta_{t+1} - x_t\beta_t) + (u_{i,t} - u_{i,t+1}). \tag{3.5.8}$$

The first term $(x_{t+1}\beta_{t+1} - x_t\beta_t)$ reflects the changes in technology, inputs and FDI. The second term $(u_{i,t} - u_{i,t+1})$ reflects changes in inefficiencies.

Suppose that the variables x_{it} and $x_{i,t+1}$ are fixed at some levels such as x_0, then the expected output growth in (3.5.8) is measured as

$$\exp(x_0(\beta_{t+1} - \beta_t)) \exp(u_{i,t} - u_{i,t+1}). \tag{3.5.9}$$

The first component $\exp(x_0(\beta_{t+1} - \beta_t))$ measures the "pure" technological change (the shift of the frontier). The second term $\exp(u_{i,t} - u_{i,t+1})$ measures the individual efficiency change. In practice inputs often change over time from $x_0 = x_{i,t}$ at period t to $x_0 = x_{i,t+1}$ at period $t+1$. To construct the technological change in such a case, we use the Fisher decomposition which is based on the geometric of the technological change for a given $x_0 = x_{i,t}$ and $x_0 = x_{i,t+1}$ (see Caves *et al.* (1982), Färe *et al.* (1994)). The technological change is hence defined as

$$\text{TC}_{i,t+1} = \exp\left\{\frac{1}{2}(x_{i,t+1} + x_{i,t})'(\beta_{t+1} - \beta_t)\right\}.$$

The change in expected GDP of country i in (3.5.8) is re-written as

$$\begin{aligned} \text{GC}_{i,t+1} &= \exp\left(\frac{1}{2}(x_{t+1} + x_t)(\beta_{t+1} - \beta_t)\right) \times \exp\left(\frac{1}{2}(x_{t+1} - x_t)(\beta_{t+1} + \beta_t)\right) \times \exp\left(u_{i,t} - u_{i,t+1}\right), \\ &= \text{TC}_{i,t+1} \times \text{IFC}_{i,t+1} \times \text{EC}_{i,t+1}, \end{aligned}$$

where $\text{EC}_{i,t+1} = \exp(u_{it} - u_{i,t+1})$ captures efficiency change, and $\text{IFC}_{i,t+1} = \exp\left(\frac{1}{2}(x_{t+1} - x_t)(\beta_{t+1} + \beta_t)\right)$ reflects the input change and FDI change.

The component $\text{IFC}_{i,t+1}$ is further decomposed into input change ($\text{IC}_{i,t+1}$) and FDI change ($\text{FDI}_{i,t+1}$). As $x_{i,t+1} = [\tilde{x}_{i,t+1}, \tilde{x}^*_{i,t+1}]$, where $\tilde{x}_{i,t+1}$ are variables containing inputs while $\tilde{x}^*_{i,t+1}$ are variables containing FDI, $\beta_{i,t+1} = [\tilde{\beta}_{i,t+1}, \tilde{\beta}^*_{i,t+1}]$ are the parameters associated with $\tilde{x}_{i,t+1}$

and $\tilde{x}^*_{i,t+1}$, $\text{IFC}_{i,t+1}$ is written as

$$\begin{aligned}\text{IFC}_{i,t+1} &= \exp\left(\frac{1}{2}(x_{t+1}-x_t)(\beta_{t+1}+\beta_t)\right),\\ &= \exp\left(\frac{1}{2}(\tilde{x}_{t+1}-\tilde{x}_t)(\tilde{\beta}_{t+1}+\tilde{\beta}_t)\right)\times\exp\left(\frac{1}{2}(\tilde{x}^*_{t+1}-\tilde{x}^*_t)(\tilde{\beta}^*_{t+1}+\tilde{\beta}^*_t)\right),\\ &= \qquad \text{IC}_{i,t+1} \qquad\qquad \times \qquad\qquad \text{FDIC}_{i,t+1}.\end{aligned}$$

To this end, the growth in expected GDP of country i is then written as

$$\text{GC}_{i,t+1} = \text{TC}_{i,t+1}\times\text{IC}_{i,t+1}\times\text{FDIC}_{i,t+1}\times\text{EC}_{i,t+1}. \tag{3.5.10}$$

The cumulated output growth and its components over the time period is given by

$$\text{CPC}_i = \text{CTC}_i\times\text{CIC}_i\times\text{CFDIC}_i\times\text{CEC}_i, \tag{3.5.11}$$

where

$$\text{CTC}_i = \prod_{t=1}^{T-1}\text{TC}_{i,t+1};\text{CIC}_i = \prod_{t=1}^{T-1}\text{IC}_{i,t+1};\text{CFDIC}_i = \prod_{t=1}^{T-1}\text{FDIC}_{i,t+1};\text{CEC}_i = \prod_{t=1}^{T-1}\text{EC}_{i,t+1}.$$

Average changes are defined as geometric average of annual changes, i.e.,

$$\begin{aligned}\text{AOGC}_i &= (\text{CTC}_i)^{\frac{1}{T-1}}\times(\text{CIC}_i)^{\frac{1}{T-1}}\times(\text{CFDIC}_i)^{\frac{1}{T-1}}\times(\text{CEC}_i)^{\frac{1}{T-1}},\\ &= \text{ATC}_i\times\text{AIC}_i\times\text{AFDIC}_i\times\text{AEC}_i.\end{aligned} \tag{3.5.12}$$

To facilitate the discussion of our results, we use average annual percentage growth . That is, $\text{AOG}_i = 100\times(\text{AOGC}_i - 1)$, $\text{ATG}_i = 100\times(\text{ATC}_i - 1)$, $\text{AIG}_i = 100\times(\text{AIGC}_i - 1)$, $\text{AFDIG}_i = 100\times(\text{AFDIC}_i - 1)$, $\text{AEG}_i = 100\times(\text{AEC}_i - 1)$.

The second column of Table 3.5.5 reports average output growth (AOG) of the economies using the actual data. The third-seventh column of Table 3.5.5 present posterior means and standard deviations of these four measures along with expected GDP growth (i.e. AEG), which is approximately equal to the sum of AIG, ATG, AFDIG and AEFFG). The standard deviations of most of the average growth measures are substantial, and therefore, our conclusions contain a considerable degree of uncertainty. This is not surprising. It would be more surprising to expect a small and noisy data set to answer the complicated questions about growth decomposition that we consider with any high degree of accuracy.

A general pattern observed from Table 3.5.5 is that input change and technological change are

Table 3.5.5: Percentage change of output growth and its decomposition for Model (iii)

Countries	AOG	AEG	AIG	ATG	AFDIG	AEFFG
Australia	3.398	3.449	2.754	0.668	-0.050	0.059
	(0.000)	(0.148)	(0.082)	(0.167)	(0.037)	(0.118)
Austria	3.059	3.437	2.296	1.503	0.032	-0.414
	(0.000)	(0.214)	(0.188)	(0.217)	(0.100)	(0.211)
Belgium	2.823	2.629	1.395	1.303	0.094	-0.178
	(0.000)	(0.193)	(0.147)	(0.419)	(0.357)	(0.186)
Canada	3.102	3.440	2.857	0.459	-0.030	0.137
	(0.000)	(0.155)	(0.066)	(0.177)	(0.044)	(0.135)
Denmark	2.348	2.681	1.316	1.936	-0.005	-0.572
	(0.000)	(0.206)	(0.165)	(0.233)	(0.094)	(0.216)
Finland	2.864	2.892	1.265	2.061	0.151	-0.594
	(0.000)	(0.214)	(0.179)	(0.238)	(0.079)	(0.234)
France	2.515	2.345	1.953	0.219	0.086	0.080
	(0.000)	(0.145)	(0.165)	(0.195)	(0.027)	(0.111)
Germany	2.776	2.166	1.565	0.360	0.024	0.208
	(0.000)	(0.211)	(0.136)	(0.230)	(0.037)	(0.211)
Greece	3.009	3.024	2.135	1.453	0.058	-0.631
	(0.000)	(0.202)	(0.145)	(0.192)	(0.014)	(0.203)
Ireland	5.591	5.441	2.537	2.482	0.192	0.150
	(0.000)	(0.203)	(0.345)	(0.405)	(0.167)	(0.163)
Italy	2.856	2.634	2.346	0.242	-0.019	0.058
	(0.000)	(0.144)	(0.172)	(0.188)	(0.013)	(0.124)
Japan	3.165	3.639	3.361	0.607	0.005	-0.340
	(0.000)	(0.224)	(0.300)	(0.284)	(0.003)	(0.197)
New Zealand	2.728	2.926	1.569	2.183	0.028	-0.856
	(0.000)	(0.219)	(0.099)	(0.321)	(0.047)	(0.284)
Netherlands	3.236	3.026	2.057	0.981	0.133	-0.159
	(0.000)	(0.267)	(0.115)	(0.691)	(0.652)	(0.287)
Norway	5.341	4.907	2.068	2.168	0.098	0.503
	(0.000)	(0.210)	(0.171)	(0.248)	(0.037)	(0.176)
Portugal	3.400	3.256	2.972	1.225	0.119	-1.054
	(0.000)	(0.239)	(0.257)	(0.273)	(0.039)	(0.223)
Spain	3.517	3.607	3.129	0.429	0.136	-0.102
	(0.000)	(0.164)	(0.221)	(0.183)	(0.039)	(0.126)
Sweden	2.558	2.933	1.444	1.346	0.149	-0.028
	(0.000)	(0.140)	(0.108)	(0.227)	(0.122)	(0.110)
Turkey	4.239	3.982	3.025	1.048	0.036	-0.154
	(0.000)	(0.223)	(0.237)	(0.389)	(0.018)	(0.269)
United Kingdom	2.494	2.306	1.463	0.461	0.152	0.215
	(0.000)	(0.164)	(0.153)	(0.242)	(0.065)	(0.166)
United States	2.814	2.975	2.510	0.315	0.081	0.057
	(0.000)	(0.224)	(0.192)	(0.332)	(0.023)	(0.170)
Mean	**3.230**	**3.224**	**2.191**	**1.117**	**0.070**	**-0.172**
std.err of the mean		**(0.171)**	**(0.141)**	**(0.155)**	**(0.014)**	**(0.084)**
95% credible interval of the mean		**(2.888; 3.559)**	**(1.915; 2.468)**	**(0.812; 1.421)**	**(0.041; 0.099)**	**(-0.336; -0.008)**

the key drivers of output growth. This confirms some of the findings in the literature on economic growth, that is output growth is driven largely by accumulation of inputs, with a lesser role for improvements in productivity (see Kumar and Russell (2002), Henderson and Russell (2005), Koop *et al.* (2000) and many others). The changes in technical efficiency and changes in FDI play minor roles. For almost all countries (e.g. Austria, Greece, Japan, New Zealand, Portugal), there is either no efficiency change or it was somewhat negative. One possible explanation for the negative contribution of efficiency change to output growth is that capital was under utilised during the period, e.g. capital was idle or not efficiently used, resulting in inefficiency in capital utilisation. The idleness of capital can be explained as a consequence of unwanted accidents that occur after a firm is built, or as a result of rational *ex ante* investment plans (see Winston (1974) for further discussion). For example, firms are built with the expectation that the expected costs of running a business is affordable, or the foreign exchange for imported inputs will be available in additional to adequate working capital.

Looking closer to some countries such as Norway and Ireland, we find that the countries achieved their fast GDP growth largely due to input growth and technological change. We also find statistical evidence suggesting efficiency improvement contributes to the fast GDP growth of Norway, but not for Ireland. For other countries such as Australia, Canada, Japan and Netherlands, they relied on input change to achieve their fast growth. Other countries (i.e. Belgium, Denmark, Finland, New Zealand) which experienced slower than average GDP growth, we find that slow growth in inputs and decrease in technical inefficiencies appeared to the culprit since technological change was above average.

FDI seems to play a minor role in output growth on the average. However, when looking closer at some specific countries, we find that FDI contributes to the economic growth of some countries such as Ireland, Norway, Finland, United Kingdom, and United States. For example, FDI contributed around 0.1% to output growth of Norway, and 0.15% to output growth of Ireland. This provides some support to the fact that the period from the mid-1990s to the mid-2000s is known as the Celtic Tiger for Ireland, which refers to the rapid real economic growth fuelled by foreign direct investment. Ireland benefited from the large scale of global FDI in the 1960s, and attracted many U.S investors. In the early 1970s, policy towards FDI became increasingly more selective, encouraging a pattern of investment in the production of modern high technology goods.

Evidence on mixing performance of the MCMC algorithm

To evaluate the mixing performance of a sampling algorithm we use the mixing factor performance factors. The mixing performance factor is defined as $1+2\sum_{l=1}^{L}\phi_l$, where ϕ_l is the sample

autocorrelation at lag length l and L, which are chosen large enough so that the autocorrelation tapers off. If a mixing performance score of a parameter obtained from 10,000 draws is equal to 5, then this implies that the draws of the parameter are equivalent to 2,000 independent draws from its posterior.

Table 3.5.6: Mixing performance factors of some selected parameters

$\tilde{\gamma}^{(1)}$	$\tilde{\delta}^{(1)}$	$\delta_0^{(1)}$	$\gamma_0^{(1)}$	$\omega_{v_i}^2$	$\omega_{\beta^{(1)}}$	$\omega_{b^{(1)}}$
61.14	8.08	9.89	13.08	1.64	15.80	15.22

Table 3.5.6 summarises the mixing performance factors for some selected parameters of our model. It is noted that for the time-varying parameter $\tilde{\delta}$, and the parameter varying across individuals $\omega_{v_i}^2$, and the parameter varying across individuals and time $\tilde{\delta}$, we choose to report the 50^{th} percentile of the mixing performance factors rather than reporting each mixing performance factor of each observation at each time period. Over all, we find that the mixing performance factors are relatively low, suggesting that the proposed sampling works well in terms of producing posterior draws that are not highly correlated.

3.6 Conclusions

Our contribution in this chapter is to extend the static parametric stochastic frontier with scaling property considered in Wang and Schmidt (2002) and Alvarez *et al.* (2006) to a dynamic one. Such a model specification allows the joint effects of environmental variables on a production function and an inefficiency components. Further, it sheds a light on drivers of technical inefficiency and the dynamic pattern of technical inefficiency over time. The time-varying parameters associated with a production provide a flexible specification to encompass a Cobb-Douglas/translog function with a linear trend. Statistical inferences (e.g. confidence intervals, standard errors) of the parameters in our proposed model are directly obtained in comparison to a nonparametric stochastic frontier model . A formal test for time-variation in inefficiencies is provided.

We apply the proposed model to investigate the role of FDI to a production process of 21 OECD countries for the period 1970-2011. Our results show that FDI plays an important role as influencing the production frontier rather than changes in the distribution of technical inefficiency. We find statistical evidence suggesting that technical inefficiencies are time-varying and that

the temporal effect should be taken into account in modelling technical inefficiencies. We further decompose output growth into input changes, technological change, efficiency change and FDI change to investigate which component contributes to the output growth. Technological change, and input change make a large contribution to output growth. Technical efficiency does not appear to improve economic growth on the average, but makes a contribution to growth of some countries such as Norway and Ireland. In addition, we find FDI makes a contribution to the growth of some of the developed economies such as Ireland, Norway, Finland and United Kingdom.

3.A Appendix A: MCMC Algorithm

In the Appendix we provide the conditional posterior density of parameters in the model (3.2.9)-(3.2.13). To facilitate further discussion, recall the model here

$$
\begin{aligned}
y_{it} &= x_{it}\beta_0 + x_{it}tb_0 + x_{it}\mathbf{w}_\beta\tilde{\beta}_t + x_{it}\mathbf{w}_b\tilde{B}_t - \exp(z_{it}\gamma_{i0} + z_{it}\mathbf{w}_{\gamma_i}\tilde{\gamma}_{it}) + v_{it}, && \text{(3.A.1)}\\
\tilde{\beta}_t &= \tilde{\beta}_{t-1} + \xi_{\beta t} \qquad \xi_{\beta_t} \sim N(0, \mathbf{I}_k), && \text{(3.A.2)}\\
\tilde{b}_t &= \tilde{b}_{t-1} + \xi_{b t} \qquad \xi_{b_t} \sim N(0, \mathbf{I}_k), && \text{(3.A.3)}\\
\tilde{B}_t &= \tilde{B}_{t-1} + \tilde{b}_{t-1}, && \text{(3.A.4)}\\
\tilde{\gamma}_{it} &= \tilde{\gamma}_{it-1} + \xi_{\gamma_{it}} \qquad \xi_{\gamma_{it}} \sim N(0, \mathbf{I}_m). && \text{(3.A.5)}
\end{aligned}
$$

As discussed in Section 3.3, we group the time invariant parameters β_0, b_0, and the time-varying parameters $\tilde{\beta}_t, \tilde{b}_t, \tilde{B}_t$ into two vectors, δ_0 and $\tilde{\delta}_t$, i.e. $\delta_0 = [\beta_0', b_0']'$, $\tilde{\delta}_t = [\tilde{\beta}'_t, \tilde{b}'_t, \tilde{B}'_t]'$, $t = 1, \ldots, T$. We also stack y_{it}, x_{it}, z_{it}, v_{it}, $\tilde{\delta}_t$, $\tilde{\gamma}_{it}$ for all observations i over t, and rewrite the measurement equation (3.A.1) as

$$
Y = \mathbf{X}_{\delta_0}\delta_0 + \mathbf{X}_{\tilde{\delta}}\tilde{\delta} - \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma}) + v, \qquad v \sim N(\mathbf{0}, \mathbf{P}_\mathbf{v}) \qquad \text{(3.A.6)}
$$

1. Sampling $\delta_0 = [\beta', b']'$ from the normal conditional posterior:

$$
\boldsymbol{\delta_0} \sim N(\hat{\delta_0}, \mathbf{P}^{-1}_{\hat{\delta_0}}),
$$

with the mean $\hat{\delta_0}$ and the precision matrix $\mathbf{P}^{-1}_{\hat{\delta_0}}$ as follows

$$
\begin{aligned}
\mathbf{P}_{\hat{\delta_0}} &= \underline{\mathbf{P}_{\delta^0}} + \mathbf{X}'_{\delta_0}\mathbf{P}_\mathbf{v}\mathbf{X}_{\delta_0},\\
\hat{\delta_0} &= \mathbf{P}^{-1}_{\hat{\delta_0}}(\mathbf{X}'_{\delta_0}\mathbf{P}_\mathbf{v}(Y - \mathbf{X}_{\tilde{\delta}}\tilde{\delta} + \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma})) + \underline{\mathbf{P}_{\delta_0}\delta_0}),
\end{aligned}
$$

where $\underline{\mathbf{P}_{\delta_0}}, \underline{\delta_0}$ are the the precision matrix (inverse variance), and the mean of the prior of δ_0, i.e. $\underline{\delta_0} = [\underline{\beta}', \underline{b}']'$, $\underline{\mathbf{P}_{\delta_0}} = \begin{bmatrix} \underline{P_\beta^{-1}} & 0 \\ 0 & \underline{P_b^{-1}} \end{bmatrix}$.

2. Sampling $\tilde{\delta} = [\tilde{\beta}', \tilde{b}', \tilde{B}']'$.

We first rewrite the state equations (3.A.2)-(3.A.4) as

$$\mathbf{K}_{\tilde{\boldsymbol{\delta}}}\tilde{\delta} = \xi \qquad \xi \sim N(\mathbf{0}, \mathbf{H}_{\tilde{\delta}}), \tag{3.A.7}$$

where $\mathbf{K}_{\tilde{\delta}} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ -\mathbf{I}_k & \mathbf{0} & \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_k & \mathbf{0} & \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_k & -\mathbf{I}_k & \mathbf{0} & \mathbf{I}_k & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_k \end{bmatrix}$; $\tilde{\delta} = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{b_1} \\ \tilde{B}_1 \\ \vdots \\ \vdots \\ \tilde{\beta}_T \\ \tilde{b_T} \\ \tilde{B}_T \end{bmatrix}$.

$$\mathbf{H}_{\tilde{\delta}} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 10^8\mathbf{I}_k & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & 10^8\mathbf{I}_k \end{bmatrix}.$$

We sample $\tilde{\delta}$ from the normal conditional posterior with mean $\tilde{\hat{\delta}}$ and the precision matrix $\mathbf{P}_{\tilde{\hat{\delta}}}^{-1}$, i.e. $\tilde{\delta} \sim N(\tilde{\hat{\delta}}, \mathbf{P}_{\tilde{\hat{\delta}}}^{-1})$, where

$$\begin{aligned} \mathbf{P}_{\tilde{\hat{\delta}}} &= \mathbf{K}_{\tilde{\delta}}'\mathbf{H}_{\tilde{\delta}}\mathbf{K}_{\tilde{\delta}} + \mathbf{X}'_{\tilde{\delta}}\mathbf{P}_{\mathbf{v}}\mathbf{X}_{\tilde{\delta}}, \\ \tilde{\hat{\delta}} &= \mathbf{P}_{\tilde{\hat{\delta}}}^{-1}(\mathbf{X}_{\tilde{\delta}}\mathbf{P}_{\mathbf{v}}(Y - \mathbf{X}_{\delta_0}\delta_0 + \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma})). \end{aligned}$$

3. Sampling ω_β

For the purpose of computation, we first reparameterise $\mathbf{X}_{\tilde{\delta}}\tilde{\delta}$ as $\mathbf{X}_{\tilde{\beta}}\omega_\beta + \mathbf{X}_{\tilde{b}}\omega_b$ where $\mathbf{X}_{\tilde{\beta}} = \begin{bmatrix} \mathbf{X}_1\text{diag}(\tilde{\beta}_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2\text{diag}(\tilde{\beta}_2) & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_T\text{diag}(\tilde{\beta}_T) \end{bmatrix}$. A similar procedure is applied for $\mathbf{X}_{\tilde{b}}$. Given the reparameterisation, ω_β is sampled from the normal conditional posterior

with the mean $\hat{\omega}_\beta$ and the precision matrix $\mathbf{P}_{\hat{\omega}_\beta}^{-1}$, i.e. $\omega_\beta \sim N(\hat{\omega}_\beta, \mathbf{P}_{\hat{\omega}_\beta}^{-1})$, where

$$\begin{aligned} \mathbf{P}_{\hat{\omega}_\beta} &= \underline{P_{\hat{\omega}_\beta}} + \mathbf{X}'_{\tilde{\beta}}\mathbf{P}_{\mathbf{v}}\mathbf{X}_{\tilde{\beta}}, \\ \hat{\omega}_\beta &= \mathbf{P}_{\hat{\omega}_\beta}^{-1}(\underline{P_{\hat{\omega}_\beta}\omega_\beta} + \mathbf{X}'_{\tilde{\beta}}\mathbf{P}_{\mathbf{v}}(Y - \mathbf{X}_{\delta_0}\delta_0 - \mathbf{X}_{\tilde{b}}\omega_b + \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma}))). \end{aligned} \quad (3.A.8)$$

4. We draw ω_b from the normal conditional posterior $\omega_b \sim N(\hat{\omega}_b, \mathbf{P}_{\hat{\omega}_b}^{-1})$, where

$$\begin{aligned} \mathbf{P}_{\hat{\omega}_b} &= \underline{P_{\hat{\omega}_b}} + \mathbf{X}'_{\tilde{b}}\mathbf{P}_{\mathbf{v}}\mathbf{X}_{\tilde{b}}, \\ \hat{\omega}_b &= \mathbf{P}_{\hat{\omega}_b}^{-1}(\underline{P_{\hat{\omega}_b}\omega_b} + \mathbf{X}'_{\tilde{b}}\mathbf{P}_{\mathbf{v}}(\mathbf{Y} - \mathbf{X}_{\delta_0}\delta_0 - \mathbf{X}_{\tilde{\beta}}\omega_\beta + \exp(\mathbf{Z}\gamma_0 + \mathbf{Z}_{\tilde{\gamma}}\tilde{\gamma}))). \end{aligned} \quad (3.A.9)$$

5. For each $\omega_{v_i}^2$, we draw $\omega_{v_i}^2$ from the inverse gamma conditional posterior:

$$\omega_{v_i}^2 \sim IG\Big(\tau_{v_i} + T - 1, S_{v_i} + \frac{1}{2}\sum_{t=2}^{T}(y_{it} - \hat{y}_{it-1})^2\Big),$$

where $y_{it} = x_{it}\beta_0 + x_{it}tb_0 + x_{it}\omega_\beta\tilde{\beta}_t + x_{it}\omega_b\tilde{B}_t - \exp(z_{it}\gamma_{i0} - z_{it}\omega_{\gamma_{it}}\tilde{\gamma}_{it})$.

6. For the non-linear parameters $\gamma_0, \tilde{\gamma}, \omega_{\gamma_i}$, the sampling algorithm is presented in Section 3.2.

3.B Appendix B: Tables

Table 3.B.1: Descriptive statistics for variables used in estimation

	GDP (in 000s)	L	K (in 000s)	FDI
Australia	469,457,022	7,915,850	1,523,100,839	2.08
Austria	191,326,249	3,491,144	621,880,626	1.93
Belgium	246,537,003	3,934,699	771,728,363	8.55
Canada	771,875,552	12,799,122	2,105,717,010	2.08
Denmark	133,609,172	2,647,057	440,202,100	1.66
Finland	112,694,104	2,339,262	448,232,295	1.54
France	1,368,321,804	23,935,329	4,393,159,479	1.33
Germany	1,863,154,487	37,606,167	6,137,226,750	0.96
Greece	171,059,356	4,097,929	574,116,274	0.77
Ireland	80,763,630	1,378,069	199,133,582	5.88
Italy	1,258,005,719	22,192,072	4,533,384,827	0.53
Japan	2,813,658,161	61,193,693	10,009,963,101	0.08
New Zealand	72,957,208	1,739,870	182,613,516	2.57
Netherlands	383,048,964	6,889,215	1,181,301,257	7.88
Norway	128,789,756	2,122,041	423,445,561	1.57
Portugal	142,134,854	4,398,743	456,053,633	1.73
Spain	734,313,237	14,644,606	2,477,709,557	1.96
Sweden	213,654,308	4,275,507	540,170,993	2.60
Turkey	514,696,384	17,998,953	1,125,752,455	0.67
U.K.	1,313,025,655	26,166,970	3,139,447,521	2.63
U.S.	8,593,822,214	118,871,391	24,282,753,048	0.94

Y is real GDP measured in million U.S. dollars at 2005 constant prices (using chained index),
L is the number of persons engaged in labour force,
K is capital stock measured in million U.S. dollars at current 2005 constant prices,
FDI is measured as a ratio of net inflows of foreign direct investment to GDP.

CHAPTER 4

Reduced Rank Regression in a Large VAR Model: Forecasting GDP and Its Impulse Response to Structural Shocks

4.1 Introduction

Chapter 4 aims to investigate some economic questions such as: How does the output growth respond to structural shocks (e.g. monetary shocks)? Will the forecast of output growth and other important macroeconomic variables be improved by using a model including many macroeconomic variables? Such a study assists policy makers in navigating changes and making better decisions around the output growth. A vector autoregression model (VAR) is often used to address these issues (see Blanchard and Quah (1989), Sims (1992), Cochrane (1998) and many others). A VAR model with a large number of variables has been found, based upon a range of point and density forecast measures, to provide better forecasts than a small VAR (see Banbura *et al.* (2010), Carriero *et al.* (2011), Koop (2013), and many others). A large VAR model can also mitigate the issues related to omitted variables. However, such a large VAR model often faces the risk of over-parametrisation, which in turn could worsen forecasting performance and impulse responses.

One of the most popular approaches addressing the over-parametrisation issue is a shrinkage prior (e.g. Minnesota prior (Doan *et al.* (1984) and Litterman (1986)), ridge regression prior (Hoerl and Kennard (1970)), and g-prior (Zellner (1986))). These priors are subjective in the sense that the shrinkage parameters are not estimated via a data-driven approach. The subjective issue is overcome by using Bayesian hierarchical priors such as shrinkage Normal-Jeffreys

(Hobert and Casella (1996)), Bayesian Lasso (Park *et al.* (2008)), hierarchical fused lasso (Tibshirani *et al.* (2005)), stochastic search variable selection (SSVS) (George *et al.* (2008), Koop (2013)) and many others. The basic idea underlying these studies is to place hyper-prior distributions on the shrinkage parameters, and the parameters are then estimated using the Bayes' rule. Most of the flexible Bayesian priors often rely on the computationally intensive Markov Chain Monte Carlo (MCMC) methods. The recursive forecasting exercise therefore can become prohibitive when the size of VAR system increases. An exception is a variant of the Minnesota prior which is based on the natural conjugate prior (see Banbura *et al.* (2010), Giannone *et al.* (2015)).

Dimension reduction is an alternative approach to handle the over-parametrisation. Lasso regression (Tibshirani (1996), Fu (1998)) reduces the number of parameters by using a double-exponential prior. The double-exponential prior shrinks the regression coefficients to zero rather than close to zero. The posterior of the regression coefficients might have more than one mode, which results in some difficulties in statistical inferences and computation (see Park *et al.* (2008) for further discussion). Another approach is to use a factor model which explains the variation of a large number of variables by a small number of factors (see Forni *et al.* (2000), Stock and Watson (2002)). Such a factor model has been applied in the literature of large VARs for forecasting and structural analysis (see Bernanke and Boivin (2003), Forni *et al.* (2003), Bernanke *et al.* (2005), Christine *et al.* (2008), and many others.). Pettenuzzo *et al.* (2016) recently adopt the Bayesian compressed regression (Guhaniyogi and Dunson (2015)) from the machine learning literature to tackle the over-parametrisation. The Bayesian compressed regression approach provides sound forecasts in their application. Reduced rank regression (RRR) (Velu *et al.* (1986), Reinsel (1983), Geweke (1996)) began to gain its popularity in forecasting large VARs (Carriero *et al.* (2011), Carriero *et al.* (2016b)). The method reduces the number of parameters significantly when the regression coefficient is a reduced-rank matrix.

In this chapter, to improve the precision of estimating such a large VAR we use a reduced rank regression model specification. We carefully investigate the performance of some popular rank selection techniques in a high dimensional set-up, and the implication for forecast performance when a misspecified rank is used.

A body of research has explored the ways of specifying reduced rank matrices that are invariant to the order of the variables in the model. For instance, some studies have used a singular value decomposition (SVD) to achieve an invariant specification in factor models (Chan *et al.* (2017), Kaufmann and Schumacher (2013)), co-integrating vector error correction models and simultaneous equations models (Strachan and Inder (2004), Koop *et al.* (2010)).

An alternative approach to solve the issue of invariant ordering is to average over orderings (see Geweke (1996), Hansen (2007), Magnus *et al.* (2010), Amini and Parmeter (2012)). Averaging over orderings appears to perform well in a small dimensional set-up, however, the computation of an averaging approach can become a challenge when the number of variables are large. For example, a system of 50 variables with the number of rank equal to 5. The total of number of orderings will be more than 2 million if we use an non-invariant approach and average over the orderings. In the application in the thesis, with 119 variables, we would need to compute over 118 million orderings for a rank 5, and for a rank of 10 the number of orderings exceed over 100 trillion.

In this chapter, we achieve the invariant ordering by using a singular value decomposition (SVD). Given the SVD specification, some of the identified parameters of the model belong to the special manifold (Stiefel manifold), i.e. they are subject to semi-orthogonal constraints. Sampling these parameters with the semi-orthogonal constraints are therefore not direct. To simplify the computation of these parameters we make use of the parameter expansion approach (see Ghosh and Dunson (2009), Koop *et al.* (2010), Chan *et al.* (2017)). We introduce two parameter expansions to map the parameters from the Stiefel manifold to the real space which is no longer subject to the semi-orthogonal constraints. More importantly, we show that the conditional posteriors of the parameters in the expanded model are normal distributions, which are simple to draw from. To improve the computation time, we adopt the computation procedure exploiting a certain Kronecker structure of covariance matrices (Carriero *et al.* (2016a)).

Rank selection of the VAR coefficient matrix is an important issue. Here we first estimate a set of models with all possible values of rank, and then choose a model with the largest marginal likelihood. In our context, the computation of a marginal likelihood is not trivial as the marginal likelihood does not have a closed form. We therefore approximate the marginal likelihood by using a number of techniques such as cross entropy, predictive likelihood, and Laplace approximation. The approaches are often used in a low dimensional set-up (see Geweke (1996), Geweke (2001), Strachan and Inder (2004), Chan and Eisenstat (2015)). The performance of the approaches in a high dimensional set-up is still not quite clear. In this paper, we conduct an extensive Monte Carlo simulation to investigate their performance in a high dimensional VAR model. The Monte Carlo results suggest that the downward bias in rank selection. The downward bias is more evident when the singular values of VAR coefficients are small (close to zero). These results are not surprising, but the implication for forecasting is important and therefore we explore this area here.

We wish to explore whether the forecast performance is affected by a misspecified rank model.

Our results suggest that models with lower rank perform better than the model with a correct rank (the benchmark model) in terms of density forecast criteria (multivariate average log predictive likelihood, average log predictive likelihood). With regard to the point forecast criteria, the models with misspecified ranks do not perform as well as the benchmark for shorter forecast horizons. They appear to perform better than the benchmark model for longer forecast horizons.

For empirical illustration, we conduct a macroeconomic forecasting exercise for a large VAR model with 119 dependent variables. Within these 119 variables, we focus on forecasting real gross domestic product, consumer price index, federal funds rate, civilian unemployment rate, industrial production growth, money stock M2, real personal consumption, producer price index, and personal consumption expenditure. We compare the forecasting performance of our proposed approach to various popular alternatives: a univariate AR model with a random walk, the Minnesota prior proposed by Banbura *et al.* (2010), a dynamic factor model (Bernanke *et al.* (2005). Our results suggest that no single model dominates in forecasting the above variables. Our proposed model provides a better forecast for gross domestic product, consumer price index, and producer price index. For the impulse response of a selection of macroeconomic variables to a contractionary monetary policy shock, the results are in line with our expectation and provide some support to economic theories.

The rest of the paper is organized as follows. Section 2 discusses the invariant specification of large VARs. Section 3 provides priors and conditional posteriors for the identified parameters of the model. Section 4 discusses the approaches used to select the rank. Section 5 presents our Monte Carlo simulation study. Section 6 discusses the empirical application and Section 7 concludes the chapter. All technical details, figures and tables are presented in the Appendix.

4.2 The Model

Consider the following VAR (s) model of $1 \times n$ stationary time series $Y_t = (y_{1,t}, \ldots, y_{n,t})$, $t = 1, \ldots, T$:

$$\begin{aligned} Y_t &= Y_{t-1}\Phi_1 + Y_{t-2}\Phi_2 + \ldots + Y_{t-s}\Phi_s + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma), \\ &= X_t\Phi + \varepsilon_t, \end{aligned} \tag{4.2.1}$$

where $\Phi_1, \ldots, \Phi_s$ are all $n \times n$ coefficient matrices; $\Phi = (\Phi_1', \ldots, \Phi_s')'$ is a $ns \times n$ matrix; $X_t = (Y_{t-1}, \ldots, Y_{t-s})$ is a $1 \times ns$ vector. The error term ε_t is normally distributed with mean of zeros and full covariance matrix Σ.

If Φ is a reduced rank matrix, the model (4.2.1) is written as a reduced rank regression model such as

$$\begin{aligned} Y_t &= \sum_{j=1}^{s} Y_{t-j} B_j A + \varepsilon_t \\ &= X_t B A + \varepsilon_t. \end{aligned} \tag{4.2.2}$$

where A is a $k \times n$ matrix, B_j is a $n \times k$ $(k < n)$ matrix with $j = 1, \ldots, s$, and $B = \begin{bmatrix} B_1 \\ \vdots \\ B_s \end{bmatrix}$ is a $ns \times k$ matrix. Such a reduced rank regression has been applied in a number of research areas such as psychology (Ter Braak (1990)), neuroscience (Vounou *et al.* (2010)), geoscience (Rasti *et al.* (2014)), and recently in the literature of large VARs (Carriero *et al.* (2011), Carriero *et al.* (2016b)).

The number of parameters in the coefficients BA are significantly reduced if rank k is much smaller than n. For instance, suppose the number of variables in the VAR system $n = 25$, and the number of lags $s = 4$, this implies that 2,500 parameters $(n^2 s)$ are required to be estimated in the full rank matrix Φ in (4.2.1). However, if the matrix Φ has the rank k of 4, only 484 $(nk(s+1) - k^2)$ parameters are required to be estimated in BA in (4.2.2).

Not only is the specification (4.2.1) useful for achieving parsimony, but also has important implications. From an economic point of view, this specification captures short run co-movements (common cycles) among stationary time series. The rank of BA matrix suggests the number of common cycles among Y_t (see Engle and Kozicki (1993), Vahid and Issler (2002)).

The parameters A and B are not uniquely defined as there always exists a non-singular matrix D such that $\Phi = BDD^{-1}A = B^{\#}A^{\#}$. To ensure the uniqueness of A and B, k^2 restrictions need to be imposed. In Bayesian analysis, researchers often consider a linear normalization, i.e. $B = [I_k \ \tilde{B}]$.[1] A drawback of such a linear normalization specification is that it is dependent on variable ordering, i.e. empirical study results can change when the order of the variables in Y_t changes. The lack of invariance is due to a non-homeomorphic transformation from the parameter spaces of A, B to the identified parameters. That is, the transformation contains the discontinuity (see Chan *et al.* (2017) for further discussion).

To obtain the invariant specification of (4.2.1) the singular value decomposition (SVD) specifi-

[1] In theory, one can estimate (4.2.2) in a frequentist framework as discussed in Reinsel and Velu (1988, Chapter 2, p.28). Such an estimation might not work well in practice if the second moment matrix $\Sigma_{XX} = 1/T(X'X)$ is close to be singular, which results in a numerical instability in the estimation of the model.

cation of Φ is considered. The invariant specification using SVD has been discussed in a number of research areas such as factor models (Chan *et al.* (2017), Kaufmann and Schumacher (2013)), co-integrating vector error correction model and simultaneous equations models (Strachan and Inder (2004), Koop *et al.* (2010)). The SVD of Φ with rank k is

$$\begin{aligned} &\Phi = U\Lambda V', \\ &\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k), \lambda_1 \geq \ldots \geq \lambda_k > 0, \\ &U \in \mathcal{V}_{k,n}, V \in \mathcal{V}_{k,ns}, \end{aligned} \tag{4.2.3}$$

where $\mathcal{V}$ denotes a Stiefel manifold, e.g. $\mathcal{V}_{k,n} = \{U(n \times k) : U'U = I_k\}$, where I_k is the $k \times k$ identity matrix. All of the parameters U, Λ, V are identified up to sign, and have $nk - \frac{k(k+1)}{2}$, k, and $nsk - \frac{k(k+1)}{2}$ free elements, respectively. The estimation of such an invariant specification is discussed in the next section.

4.3 Model Estimation

This section first elicits the priors for (U, Λ, V) and the covariance variance matrix Σ in (4.2.3). Due to semi-orthogonal constraint of U and V, sampling the parameters from their respective conditional posteriors is not direct. To handle the semi-orthogonal constraints and simplify the sampling computation of the parameters, we follow Chan *et al.* (2017) and introduce two parameter expansions to map the parameters (U, Λ, V) to (A, B) which belong to a real space. We provide the priors of identified parameters A, B in the expanded model and provide the link between the priors of (A, B) and those of (U, Λ, V) . The conditional posterior densities of the identified parameters in the expanded model are also presented. We assume rank k of the VAR coefficient matrix is known in this section. The selection of k is discussed in Section 4.4.

4.3.1 Parameter Expansions and The priors

For the inverse covariance variance matrix Σ^{-1} we choose a Wishart prior, i.e. $W(S_0, v_0)$ with the scale S_0 and the degrees of freedom v_0. For the parameters U and V we choose a non informative priors (i.e. a Uniform distribution over the Stiefel manifold). For Λ, we want to choose a prior which can simplify the computation of (A, B) in the expanded model. For these

purposes, we adopt the priors proposed in Chan *et al.* (2017)

$$\frac{p(\Lambda, V|c_\lambda, \Omega)(d\Lambda)(V'dV)(U'dU)c_C}{c_\Lambda c_U}, \tag{4.3.1}$$

$$p(\Lambda, V|c_\lambda, \Omega) = |\Lambda|^{n-k} \prod_{i<j}^{k} (\lambda_i^2 - \lambda_j^2) \exp\Big\{ -\frac{c_\lambda}{2} \text{tr}(V'\Omega^{-1}V\Lambda^2) \Big\}, \tag{4.3.2}$$

$$\frac{c_\Lambda}{c_C} = \int p(\Lambda, V|\Omega)(d\Lambda)(V'dV), \tag{4.3.3}$$

$$c_U = \int (U'dU) = \frac{2^k \pi^{nsk/2}}{\Gamma_k(\frac{np}{2})}, \tag{4.3.4}$$

$$c_C = \int (C'dC) = \frac{2^k \pi^{k^2/2}}{\Gamma_k(\frac{k}{2})}, \tag{4.3.5}$$

$$\Gamma_k(\frac{m}{2}) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\Big[\frac{m-i+1}{2}\Big], \tag{4.3.6}$$

where $\text{tr}(M)$ is the trace of a square matrix M. The constants c_U, c_C, c_Λ are normalising constants to ensure the priors of U, Λ, V are all proper. The forms of c_U, c_C are obtained from James (1954, pp. 58-59) while c_Λ is defined in (4.A.5). The matrix Ω can be fixed at the identity matrix for a prior that is invariant to ordering only, or equal to Σ for a prior that is invariant to both ordering and scale transformation.

Both U and V are subject to the orthogonal constraints which complicate the procedure of sampling the parameters. To overcome this issue, we make use of parameter expansions developed in Chan *et al.* (2017) to map (U, Λ, V) to (A, B) in a real space. The parameter expansion approach is also used in a number of studies to obtain the efficient computation (see Ghosh and Dunson (2009), Koop *et al.* (2010), Chan *et al.* (2017)).

The mapping from (U, Λ, V) to (A, B) is established by introducing two parameter expansions. The first parameter expansion is an orthogonal matrix $C \in \mathcal{O}(k)$ (i.e. $C'C = I_k$). The second is a $k \times k$ upper positive definite matrix κ. The mapping of (U, Λ, V) and (A, B) is described as

$$\begin{aligned} U\Lambda V &= (UC')(C\Lambda V), \quad \text{where } C \in \mathcal{O}(k) \\ &= B^*A^* = (B^*\kappa)(\kappa^{-1}A) = BA. \end{aligned}$$

The parameter expansions C, κ are non-identified parameters, i.e. they do not change the likelihood function and are not identified under the joint posterior density.

The choice of the prior $p(\Lambda, V|c_\lambda, \Sigma)$ in (4.3.2) implies that the marginal prior of A (integrating out B) is a matrix variate t-distribution with the degree of freedom $ns - n - k + 1$, means of zero, and the variance matrices $\left(\Sigma, I_k \frac{1}{c_\lambda(ns-n-k-1)}\right)$ (see Gupta and Nagar (2000, p.134) for the definition of a matrix variate t-distribution). The shrinkage parameter c_λ is to permit the shrinkage of A in (4.2.2) towards zero. We estimate the shrinkage parameter via a data-driven approach, i.e. we place a Gamma prior on the shrinkage parameter c_λ, i.e. $c_\lambda \sim G(\underline{\alpha}, \underline{\beta})$, where G denotes a Gamma distribution.

The priors of A and B are given in the following theorem.

Theorem 4.3.1 *The prior* (U, Λ, V) *in* (4.3.1)-(4.3.6) *implies that the priors for* A *and* B *are* ***for the case*** $s = 1$*:*

$$p(B|A, \Sigma) \sim N(\mathbf{0}, (c_\lambda A \Sigma^{-1} A')^{-1} \otimes I_{ns}),$$
$$p(A|B) \sim N\left(\mathbf{0}, \frac{1}{c_\lambda}\Sigma \otimes (B'B)^{-1}\right);$$

for the case $s > 1$*:*

$$p(B|A, \Sigma) \sim N(\mathbf{0}, (I_k + c_\lambda A \Sigma^{-1} A')^{-1} \otimes I_{ns}),$$
$$p(A|B) \sim N\left(\mathbf{0}, \frac{1}{c_\lambda}\Sigma \otimes (B'B)^{-1}\right).$$

Proof: Our proof is presented in Appendix 4.A.1. The proof is based on a number of results in Chan *et al.* (2017).

4.3.2 The Posteriors

In this section, we present the MCMC algorithm to estimate parameters $(A, B, \Sigma^{-1}, c_\lambda)$. For notation convenience, we stack all the observations over time periods and denote $Y = (Y_1', \ldots, Y_T')'$, $X = (X_1, \ldots, X_T)$, and rewrite the reduced rank regression model as follows

$$Y = XBA + \varepsilon \qquad \varepsilon \sim N(0, \Sigma \otimes I_T). \tag{4.3.7}$$

The likelihood function of the model is :

$$p(Y|A, B, \Sigma) \propto |\Sigma|^{-T/2} \exp -\frac{1}{2}\text{tr}\left(\Sigma^{-1}(Y - XBA)'(Y - XBA)\right). \tag{4.3.8}$$

Following Bayes'rule, the posterior is obtained by incorporating the priors and the likelihood function:

$$p(c_\lambda, A, B, \Sigma^{-1}|Y) \propto p(A, B|c_\lambda, \Sigma)p(c_\lambda)p(\Sigma^{-1})p(Y|A, B, \Sigma). \tag{4.3.9}$$

The conditional posterior densities of $(A, B, \Sigma^{-1}, c_\lambda)$ are presented below.

Sampling parameter A

The conditional posterior of A

$$p(\text{vec}(A)|B, \Sigma, c_\lambda, Y) \propto N(\text{vec}(\overline{A}), \overline{V_A}),$$

where vec(A) is vectorisation of a matrix A

$$\begin{aligned} \overline{V_A} &= \Sigma \otimes P_A^{-1}, \text{where } P_A = c_\lambda B'B + B'X'XB, \\ \overline{A} &= P_A^{-1}B'X'Y. \end{aligned} \tag{4.3.10}$$

To compute the mean of $\overline{A}$, we first take a Cholesky decomposition C_{P_A} of P_A, i.e. $C'_{P_A}C_{P_A} = P_A$ and then follow the below procedure

$$\begin{aligned} P_A\overline{A} &= B'X'Y, \\ C'_{P_A}C_{P_A}\overline{A} &= B'X'Y, \\ C_{P_A}\overline{A} &= (C'_{P_A})^{-1}B'X'Y, \\ \overline{A} &= C_{P_A}^{-1}Z_A, \text{ where } Z_A = (C'_{P_A})^{-1}B'X'Y. \end{aligned} \tag{4.3.11}$$

A conventional way to sample A from its conditional posterior is

$$\text{vec}(A) = \text{vec}(\overline{A}) + \text{chol}(V_A)z,$$

where chol(V_A) denotes Cholesky decomposition of a matrix V_A, z is a $nk \times 1$ vector of of random variables drawn from $N(0, I_{nk})$. To speed up the computation time of sampling A when sample size n increases, we use the computation algorithm (considered in Carriero *et al.* (2016b)) which exploits the Kronecker structure of the covariance matrix rather than taking the inverse matrix V_A, i.e.

$$A = \overline{A} + C_{P_A}^{-1}ZC_\Sigma, \tag{4.3.12}$$

where Z is a $k \times n$ matrix of independent $N(0, I_{nk})$ random variables. The expression in (4.3.12) implies that $a = \text{vec}(A) \sim N(\text{vec}(\overline{A}), \Sigma \otimes P_A^{-1})$ (see Bauwens and Richard (2000, pp. 301-302)). For a given matrix $r = ns + 1$, the time complexity implied by (4.3.12) is of order

$O(n^3s^3)$ while the conventional approach is of order $O(n^6s^3)$.

Sampling parameter B

The conditional posterior of B

$$p(\text{vec}(B)|A, \Sigma, c_\lambda, Y) \propto N(\text{vec}(\overline{B}), \overline{V_B}),$$

where

$$\begin{aligned}
\overline{B} &= \overline{V}_B X'Y\Sigma^{-1}A' \\
\overline{V_B} &= [A\Sigma^{-1}A' \otimes (X'X + c_\lambda I_{ns})]^{-1} \quad \text{for the case } s = 1 \\
\overline{V_B} &= [A\Sigma^{-1}A' \otimes (X'X + c_\lambda I_{ns}) + I_{nsk}]^{-1} \quad \text{for the case } s > 1
\end{aligned}$$

Since the covariance matrix $\overline{V}_B$ does not have a Kronecker structure for $s > 1$, a conventional approach is applied to draw $\text{vec}(B)$ from its normal conditional posterior $N(\text{vec}(\overline{B}), \overline{V_B})$, i.e. $\text{vec}(B) = \text{vec}(\overline{B}) + C_{V_B}^{-1}z$, where z is a $nsk \times 1$ matrix of independent $N(0, I_{nsk})$.

Sampling parameter Σ

The conditional posterior of Σ^{-1}

$$p(\Sigma^{-1}|A, B, c_\lambda, Y) \propto W(\overline{S}, \overline{v}),$$

where $\overline{v} = T + k + v_0$, $\overline{S} = \Big[c_\lambda A'B'BA + (Y - XBA)'(Y - XBA) + S_0^{-1}]^{-1}$.

Sampling parameter c_λ

The conditional posterior of shrinkage parameter c_λ is

$$p(c_\lambda|A, B, c_\lambda, Y) \propto G(\overline{\alpha}, \overline{\beta}),$$

where $\overline{\alpha} = \underline{\alpha} + \frac{1}{2}\text{tr}(\Sigma^{-1}A'B'BA)$, $\overline{\beta} = \underline{\beta} + \frac{nk}{2}$.

4.4 Determining the Rank

One of the common approaches for rank selection of a matrix is based on a sequential test. A sequential test is well-developed in the frequentist framework and in a small set-up (see Gill

and Lewbel (1992), Engle and Kozicki (1993), Cragg and Donald (1997), Robin and Smith (2000), Camba-Mendez *et al.* (2003), and references therein). The test statistics in the previous studies require a root-T consistent estimator of the coefficient regression matrix (Φ in our context). Information criteria such as Akaike (AIC), Hannan-Quinn (HQ) and Schwarz (SC) is an alternative way of selecting k (see Vahid and Issler (2002), Lütkepohl (2007)). Information criteria can be used to jointly select the rank k and the number of lags s.

In Bayesian analysis, a Bayes factor or a marginal likelihood is a natural way to select the rank. In the approach, a range of RRR models with several possible values of k is considered. The marginal likelihood of each model is estimated, and the model with the largest marginal likelihood is chosen. Similar to information criteria, the number of lags can be jointly selected with the rank. Comparing the performance of information criteria and marginal likelihood is worth to be investigated.

The marginal likelihood, under model $\mathscr{M}_l$ with rank k_l, is calculated by integrating the joint density of the observables (Y) and parameters (θ_l) with respect to θ_l, i.e.

$$p(Y|k_l) = p(Y|\mathscr{M}_l) = \int p(Y, \theta_l|\mathscr{M}_l)d\theta_l = \int p(Y|\theta_l, \mathscr{M}_l)p(\theta_l|\mathscr{M}_l)d\theta_l, \tag{4.4.1}$$

where $\theta_l = [A_l, B_l, \Sigma_l^{-1}]$, $p(Y|\theta_l, \mathscr{M}_l)$ is the likelihood function of model $\mathscr{M}_l$, $p(\theta_l|\mathscr{M}_l)$ is the prior of θ_l.

In our context an analytical solution of the marginal likelihood often does not exist. We hence approximate the marginal likelihood by using a number of techniques such as adaptive importance sampling (Chan and Eisenstat (2015)), predictive likelihood (Geweke and Amisano (2010)), and Laplace approximation (Tierney and Kadane (1986)). The approaches have been applied in a low dimension set-up, however, their performance in a high dimensional set up is not quite clear. Their performance in selecting the rank is investigated in Section 4.5, while this section discusses the approaches.

4.4.1 Cross-Entropy Method

We first integrate the joint density of Y and unknown parameters $[A, B, \Sigma^{-1}]$ with respect to $[A, \Sigma^{-1}]$. The joint density of Y and the remaining parameter B is stated in Theorem 4.4.1.

Theorem 4.4.1 *The joint density of Y and B conditional on model $\mathscr{M}$ is*[2]

[2]For notation convenience, we suppress the subscript l in k here.

For the case $s = 1$

$$p(B, Y|k_l) = p(B, Y|\mathscr{M}_l) = c_1^*|B'PB|^{\frac{T+v_0-n}{2}}|B'QB|^{-\frac{T+v_0}{2}}, \tag{4.4.2}$$

where $c_1^* = c_\lambda^{\frac{nk}{2}}\pi^{\frac{-nsk-nT}{2}}\Gamma_k(\frac{ns}{2})\frac{\Gamma_n(\frac{T+v_0}{2})}{\Gamma_n(\frac{v_0}{2})}|S_0|^{-\frac{v_0}{2}}$; $P = c_\lambda I_{ns} + X'X$; $Q = P - X'Y(S_0^{-1} + Y'Y)^{-1}Y'X$.

For the case $s > 1$

$$p(B, Y|k_l) = p(B, Y|\mathscr{M}_l) = c_1|B'PB|^{\frac{T+v_0-n}{2}}|B'QB|^{-\frac{T+v_0}{2}}\exp -\frac{1}{2}tr(B'B), \tag{4.4.3}$$

where $c_1 = c_\lambda^{\frac{nk}{2}}2^{\frac{-nsk+nk}{2}}\pi^{\frac{-nsk-nT}{2}}\frac{\Gamma_k(\frac{ns}{2})}{\Gamma_k(\frac{ns-n}{2})}\frac{\Gamma_n(\frac{T+v_0}{2})}{\Gamma_n(\frac{v_0}{2})}|S_0|^{-\frac{v_0}{2}}$; $P = c_\lambda I_{ns} + X'X$; $Q = P - X'Y(S_0^{-1} + Y'Y)^{-1}Y'X$.

Proof. Our proof is presented in Appendix 4.A.2.

The marginal likelihood under model $\mathscr{M}_l$ is calculated as

$$p(Y|k_l) = p(Y|\mathscr{M}_l) = \int p(B, Y|k_l)dB. \tag{4.4.4}$$

Such a marginal likelihood of (4.4.2) can be approximate by an importance sampling estimator:

$$p(Y|k_l) = \int \frac{p(B, Y|k_l)}{g(B)}g(B)d(B) \approx \frac{1}{M}\sum_{m=1}^{M}\frac{p(B^{(m)}, Y|k_l)}{g(B^{(m)})} = \hat{p}(Y|k_l), \tag{4.4.5}$$

where $B^{(m)}$, $m = 1, \ldots, M$, are draws from an importance density $g(B)$ which dominates $p(B, Y|k_l)$, i.e. $g(B) \geq p(B, Y|k_l)$. The performance of the estimator, particularly its variance, is critically dependent upon the choice of the importance density $g(B)$. In our context, the optimal choice of $g(.)$, denoted as $g^*(.)$, is $p(Y|B, k)p(B)/p(Y|k)$. The optimal choice is rarely obtained as it contains the marginal likelihood term, $p(Y|k)$, which we want to estimate. The idea underpinning the cross entropy approach is to search for a proposed importance density "close" enough to the optimal density g^*. The closeness between the proposed density and the optimal density is defined by the cross-entropy distance (also known as Kullback-Leibler divergence distance).[3]

Let us denote a parametric family $\mathscr{T} \equiv \{g(B; v)\}$ which is indexed by some parameter vector v that we want to locate importance density, e.g. a mean and a variance. The cross-entropy

[3]Further discussion about the cross-entropy method is referred to in Rubinstein and Kroese (2004).

distance is then defined as

$$\mathcal{D}(g^*, g) = \int g^*(B) \log \frac{g^*(B)}{g(B; v)} dB. \tag{4.4.6}$$

The closer $g(.)$ to $g^*(.)$ implies that the Kullback-Leibler is minimised, which is equivalent to

$$v^*_{CE} = \arg\max_v \int g^*(B) \log g(B; v) dx.$$

In our context, it is

$$v^*_{CE} = \arg\max_v \int p(B, y|k) \log g(B; v) dB \approx \arg\max_v \frac{1}{M} \sum_{m=1}^{M} \log(g(B^{(m)}; v),$$

where $B^{(m)}$, $m = 1, \ldots, M$, are draws from the conditional posterior density using the MCMC algorithm.

Our choice for $g(B; v)$ is based on the following criteria i) generating random samples from a parametric family should be easy and direct, and ii) the optimisation should be easily solved either analytically or numerically. We choose a matrix variate t proposal density, i.e. $B \sim g(B) \equiv t(df, \hat{B}, \hat{V}_B, I_k)$ with degree of freedom df, location vector $\hat{B}$, and scale matrix $\hat{V}_B$, and I_k:

$$\hat{B} = \frac{1}{M} \sum_{m=1}^{M} B^{(m)}; \; \hat{V}_B = \frac{1}{kM} \sum_{m=1}^{M} (B^{(m)} - \hat{B})(B^{(m)} - \hat{B})'.$$

Given the proposal density $g(.)$, we then apply the importance sampling estimator in (4.4.5) to evaluate the marginal likelihood. The procedure approximating the marginal likelihood is summarised in the following steps:

1. Use the outputs from the MCMC described in Section 4.3 to obtain the location vector $\hat{B}$ and the scale matrix $\hat{V}_B$. Then obtain the Cholesky decomposition of $\hat{V}_B$, denoted as C_B, ($\hat{V}_B = C'_B C_B$).

2. Sample $u \sim N(0, I_{nkn})$ and $r \sim G(df/2, df/2)$. Then set $v \equiv u/\sqrt{r} \sim t(df, 0, I_{nkn})$.

3. Solve $C_B x = v$ for x, i.e. $x = C_B^{-1} v$, and then take $B = \hat{B} + x \sim t(df, \hat{B}, \hat{V}_B, I_k)$.

4. Use the draws from step 3 to evaluate the marginal likelihood (4.4.5).

4.4.2 Predictive Likelihood Method

The second method of evaluating the marginal likelihood we consider here is predictive likelihood. The method has been applied in a number of studies such as Geweke (1996), Geweke (2001), Geweke and Amisano (2010), Warne *et al.* (2017), etc.

To facilitate the discussion in the section, we denote the dataset at period T as $Y_T = \{y_1, \ldots, y_T\}$ (the subscript T is introduced in Y). For the dataset Y_T, the marginal likelihood of the model $\mathscr{M}_l$ can alternatively be decomposed as the product of the predictive likelihood, i.e.

$$p(Y_T|\mathscr{M}_l) = \prod_{t=1}^{T} p(y_t|Y_{t-1}, \mathscr{M}_l), \tag{4.4.7}$$

where $p(y_t|Y_{t-1}, \mathscr{M}_l)$ is one-step ahead predictive likelihood, evaluated at time t

$$\begin{aligned} p(y_t|Y_{t-1}, \mathscr{M}_l) &= \int p(y_t|Y_{t-1}, \theta_l, \mathscr{M}_l) p(\theta_l|Y_{t-1}, \mathscr{M}_l) d\theta_l, \\ &\approx (1/M) \sum_{m=1}^{M} p(y_t|Y_{t-1}, \theta_l^{(m)}, \mathscr{M}_l), \end{aligned}$$

The cost of this approach is that the posterior simulator is computed at each time period t.

The corresponding generalization of (4.4.7) is

$$p(Y_T|Y_S, \mathscr{M}_l) = \prod_{t=S+1}^{T} p(y_t|Y_{t-1}, \mathscr{M}_l). \tag{4.4.8}$$

for $S < T$, $Y_S = \{y_1, \ldots, y_s\}$. Equation (4.4.8) is identical to equation (4.4.7) when $S = 0$. The right hand side of (4.4.8) is the ratio of the accumulation of the predictive likelihood function, which starts at period $t = S + 1$ rather than $t = 1$.

4.4.3 Laplace Approximation Method

The application of Laplace approximation for evaluating integrals over a space (e.g. Lebesgue space) is often straightforward when the mode and the Hessian matrix of a given kernel function are easily obtained (see Tierney and Kadane (1986), Breslow and Lin (1995), Chickering and Heckerman (1996)). The approach provides a faster computation than the MCMC and numerical integration methods. However, the approach does not perform well if the kernel function is not smooth and is not peaked around the mode.

Instead of evaluating the marginal likelihood $p(Y|k) = \int p(B, Y|k)dB$ stated in Theorem 4.4.1, we evaluate the marginal likelihood $\int p(\tilde{B}, Y|k)(dv_{np}^k)$, where dv_{np}^k is the measure of the Stiefel manifold (its form is discussed later). It is noted that the two marginal likelihood expressions are equivalent. The key difference between the expressions is that $\tilde{B}$ does not contain the parameter expansions C, κ while B does. As discussed in Section 4.3.1, the parameter expansions do not enter to the likelihood function, and are not identified under the posterior density. Evaluating the marginal likelihood using either of the two expressions should provide similar results. The reason we choose to evaluate $\int p(\tilde{B}, Y|k)(dv_{ns}^k)$ is that the mode and the negative Hessian matrix of the joint density $p(\tilde{B}, Y|k)$ are easier to derived.

The space we integrate the integral $p(\tilde{B}, Y|k)$ is over the Stiefel manifold, $\mathcal{V}_{k,ns}$, and therefore standard rules are not directly applied to calculate the integral. More specifically, for any matrix $ns \times k$ $E \in \mathcal{V}_{k,ns}$, there are only $ns - k(k+1)/2$ free elements in E, and therefore, if one wants to compute the second differentials of a given function with respect to a matrix E (i.e. the Hessian matrix), this must be subject to $ns - k(k+1)/2$ free elements in E. Due to this, the calculation of differentials over the Stiefel manifold becomes more challenging. Our derivations of the mode and the negative Hessian matrix are built upon a number of results from James (1954), Strachan and Inder (2004), Villani (2005), Muirhead (2005, Chapter 9), and Magnus and Neudecker (2007, Chapter 9-10).

Let recall the SVD of the matrix $\Phi = U\Lambda V'$, where $U \in \mathcal{V}_{k,ns}, V \in \mathcal{V}_{k,n}$. It is known that for a matrix $U \in \mathcal{V}_{k,ns}$, it can be expressed as $U = \tilde{B}\overline{C}$, where $\overline{C} \in \mathcal{O}(k)$ (see Chikuse (2003)). This implies that we rotate the vector of $\tilde{B}$ within the plane $\mathfrak{b} = sp(\tilde{B})$, which is in the Grassmann manifold, $\mathcal{G}_{k,ns}$.[4] The matrix U spans the same space as $\tilde{B}$, i.e. $\mathfrak{b} = sp(\tilde{B}) = sp(U)$. Its measure, prior, and posterior are preserved under the transformation (see James (1954)). The matrix can then be re-parametrised as

$$\begin{aligned} \Phi &= \tilde{B}\overline{C}\Lambda V' \\ &= \tilde{B}\tilde{A}, \qquad \text{where } \tilde{A} = \overline{C}\Lambda V'. \end{aligned} \tag{4.4.9}$$

In other words, $sp(\tilde{B})$ has a uniform prior over the Grassmann manifold. The idea of doing the translation is to map (U, Λ, V) to $(\tilde{A}, \tilde{B})$, and to obtain a standard form for the prior of A.

The forms of the priors, the likelihood function, the posteriors, and the joint density of $\tilde{B}$ and

[4] The Grassmann manifold is defined as the space whose points are k-planes (k-dimensional hyperplanes containing the origin) $\mathcal{V}_{k,np}$ (see James (1954) or Chikuse (2003) for further discussion about the Grassmann manifold, Stiefel manifold, and the orthogonal manifold).

Y are presented in the two following theorems. The derivations of the priors and posteriors are similar to these presented in Section 4.3 and therefore we omit.

Theorem 4.4.2 *Given the priors of the parameters U, Λ, V as described in Section 2, the prior of $\tilde{B}$ is a uniform, the prior of $\tilde{A}$ is normal, and the prior of Σ^{-1} is Wishart:*

$$\frac{p(\tilde{A}|c_\lambda,\Sigma)p(\Sigma^{-1})(d\tilde{A})(\tilde{B}'d\tilde{B})}{c_\Lambda c_{\tilde{B}} c_{\Sigma^{-1}}},$$

$$p(\tilde{A}|c_\lambda,\Sigma) \propto \exp\Big\{-\frac{c_\lambda}{2}tr(\Sigma^{-1}\tilde{A}'\tilde{A})\Big\},$$

$$c_\Lambda = \int p(\tilde{A}|c_\lambda,\Sigma)(d\tilde{A}) = \Big(\frac{2\pi}{c_\lambda}\Big)^{nk/2}|\Sigma|^{k/2},$$

$$c_{\tilde{B}} = \int \tilde{B}'d(\tilde{B}) = \frac{\pi^{\frac{k(ns-k)}{2}}\Gamma_k(k/2)}{\Gamma_k(np/2)},$$

$$p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{\frac{v_0-n-1}{2}}\exp\Big\{-\frac{1}{2}trS_0^{-1}\Sigma^{-1}\Big\},$$

$$c_{\Sigma^{-1}} = \int p(\Sigma^{-1})d\Sigma^{-1} = 2^{nv_0/2}\Gamma_n(\frac{v_0}{2})|S_0|^{v_0/2}.$$

The likelihood function is

$$p(y|\tilde{A},\tilde{B},\Sigma^{-1}) = (2\pi)^{-Tn/2}|\Sigma|^{-T/2}\exp\Big\{-\frac{1}{2}tr\Sigma^{-1}(Y-X\tilde{B}\tilde{A})'(Y-X\tilde{B}\tilde{A})\Big\}.$$

The posterior density is

$$p(\tilde{A},\tilde{B},\Sigma^{-1}|Y) \propto c_2|\Sigma^{-1}|^{\frac{T+v_0+k-n-1}{2}}\exp\Big\{-\frac{c_\lambda}{2}tr\Sigma^{-1}\tilde{A}'\tilde{A}\Big\}\exp\Big\{-\frac{1}{2}trS_0^{-1}\Sigma^{-1}\Big\}$$
$$\times\exp\Big\{-\frac{1}{2}tr\Sigma^{-1}(Y-X\tilde{B}\tilde{A})'(Y-X\tilde{B}\tilde{A})\Big\}, \tag{4.4.10}$$

where $c_2 = c_\lambda^{nk/2}2^{-nk/2}\pi^{-k(n+ns+k)/2}\frac{\Gamma_k(ns/2)}{\Gamma_k(k/2)}$.

Theorem 4.4.3 *The joint density of $\tilde{B}$ and Y conditional on model $\mathscr{M}$ is*

$$p(\tilde{B},Y|k) = p(\tilde{B},Y|\mathscr{M}) \propto h(\tilde{B}|Y,k)$$
$$= c_k|\tilde{B}'P\tilde{B}|^{\frac{T+v_0-n}{2}}|\tilde{B}'Q\tilde{B}|^{-\frac{T+v_0}{2}}, \tag{4.4.11}$$

where $h(\tilde{B}|Y,k) = c_k|\tilde{B}'P\tilde{B}|^{\frac{T+v_0-n}{2}}|\tilde{B}'Q\tilde{B}|^{-\frac{T+v_0}{2}}$; $c_k = (c_\lambda)^{\frac{nk}{2}}\pi^{\frac{-kns+k^2}{2}}\frac{\Gamma_k(ns/2)}{\Gamma_k(k/2)}$; $P = c_\lambda I_{ns} + X'X$; $Q = P - X'Y(S_0^{-1}+Y'Y)^{-1}Y'X$.

Proof: The proof follows the same procedure as described in Theorem (4.4.1).

The marginal likelihood $p(Y|k)$ can be computed by integrating $\tilde{B}$ out of $p(\tilde{B}, Y|k)$ over the Stiefel manifold. That is,

$$p(Y|k) = \int_{\mathcal{G}_{k,ns}} p(\tilde{B}, Y|k)(dg_{ns}^k) = \int_{\mathcal{G}_{k,ns}} h(\tilde{B}|Y,k)(dg_{ns}^k),$$

where $\tilde{B} = [\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_k]$, $dg_{ns}^k = \mathbf{\Lambda}_{w=1}^{k} \mathbf{\Lambda}_{j=ns-k+1}^{np} \tilde{b}_w' d\tilde{b}_s$, $\mathbf{\Lambda}$ denotes exterior product differential forms (see Muirhead (2005, Chapter 2)).

It is known that an integral of $h(\tilde{E})$ with respect to $\mathcal{G}_{k,ns}$ can be obtained by integrating over the Stiefel manifold $\mathcal{V}_{k,ns}$ with the adjustment by the volume of the orthogonal group, Vol$(\mathcal{O}_k)$ (see Strachan and Inder (2004, p.315)). In our case, that is

$$\int_{\mathcal{G}_{k,ns}} h(\tilde{B}|Y,k) dg_k^{ns} = \frac{\int_{\mathcal{V}_{k,np}} h(\tilde{B}|Y,k) dv_k^{ns}}{\int_{\mathcal{O}_k} dv_k^k}, \tag{4.4.12}$$

where $k(.) = c_k f^{T+v_0} g$, $f = |\tilde{B}' P \tilde{B}|^{\frac{1}{2}} |\tilde{B}' Q \tilde{B}|^{-\frac{1}{2}}$, $g = |\tilde{B}' P \tilde{B}|^{\frac{-n}{2}}$, $dv_k^{np} = \mathbf{\Lambda}_{w=1}^{ns} \mathbf{\Lambda}_{j=w+1}^{ns} \tilde{b_w} d\tilde{b}_s$, $dv_k^k = \mathbf{\Lambda}_{w=1}^{k} \mathbf{\Lambda}_{j=w+1}^{k} \tilde{b_w} d\tilde{b}_s$.

Following the Poincaré theorem, for $\tilde{B} \in \mathcal{V}_{k,ns}$, then $\prod_{j=1}^{k} \lambda_{ns-k+j}(W) \leq |\tilde{B}' W \tilde{B}| \leq \prod_{j=1}^{k} \lambda_j(W)$, where $W = P, Q$; $\lambda_j(W)$ are eigenvalues with positive and finite values. The kernel density $r(.)$ therefore has a finite upper bound. These are sufficient conditions to ensure that the conditional posterior for $\tilde{B}$ is proper and all finite moments (see Strachan and Inder (2004) for further discussion).

To this end, the marginal likelihood in (4.4.12) can be approximated by

$$c_k g(\hat{B}^*) f^{T+v_0}(\hat{B}^*) \Big(\frac{2\pi}{T+v_0} \Big)^{(ns-k)k/2} |\Psi|^{-1/2},$$

where $\hat{B}^*$ is the mode of the function f, Ψ is the negative Hessian matrix of $\ln(f)$. The forms of the mode and the negative Hessian matrix are presented in 4.B.

4.5 Monte Carlo Experiment

In this section we perform extensive Monte Carlo simulation and wish to investigate under what circumstances cross entropy (CE), predictive likelihood and Laplace approximation fail to select the correct rank, and whether the forecast performance of the model is affected by a misspecified rank.

4.5.1 Monte Carlo Design

The Monte Carlo design is based on the following data generating process (DGP)

$$\begin{aligned} Y_t &= Y_{t-1}\Phi_1 + Y_{t-2}\Phi_2 + \ldots + Y_{t-s}\Phi_s + \varepsilon_t \qquad \varepsilon_t \sim N(0,\Sigma), \\ &= X_t\Phi + \varepsilon_t, \\ &= X_t U\Lambda V' + \varepsilon_t, \end{aligned}$$

where the initial values of X_t in period $t = 1$ are randomly drawn from the standard normal distribution, i.e. $X_1 \sim N(0, I_n)$ with the number of variables $n = 25, 35,$ and 50. The number of time periods T is set at 250 and the number of lags s is 2. The inverse covariance matrix Σ of the random error term ε_t is drawn from a Wishart distribution, $\Sigma^{-1} \sim W(100I_n/n, n)$. The rank k of the coefficient Φ is set at $\{2, 3, 4, 5, 6\}$. Without loss of generality we set the semi-orthogonal matrices U and V as follows $U = \begin{bmatrix} I_k \\ 0_{ns-k,k} \end{bmatrix}$, $V = \begin{bmatrix} I_k \\ 0_{n-k,k} \end{bmatrix}$.[5] For the i-th element λ_i of the diagonal Λ ($\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$), we generate the elements following the DGP below[6]

$$\lambda_i = 0.97 - a(i-1), \tag{4.5.1}$$

where $i = 1, \ldots, k$. The slope a is chosen from a set of values which i) ensures that the VAR system is stable, ii) reflects how fast the magnitude of the singular values decreases. The purpose of generating such DGPs is to examine whether the magnitudes of the singular values play a role in the rank selection using CE, predictive likelihood and Laplace approximation approaches. For example, for the case $k = 5$, the slope a is chosen from the set $\{0.01, 0.025,\ 0.05, 0.075, 0.1, 0.125,\ 0.15, 0.175, 0.2\}$, which corresponds to the cases where the singular values decrease at a slow rate ($a = 0.01$, DGP1) and at a fast rate ($a = 0.2$, DGP9). The largest and smallest singular values in DGP1 ($a = 0.01$) are all close to 1, i.e. $\lambda_1 = 0.97, \lambda_2 = 0.96, \lambda_3 = 0.95, \lambda_4 = 0.94, \lambda_5 = 0.93$. On the other hand, the largest singular value of DGP9 ($a = 0.2$) is close to 1 while the smallest one approaches to 0, i.e. $\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$.

[5] One can generate the semi-orthogonal matrices by using the polar decomposition. For any $np \times k$ matrix of a random Z which has a density function, and has the rank of k almost everywhere, the unique polar decomposition of Z is $U_Z T_Z^{1/2}$, where $U_Z = Z(Z'Z)^{-1/2} \in \mathcal{V}_{k,ns}$ and $T_Z = Z'Z$. We also generate data using the approach and find that the conclusions are robust to our main Monte Carlo experiments.

[6] We also generate the singular values λ_i following the DGP: $\lambda_i = 0.97 - ai$. The DGP implies that the magnitudes of singular values change according to the slope a, whereas the first singular value is fixed regardless the value of a in (4.5.1). The main conclusions remain the same as they are obtained from our main Monte Carlo experiments.

Over all, the total number of DGPs for a given k and n is 9. For each DGP, we replicate 500 datasets. The total number of experiments is 4500 for a given k and n. The Monte Carlo experiments are performed on a Euramoo computer cluster at The University of Queensland.

4.5.2 The Performance of CE, Predictive Likelihood, and Laplace Approximation in Rank Selection

The common pattern we observe from the extensive Monte Carlo exercise is that both the dimensions of VAR systems and the magnitude of singular values affect the performance of the approaches. There is a downward bias in rank selection in a high dimensional set-up, and the downward bias is more distinct when the singular values of VAR coefficients are small (approaching to zero).

Our discussion below is not subject to the DGPs with the true rank of 5, but emerges from the simulation study we conduct. Table 4.5.1 reports the means and the standard errors of the estimated ranks (over 500 replicated data) for DGP1-DGP9 with the true rank $k = 5$ using CE, predictive likelihood, and Laplace approaches. Column (i) of the table reports the results for $n = 25$, column (ii) is for $n = 35$, and column (iii) is for $n = 50$. Figure 4.5.1 summarises the estimated rank $\hat{k}$ according DGPs. The results for other ks are presented in Table 4.D.1-Table 4.D.4.

Figure 4.5.1 shows that the estimated rank k appears to be underestimated when the slope of eigenvalues a increases (the magnitude of the singular values decreases at a faster rate). For instance, when the slope a is 0.01 which corresponds to DGP1 (all the eigenvalues are large and the smallest singular value λ_5 is 0.93), all of these approaches correctly estimate the true rank (e.g. 4.99 (0.12) for CE, 4.76 (0.58) for predictive likelihood, and 5 (0.00) for Laplace approximation given $n = 25$) . However, when the slope a is large (e.g. a= 0.2 associated with DGP9 where the magnitude of the smallest singular value is at 0.17), the estimated rank is much lower than the true value (e.g. 2.59 (0.53) for CE, and 2.25 (0.44) for Laplace approximation). The similar patterns are found for $n = 35$ and $n = 50$. Table 4.5.1 reveals that standard errors of the mean estimated k obtained from predictive likelihood approach are larger than those obtained from the others, which indicate more uncertainty in selecting the rank.

To investigate the performance of the approaches when the size of VAR (n) grows, we keep all DGPs the same, but increase the sample size from $n = 25$ to 35 and 50. Reading the results of each row across columns (i), (ii), and (iii) of Table 4.5.1, we see that the average mean of

Table 4.5.1: The means and standard errors (in parenthesis) of selected ranks for DGPs with $k = 5$

DGP	Methods	(i) n=25	(ii) n=35	(iii) n=50
DGP1	**CE**	4.99(0.08)	4.99(0.12)	4.90(0.29)
	Pred.Like	4.76(0.58)	4.64(0.68)	3.81(1.15)
	Laplace	5.00(0.00)	5.00(0.00)	5.00(0.04)
DGP2	**CE**	4.98(0.13)	4.91(0.28)	4.75(0.44)
	Pred.Like	4.53(0.85)	4.37(0.93)	3.42(1.24)
	Laplace	5.00(0.00)	5.00 (0.00)	4.99(0.13)
DGP3	**CE**	4.88(0.33)	4.69(0.46)	4.32(0.53)
	Pred.Like	4.12(1.05)	3.68(1.16)	2.53(1.11)
	Laplace	5.00(0.00)	5.00(0.00)	4.12(0.88)
DGP4	**CE**	4.60(0.49)	4.34(0.58)	3.77(0.54)
	Pred.Like	3.45(1.22)	2.98(1.21)	2.17(0.82)
	Laplace	5.00(0.00)	4.26(0.49)	2.69(0.60)
DGP5	**CE**	4.12(0.58)	3.57(0.54)	3.17(0.43)
	Pred.Like	2.87(1.13)	2.39(0.97)	1.83(0.64)
	Laplace	4.52(0.67)	3.16(0.64)	2.03(0.34)
DGP6	**CE**	3.69(0.58)	3.31(0.49)	2.93(0.42)
	Pred.Like	2.55(0.97)	2.21(0.80)	1.71(0.55)
	Laplace	3.33(0.54)	2.51(0.53)	1.81(0.40)
DGP7	**CE**	3.23(0.47)	2.84(0.51)	2.47(0.50)
	Pred.Like	2.10(0.76)	1.87(0.62)	1.59(0.53)
	Laplace	2.96(0.20)	2.01(0.25)	1.41(0.49)
DGP8	**CE**	2.95(0.51)	2.61(0.50)	2.23(0.42)
	Pred.Like	2.05 (0.79)	1.79 (0.63)	1.61(0.68)
	Laplace	2.80(0.40)	1.87 (0.33)	1.19(0.39)
DGP9	**CE**	2.59 (0.53)	2.27(0.45)	2.07(0.26)
	Pred.Like	2.05(0.79)	1.79(0.63)	1.64(0.77)
	Laplace	2.25(0.44)	1.73(0.44)	1.05(0.23)

DGP1: $\lambda_1 = 0.97, \lambda_2 = 0.96, \lambda_3 = 0.95, \lambda_4 = 0.94, \lambda_5 = 0.93$;
DGP2: $\lambda_1 = 0.97, \lambda_2 = 0.95, \lambda_3 = 0.93, \lambda_4 = 0.91, \lambda_5 = 0.89$;
DGP3: $\lambda_1 = 0.97, \lambda_2 = 0.93, \lambda_3 = 0.89, \lambda_4 = 0.85, \lambda_5 = 0.81$;
DGP4: $\lambda_1 = 0.97, \lambda_2 = 0.91, \lambda_3 = 0.85, \lambda_4 = 0.79, \lambda_5 = 0.71$;
DGP5: $\lambda_1 = 0.97, \lambda_2 = 0.88, \lambda_3 = 0.79, \lambda_4 = 0.70, \lambda_5 = 0.61$;
DGP6: $\lambda_1 = 0.97, \lambda_2 = 0.86, \lambda_3 = 0.75, \lambda_4 = 0.64, \lambda_5 = 0.53$;
DGP7: $\lambda_1 = 0.97, \lambda_2 = 0.82, \lambda_3 = 0.67, \lambda_4 = 0.52, \lambda_5 = 0.37$;
DGP8: $\lambda_1 = 0.97, \lambda_2 = 0.80, \lambda_3 = 0.63, \lambda_4 = 0.46, \lambda_5 = 0.29$;
DGP9: $\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$.

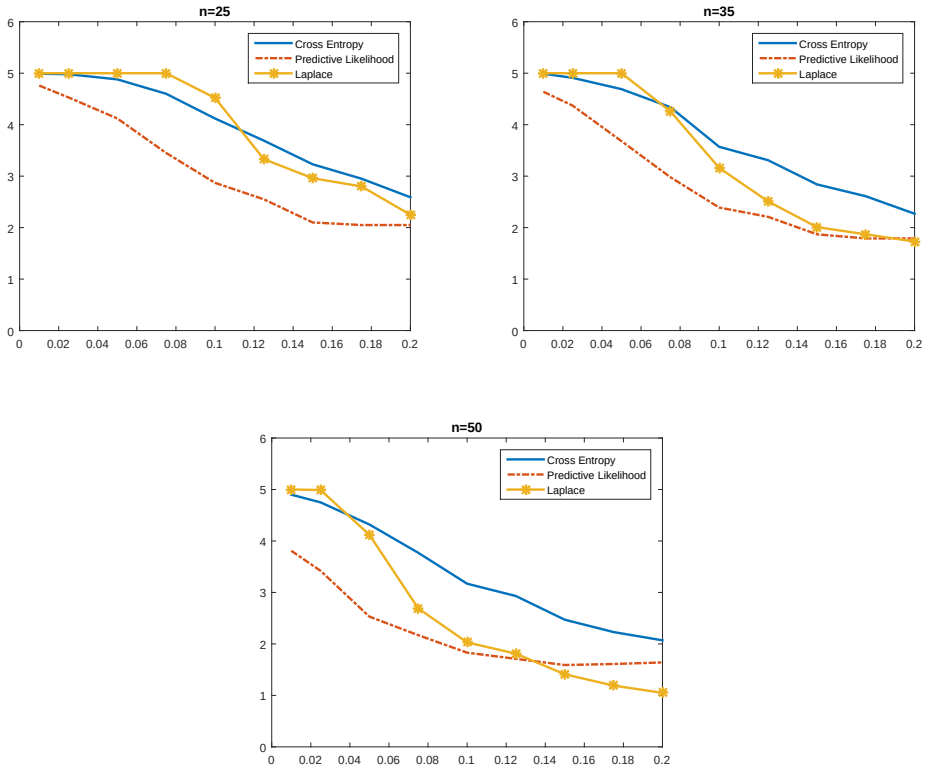


Figure 4.5.1: Estimated rank versus slope a

the estimated k tends to decrease as the size n grows. The rate of decrease in the estimate rank is faster when the magnitude of singular values is smaller. The conclusions hold when the true rank of the $VAR(n)$ system increases.

We also want to point out that the results in Table 4.5.1 and Tables 4.D.1-4.D.4 are based on DGPs where we ensure the existence of the mode of the kernel density considered in Laplace approximation. We go further and generate data without imposing the restriction to examine whether Laplace approximation still performs well. The results show that the number of times that the mode is not identified or the negative Hessian matrix is not a positive-definite matrix increases when the magnitudes of singular values are small, and the dimensions of VARs increase. For CE and predictive likelihood approaches, similar conclusions about the downward bias in the estimation of the rank still hold for such DGP without imposing the restrictions on the existence of a single mode and a positive definite matrix of a negative Hessian matrix.

With respect to computation time, the fastest computation time is Laplace, followed by CE and then predictive likelihood. This is because Laplace does not require MCMC iterations

to evaluate the marginal likelihood while the others do. Predictive likelihood requires fairly intensive computation as it involves the recursive forecast exercise. That is, at each time period we need to use MCMC iterations to evaluate a one-step predictive likelihood.

Overall, from the Monte Carlo simulation study we prefer CE approach for selecting the rank of the VAR coefficients, given its advantage in computation time and its accuracy in estimating the rank of VAR coefficients

4.5.3 Forecasting Performance

The rank selection results of Monte Carlo simulation exercises suggest that CE, predictive likelihood and Laplace approximation tend to underestimate the rank model of the VAR coefficients in a high dimensional set-up. A question arising is whether a misspecified rank provides better (or worse) forecast than the model with a true rank based on some point and density forecast measures.

Given the DGPs set-up in our Monte Carlo design, the first k variables of a VAR system are associated with the first non-negative k singular values, and the remaining $n - k$ variables are associated with the singular values of 0. We conduct and present the recursive forecast results for the first k variables and the first three variables of the remaining $n - k$ variables. Our aim is to investigate whether the forecast performance of RRR is associated with the magnitude of singular values.

The recursive forecast exercise is summarised as follows. We first use the sample including 203 periods to obtain the estimates used to predict outcomes for the periods from $T = 204$ ($h = 1$) to $T = 211$ ($h = 8$). For the next forecast, $T = 204$ is included in the sample, and the same procedure is repeated to predict outcomes for the periods $T = 205$ to $T = 212$. The procedure is conducted until $T = 242$.

To assess the forecast accuracy, researchers have proposed a number of criteria such as mean squared forecast errors (MSFE), median absolute forecast errors (MAFE), root mean squared forecast errors (RMSFE). Further discussion about some limitation of the point forecast measures in assessing forecast accuracy is referred to Hyndman and Koehler (2006). In this chapter we consider MSFE and MAFE which are are widely used in the literature (see Banbura *et al.* (2010), Carriero *et al.* (2011), Koop (2013) and many others). We report these forecast measures obtained from RRR with a misspecified rank relative to that obtained from a model with a correct rank (the benchmark model). For each variable i, the h-step-ahead point forecast

measures of MAFE and MSFE obtained from model j relative to that from the benchmark are

$$\text{MAFE}_{j,i,h} = \frac{\sum_{t=\underline{t}}^{\bar{t}-h} |\hat{y}_{j,i,t+h|t} - y_{i,t+h}|}{\sum_{t=\underline{t}}^{\bar{t}-h} |\hat{y}_{bm,i,t+h|t} - y_{i,t+h}|},$$
$$\text{MSFE}_{j,i,h} = \frac{\sum_{t=\underline{t}}^{\bar{t}-h} (\hat{y}_{j,i,t+h|t} - y_{i,t+h})^2}{\sum_{t=\underline{t}}^{\bar{t}-h} (\hat{y}_{bm,i,t+h|t} - y_{i,t+h})^2},$$

where $\underline{t}$ and $\bar{t}$ denote the start and the end of the out-of-sample period, $\hat{y}_{j,i,t+h|t}$, and $\hat{y}_{bm,i,t+h|t}$ are the predicted values of model j and the benchmark model for variable i at forecast horizon h using the information up to t.

The above forecast measures are designed to assess the forecast accuracy for a single variable. To assess the forecast accuracy for multivariate variables, we consider the multivariate loss function of Christoffersen and Diebold (1998), and calculate the multivariate weighted mean squared forecast error (WMSFE) of model j relative to that of the benchmark model for a h-step-ahead point forecast

$$\text{WMSFE}_{j,h} = \frac{\sum_{t=\underline{t}}^{\bar{t}-h} (e'_{j,t+h} \times W \times e_{j,t+h})}{\sum_{t=\underline{t}}^{\bar{t}-h} (e'_{bm,t+h} \times W \times e_{bm,t+h})}, \tag{4.5.2}$$

where $e_{j,t+h}$ and $e_{bm,t+h}$ are a $(k+3) \times 1$ vector of forecast errors of jointly $k+3$ variables, and W is the $(k+3) \times (k+3)$ matrix of weight. Following Carriero *et al.* (2011), we set the matrix W to be a diagonal matrix which features on the diagonal the inverse of the variance of the series to be forecast.

In Bayesian analysis another popular measure of forecast accuracy is forecast density (e.g. average log predictive likelihood (ALPL), and multivariate average log predictive likelihood (MVALPL)). A great advantage of the density forecast measures is that they evaluate the forecast performance of the entire predictive density rather that at a particular point estimate. The ALPL obtained from model j relative to that from the benchmark model is

$$\text{ALPL}_{j,h} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{t=\underline{t}}^{\bar{t}-h} (\text{LPL}_{j,i,t+h} - \text{LPL}_{bm,i,t+h}), \tag{4.5.3}$$

where $\text{LPL}_{j,i,t+h}$ and $\text{LPL}_{bm,i,t+h}$ are the log predictive scores obtained from model j and the benchmark model for variable i at time $t+h$.

The multivariate average log predictive likelihood from model j relative to that from the bench-

mark model is

$$\text{MVALPL}_{j,h} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{t=\underline{t}}^{\bar{t}-h} (\text{MVLPL}_{j,t+h} - \text{MVLPL}_{bm,t+h}), \tag{4.5.4}$$

where $\text{MVLPL}_{j,t+h}$ and $\text{MVLPL}_{bm,t+h}$ denote the multivariate log predictive likelihoods of model j and the benchmark at time period $t + h$.

To test if the point forecast and density forecast measures of a misspecified rank model are statistically different from these obtained from the benchmark (e.g. $H_0 : \text{MSFE}_j = \text{MSFE}_{bm}$, $H_1 : \text{MSFE}_j \neq \text{MSFE}_{bm}$), the modified Diebold and Mariano (1995)'s test statistic proposed by Harvey *et al.* (1997) is used. To make a decision on rejecting the null hypothesis, the modified Diebold and Mariano (1995)'s test statistic is compared with critical values from the Student's t distribution with $(T - 1)$ degree of freedom rather than the standard normal distribution.

We conduct the recursive forecast exercises for 500 replicated data for each of the nine DGPs. To summarize the forecast results over 500 replicated data for each DGP, we choose to report the 50% percentile of the estimated distribution of the metric measures. We also compute the 25% and 75% percentiles of the measures, and find the conclusions obtained from the percentiles are robust to those obtained from the 50% percentile.

The results presented below are obtained from a small subset of our extensive Monte Carlo simulation, DGP9 with the true rank $k = 5$.[7] The conclusions are not subject to the particular cases, but are the common patterns that are observed from our Monte Carlo simulation.

Table 4.5.2 reports average WMSFE and average MSFE over 500 replicated data for $k + 3$ forecast variables at a number of forecast horizons $h = \{1, 2, 4, 6, 7, 8\}$ for DGP9 with $k = 5$. The measures smaller than 1 indicate that the other models are preferred to the benchmark. The ***, **, and * indicate whether the forecast errors obtained from the models are statistically different from those obtained from the benchmark model at the 1%, 5%, and 10% level of significance, respectively.

Based on the point forecast measure of WMSFE for the $k + 3$ variables, the benchmark model outperforms the models with lower rank for shorter forecast horizons (e.g. $h = 1, 2$).[8] However, the models with lower rank perform as well as, or even better than the benchmark model for longer forecast horizons (e.g. $h = 4, 6, 8$). Similar results are found for MSFE for an individual variable. For the higher rank models, there is no statistical evidence suggesting that they

[7]The results for other DGPs are available upon the request.

[8]In our discussion a shorter forecast implies that $h < 4$, which is equivalent to a year forecast ahead using quarterly data, while a longer forecast horizon implies $h > 4$.

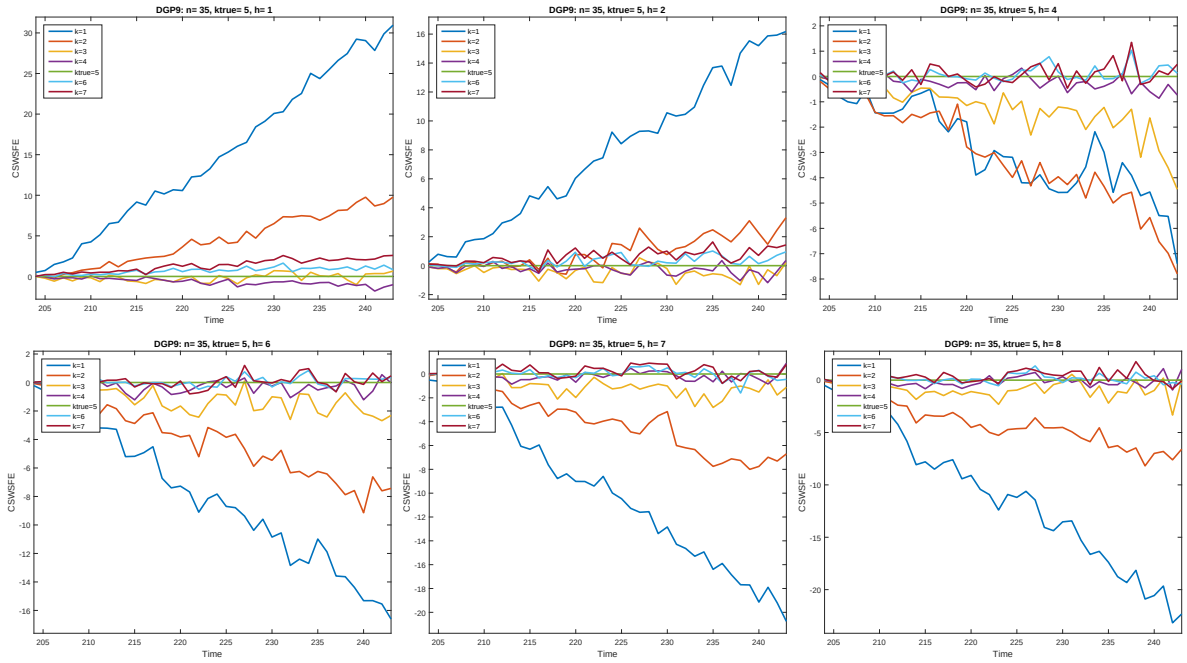


Figure 4.5.2: Cumulative sum of weighted of squared forecast errors differentials, $n = 35$, $k = 5$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$)

perform better or worse than the benchmark (see Table 4.5.2 for variables 6, 7, and 8).

All of the above measures only provide the forecast accuracy based on the average of the measures over the forecast periods, but not for each time forecast period. To better understand the forecast performance at each time forecast period, we plot the cumulative sum of WSFE differentials (CSWSFE) (Figure 4.5.2), and the cumulative sum of squared forecast error differentials (CSSFE) (Figures 4.D.1-4.D.4) during the forecast period T=204-242.[9] If the metric values are smaller than zero, this implies that the other RRR models are beating the benchmark. The figures confirm our findings in Table 4.5.2

The quality of the forecast density (AMVLPL), and average log predictive likelihood (ALPL) are reported in Tables 4.5.3 for DGP9, with $h = \{1, 2, 4, 6, 7, 8\}$ over 500 replicated datasets. The metric greater than zero suggests that the lower models provide better forecast than the benchmark model. The lower rank models seem to outperform the benchmark for a h-step ahead forecast with $h \geq 2$. The higher rank model appear to perform worse than the benchmark

[9]Due to space limitations, we present only the results for CSWSFE and CSSFE for some variables. The full results are available upon the request. The cumulative sum of WSFE differentials between model j and the benchmark is $\text{CSWSFE}_{j,h} = \sum_{t=\underline{t}}^{\bar{t}-h}(e'_{j,t+h}We_{j,t+h} - e'_{bm,t+h}We_{bm,t+h})$. The cumulative sum of SFE differentials between model j and the benchmark for variable j is $\text{CSSFE}_{i,j,h} = \sum_{t=\underline{t}}^{\bar{t}-h}(e^2_{i,j,t+h} - e^2_{i,bm,t+h})$.

Table 4.5.2: Out of sample point forecast performance (WMSFE and MSFE) of RRR with rank $k = \{1, 2, 3, 4, 6, 7\}$ relative to RRR with rank $k = 5$ across forecast horizons $h = \{1, 2, 4, 6, 7, 8\}$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17.$), $n = 35, ktrue = 5$

Variable	k=1	k=2	k=3	k=4	k=6	k=7	k=1	k=2	k=3	k=4	k=6	k=7
WMSFE/MSFE				**h=1**						**h=2**		
Multivariate	1.120***	1.038***	1.003	0.996	**1.003***	1.010**	1.052	1.011***	**1.001**	1.001	1.003	1.005
Var.1	0.984	**0.981**	0.992	0.989	0.998	1.000	**0.975**	0.990	0.999	0.998	1.003	1.004
Var.2	2.037***	1.016	0.999	**0.997**	1.002	1.010	1.413***	0.998	**0.994**	0.999	1.005	1.009
Var.3	1.291***	1.280***	1.024	1.005	**0.997**	0.995	1.056***	1.056***	**0.996**	0.998	1.005	1.005
Var.4	1.048**	1.052**	1.032	1.001	**0.990**	0.993	1.020**	1.005**	1.007	**1.002**	1.006	1.002
Var.5	0.990	**0.989**	0.996	1.000	1.007	1.011	1.001	**0.994**	0.998	1.000	1.002	1.003
Var.6	0.976	**0.972**	0.983	0.997	1.008	1.014*	1.000	**0.996**	1.000	0.998	1.000	1.002
Var.7	**0.975****	0.978**	0.979	0.989	1.005	1.014	0.990	**0.998**	0.998	0.999	1.000	1.002
Var.8	0.962**	**0.961****	0.970	0.982	1.005	1.016	1.014	0.999	**0.996**	0.996	0.999	1.003
				h=4						**h=6**		
Multivariate	0.979***	**0.977*****	0.987	0.998	1.000	1.001	**0.954*****	0.980***	0.994***	1.000	1.001	1.001
Var.1	0.987	**0.984**	0.991	0.999	1.004	1.003	**0.998**	1.006	1.004	1.006	0.996	1.001
Var.2	**0.949**	1.012	1.007	1.008	0.993	1.001	**0.797****	1.008	1.006	0.999	0.992	0.989
Var.3	0.952**	**0.946****	1.006**	0.996	1.004	1.006	**0.897*****	0.904***	1.013	0.991	0.991	0.999
Var.4	0.960***	**0.954*****	0.963***	1.000	1.006	1.007	**0.937*****	0.930***	0.951***	0.993	1.004	1.005
Var.5	0.998	0.979	**0.987**	0.995	1.000	0.997	1.001	**0.987**	0.997	1.003	1.006	1.006
Var.6	**0.998**	1.003	1.001	1.001	0.999	0.999	**1.001**	1.005	1.008	1.010	0.995	0.998
Var.7	**0.998**	1.005	1.001	1.003	0.996	0.999	**0.993**	1.000	1.000	1.002	0.998	0.995
Var.8	1.016***	**0.998**	0.994	0.998	0.999	0.996	1.021***	**0.995**	1.000	1.001	0.993	0.992
				h=7						**h=8**		
Multivariate	**0.944*****	0.982***	0.997	1.002	0.999	1.002	**0.941*****	0.983***	1.000	1.003	0.999	1.000
Var.1	0.991	0.993	**0.983**	0.990	0.994	1.002	1.007	1.012	0.995	1.005	**0.995**	0.997
Var.2	**0.761****	1.005	1.008	1.004	0.992	0.995	**0.751****	1.007	1.010	1.008	0.998	1.002
Var.3	**0.892*****	0.895***	1.005	0.993	0.999	1.000	**0.891*****	0.899***	1.003	0.992	1.002	0.998
Var.4	0.941***	**0.931*****	0.955***	0.995	1.011	1.013	0.945***	**0.931*****	0.954***	0.996	1.006	1.009
Var.5	1.008	**0.982*****	0.999	1.004	0.997	0.997	1.009	**0.983*****	0.992**	1.000	0.997	0.998
Var.6	**0.995**	1.002	1.002	1.003	1.003	0.999	0.999	1.001	1.000	1.001	**0.997**	0.998
Var.7	1.000	1.006	1.002	1.008	0.999	0.999	1.006	1.007	1.004	1.003	**0.998**	1.001
Var.8	1.019***	**0.995**	0.998	1.000	0.996	0.996	1.021***	0.999	1.003	1.006	**0.995**	0.996

Bold numbers indicate the lowest WMSFE and MSFE across all the models. * indicates significance at 10% level; ** indicates significance at 5% level; and *** indicates significance at 1% level.

Table 4.5.3: Out of sample density forecast performance (MVALPL and APLPL) of RRR with rank $k = \{1, 2, 3, 4, 6, 7\}$ relative to RRR with rank $k = 5$ across forecast horizons $h = \{1, 2, 4, 6, 7, 8\}$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17.$), $n = 35, ktrue = 5$

Variable	k=1	k=2	k=3	k=4	k=6	k=7	k=1	k=2	k=3	k=4	k=6	k=7
MVALPL/APLPL				**h=1**						**h=2**		
Multivariate	-2.446***	-1.928***	-1.210***	-0.533***	-1.158**	-2.180***	**6.830*** **	4.803***	3.224***	1.562***	-0.956	-1.848*
Var.1	-0.524***	-0.330***	-0.181***	-0.088***	-0.013*	-0.024**	**0.501*** **	0.386***	0.246***	0.072***	-0.018	-0.041
Var.2	-0.792***	-0.459***	-0.246***	-0.092***	-0.007	-0.018**	**0.403*** **	0.374***	0.256***	0.156***	-0.020	-0.041
Var.3	-0.703***	-0.706***	-0.412***	-0.231***	-0.013	-0.028**	**0.422*** **	0.324***	0.319***	0.100***	-0.021	-0.032
Var.4	-0.536***	-0.537***	-0.393***	-0.180***	-0.020	-0.038**	**0.517*** **	0.450***	0.227***	0.105***	-0.020*	-0.042*
Var.5	-0.554***	-0.544***	-0.589***	-0.499***	-0.019**	-0.032*	**0.596*** **	0.497***	0.437***	0.261***	-0.023	-0.034
Var.6	0.051***	0.044***	0.032***	0.017*	-0.029**	-0.065***	**0.039**	0.036***	0.032**	0.011*	-0.029	-0.055**
Var.7	0.056***	0.046***	0.029***	0.015*	-0.037**	-0.067***	**0.050***	0.038*	0.020	0.008	-0.031**	-0.049**
Var.8	0.051***	0.041***	0.029***	0.019*	-0.033***	-0.070***	**0.038*** **	0.041**	0.027	0.011	-0.017	-0.034**
				h=4						**h=6**		
Multivariate	**26.451*** **	19.217***	12.876***	6.217***	-0.697	-1.342	**48.425*** **	35.698***	21.990***	11.054***	-0.390	-1.188
Var.1	**2.433*** **	1.783***	1.095***	0.396***	-0.022	-0.066	**4.253*** **	3.015***	1.855***	0.790***	-0.006	-0.077
Var.2	**2.715*** **	2.049***	1.303***	0.554***	-0.010	-0.052	**4.537*** **	3.288***	1.914***	0.742***	0.056	-0.003
Var.3	**2.447*** **	2.132***	1.580***	0.809***	-0.020	-0.048	**4.129*** **	3.592***	2.564***	1.256***	-0.043	-0.020
Var.4	**2.644*** **	2.453***	1.841***	0.859***	-0.018*	-0.034*	**4.275*** **	3.964***	2.941***	1.170***	-0.042*	-0.077**
Var.5	**2.633*** **	2.472***	2.258***	1.625***	-0.061	-0.077	**4.253*** **	4.055***	3.685***	2.712***	0.009	-0.065
Var.6	**0.020**	0.013***	0.017**	0.003*	-0.007	-0.026**	**0.025**	0.013	0.015	0.000	-0.009*	-0.023**
Var.7	**0.039***	0.031	0.025	0.014	-0.017**	-0.036**	**0.048*** **	0.038	0.019	0.003	-0.009*	-0.030**
Var.8	**0.027*** **	0.014***	0.004	-0.003	-0.025	-0.038**	**0.027****	0.016	0.008	0.007	-0.012	-0.027**
				h=7						**h=8**		
Multivariate	**60.712*** **	44.889***	26.596***	12.929***	-0.406	-0.887	**73.473*** **	54.104***	32.617***	15.893***	-0.364	-0.966
Var.1	**4.967*** **	3.477***	2.180***	0.893***	-0.040	-0.085	**5.696*** **	3.966***	2.538***	0.964***	0.025	0.002
Var.2	**5.364*** **	3.906***	2.219***	0.734***	0.059	0.054	**6.152*** **	4.352***	2.447***	0.877***	0.010	0.013
Var.3	**4.788*** **	4.173***	2.931***	1.385***	-0.062	-0.065	**5.283*** **	4.663***	3.146***	1.421***	-0.076	-0.082
Var.4	**5.045*** **	4.664***	3.490***	1.337***	-0.029	-0.056	**5.648*** **	5.174***	3.854***	1.424***	-0.021	-0.091
Var.5	**4.843*** **	4.609***	4.215***	3.136***	-0.038	-0.127	**5.492*** **	5.243***	4.748***	3.583***	-0.038	-0.107
Var.6	**0.024**	0.024	0.021	0.004	-0.007*	-0.018*	**0.031**	0.022	0.024	0.008	-0.019***	-0.025**
Var.7	**0.042****	0.036	0.020	0.012	-0.008	-0.024**	**0.045***	0.037	0.028	0.013	-0.015**	-0.035**
Var.8	**0.020****	0.015	-0.001	0.002	-0.014	-0.031*	**0.025***	0.017	0.005	0.002	-0.013	-0.022*

Bold numbers indicate the highest MVALPL and ALPL across all the models. * indicates significance at 10% level; ** indicates significance at 5% level; and *** indicates significance at 1% level.

here (the density forecast are negative). Our understanding here is that the cumulative log predictive likelihood for a one-step ahead is an approximation of the marginal likelihood. Since the true model is known in the Monte Carlo simulation, one would expect that the benchmark yields the highest marginal likelihood.[10] In other words, the benchmark model performs best for a one-step forecast ahead.

Over all, a number of conclusions are drawn from the Monte Carlo simulation. The models with lower rank perform better in term of density forecast criteria. For shorter forecast horizons, the models with misspecified rank do not perform as well as the model with the correct rank based on the point forecast criteria (MSFE and MAFE). However, for longer forecast horizons they appear to perform better than the benchmark model. With respect to the density forecast, the lower rank models seem to outperform the benchmark for a h-step ahead forecast with $h \geq 2$.

4.6 Empirical Study

To evaluate the forecasting performance of the RRR model, we compare its forecast performance with the forecast performance of a number of models that are widely used in the forecasting literature for high dimensional data. These include the dynamic factor model (e.g., Forni *et al.* (2000), the Minnesota prior model (Banbura *et al.* (2010)), AR model with a random walk. The details of the factor model and the Minnesota prior are referred to Bernanke *et al.* (2005) and Banbura *et al.* (2010), respectively. A brief description of the approaches is presented in Section 4.6.2.

4.6.1 Data

The data used in this empirical study is from the FED_QD database of quarterly U.S. variables from 1960Q1 to 2014Q3. We work with a VAR system including 119 variables. The list of the variables used in the paper is presented in Section 4.C; a full description of the variables is referred to McCracken and Ng (2015). We transform the variables to stationary variables using the transformation codes provided in McCracken and Ng (2015).[11]

[10] In the computation, we set the priors of the identified parameters of the model at the true values.

[11] Similar to Pettenuzzo *et al.* (2016), we remove the series non-borrowed reserves, as it was too volatile during the Great Recession.

4.6.2 Forecasting Models

Minnesota Priors

The normal conjugate prior is chosen for the coefficient Φ and the inverse covariance variance matrix Σ

$$\begin{aligned} \text{vec}(\Phi)|\Sigma &\sim N(\text{vec}(\underline{\Phi}), \Sigma \otimes \underline{V}), \\ \Sigma^{-1} &\sim W(\underline{S}, \underline{v}), \end{aligned}$$

where $\underline{\phi}, \underline{V}, \underline{S}$, and $\underline{v}$ are prior hyperparameters.

Banbura *et al.* (2010) introduce two fictitious variables $\underline{Y}$ and $\underline{X}$ with dimension of $(ns+n+1)\times n$ and $(ns+n+1)\times(ns+1)$ in to the system (4.2.1), and show that the prior hyperparameters can be written as $\underline{\Phi} = (\underline{X}'\underline{X})^{-1}\underline{X}\underline{Y}$, $\underline{V} = (\underline{X}'\underline{X})^{-1}$, $\underline{S} = (\underline{Y} - \underline{X}\underline{\Phi})'(\underline{Y} - \underline{X}\underline{\Phi})$. For given specific forms of $\underline{Y}$ and $\underline{X}$, Banbura *et al.* (2010) show that their proposed priors coincide with the traditional Minnesota (Litterman (1986)) (except that the covariance matrix Σ is treated as unknown and a single scalar is used for shrinkage instead of the two scalars used for shrinkage in the traditional implementation). The forms of the two fictitious $\underline{Y}$ and $\underline{X}$ are

$$\underline{Y} = \begin{pmatrix} \frac{\text{diag}(\underline{\delta}_1\omega_1,\ldots,\underline{\delta}_n\omega_n)}{\lambda} \\ 0_{(ns-n+1)\times n} \\ \text{diag}(\omega_1,\ldots,\omega_n) \end{pmatrix}, \underline{X} = \begin{pmatrix} J_s \otimes \frac{\text{diag}(\omega_1,\ldots,\omega_n)}{\lambda} & 0_{ns\times 1} \\ 0_{1\times ns} & v \\ 0_{n\times ns} & 0_{n\times 1} \end{pmatrix}, \tag{4.6.1}$$

where $J_s = \text{diag}(1, 2, \ldots, s)$, $\Sigma = \text{diag}(\omega_1^2, \ldots, \omega_n^2)$. $\underline{\delta_i} = 1$ if the ith is believed to exhibit persistence; and $\underline{\delta_i} = 0$ if the ith is believed to exhibit little persistence. The first block of the fictitious variables reflects a prior belief on the autoregressive coefficients, the second block reflects the prior for the intercept and the third block is the prior for the covariance matrix. The parameter v in the second block of $\underline{X}$ is chosen at a small value, implying a fairly noninformative prior for the intercept. The shrinkage parameter λ is chosen as follows. We first estimate the VAR model using the training sample (which is half of the sample size), and use the resulting parameters to conduct an in-sample forecast within this training sample. The "optimal" parameter λ is chosen so that the fit of the in-sample forecast is as close to as that of the small VAR including only the forecast variables of interest (n_f). That is, we set a grid search for λ, e.g. $[0.01, 0.05, 0.1, 0.5, 1]$, and choose the "optimal" choice of λ by minimizing the

$$|\text{Fit}_{\lambda n} - \text{Fit}_{\infty n_f}|, \tag{4.6.2}$$

where $\text{Fit}_{\lambda n} = \frac{1}{n_f}\sum_{i=1}^{n_f}\frac{\text{MSFE}(i,\lambda,n)}{\text{MSFE}(i,0,n_f)}$, $\text{MSFE}(i,\lambda,n)$ is the MSFE of variable i using shrinkage parameter λ in a VAR with n variables. If we stack the actual data and the fictitious variables as $\overline{Y} = (Y', \underline{Y}')'$, $\overline{X} = (X', \underline{X}')'$, then the conditional posteriors of Φ and Σ^{-1} have the following forms

$$\begin{aligned}\text{vec}(\Phi)|\Sigma &\sim N(\text{vec}(\overline{\Phi}), \Sigma \otimes \overline{V}),\\ \Sigma^{-1} &\sim W(\overline{S}, \overline{v}),\end{aligned}$$

where $\overline{V} = (\overline{X}'\overline{X})^{-1}$, $\overline{\Phi} = (\overline{X}'\overline{X})^{-1}\overline{X}'\overline{Y}$, $\overline{S} = ((\overline{Y} - \overline{X\Phi})'(\overline{Y} - \overline{X\Phi}))^{-1}$, $\overline{v} = T + \underline{v}$, $\overline{Y} = (Y', \underline{Y}')$, $\overline{X} = (X', \underline{X}')'$.

Dynamic Factor Model

The idea underpinning a dynamic factor model is that a small number of factors can summarize the information in a large system of n variables (see Forni *et al.* (2000), Stock and Watson (2002)). The factor has the following form

$$\begin{aligned}Y_t &= \lambda_0 + \lambda_1 F_t + \epsilon_t^Y, \quad & \epsilon_t^Y &\sim N(0, \Omega^Y),\\ F_t &= \theta_1 F_{t-1} + \epsilon_t^F, & \epsilon_t^F &\sim N(0, \Omega^F),\end{aligned} \tag{4.6.3}$$

where F_t is a $q \times 1$ vector of unobserved latent factors, $q \ll n$. The vector of factors is assumed to follow VAR(1) process. The covariance-variance Ω^Y is a diagonal matrix, ϵ_t^F independent of $\epsilon_{\tilde{t}}^F$ at all t and $\tilde{t}$. To select the number of factors, we first specify the maximum possible number of factors to be $q^{max} = \sqrt{n}$. At each point in time, we use IC criteria proposed by Bai and Ng (2002) to determine the number of factors. We use Bayesian methods with non-informative priors to estimate and forecast the models.

4.6.3 Forecast Results

We conduct a recursive forecast and present the forecast results for nine variables of interest: real gross domestic product (GDPC96), consumer price index (CPIAUCSL), federal funds rate (FEDFUNDS), civilian unemployment rate (UNRATE), industrial production growth (INDPRO), money stock M2 (M2REALx), real personal consumption (PCECC, producer price index (PPIFGS), and personal consumption expenditure (PCECTPI). It is noted that one can choose to forecast other variables among the 119 variables, our choice of the nine variables is

purely based our interest, and the variables are also the most commonly forecast in the literature of macroeconomics.

In the recursive forecasting exercise, we use the sample from 1960Q1-1986Q4 to estimate the initial parameter estimates, which are then used to predict outcomes for 1987Q1 ($h = 1$) through 1990Q4 ($h = 12$). For the next forecast, 1987Q1 is included in the sample, and the same procedure as above is repeated to predict outcomes for 1988Q2-1990Q1. The process is recursively proceeded until 2015Q4, and a time series of forecasts is generated for up to three years $h = 1, \ldots, 12$.

To assess the forecast accuracy of models, we use both point and density forecast measures for a single variable (MSFE and ALPL), and for multivariate variables (WSFE and WALPL). These measures are described in Section 4.5.3. The benchmark model we choose here is AR(1) with a random walk. Tables 4.6.1 (4.6.2) summarizes WMSFE, MSFE (MVALPL and ALPL) over a number of forecast horizons $h = \{1, 4, 6, 8, 10, 12\}$ for the nine variables of interest. The value of WMSFE/MSFE less than 1 indicates that the other models (RRR, DFM, and Minnesota prior) perform better forecast than the benchmark, while the value of MVAPLP/APLP smaller than 0 suggests the benchmark is preferred.

Following Koop (2013), we choose the lag length for the VAR system of quarterly data is equal to 4. The cross-entropy is used to determine the rank of the VAR coefficients, and the result suggests the rank k equal to 3.

Our results suggest no single approach dominant in forecasting the variables of interest. Some models perform well in some cases and do not perform well in the others. In general, the point forecast and the density forecast measures achieve the same conclusions (with the exception of interest rates). Both WMSFE and WALPL in Tables 4.6.1-4.6.2 suggest that DFM provides the best forecast at shorter horizons $h = 1, 4, 6$ while RRR performs better forecast at longer forecast horizons $h = 8, 10, 12$. We look closer at the forecast performance for each individual variable of interest: DFM is found to be preferable for one-step ahead forecast for unemployment rate, and industrial production index, whereas RRR performs well for the two-step ahead forecast onwards. Minnesota prior provides a better forecast for money stock (M2), while DFM provides the best forecast for producer price index and personal consumption. As for the federal funds rate, it appears that none of the models outperforms the benchmark (AR(1) with a random walk) in terms of the point forecast measures. However, the results are mixed in term of density forecast measure. DFM performs well in forecasting the federal funds rate up to the forecast horizon $h = 6$.

Table 4.6.1: Out of sample point forecast performance (WMSFE/MSFE) of BRRR, BGR, and DFM relative to a random walk model

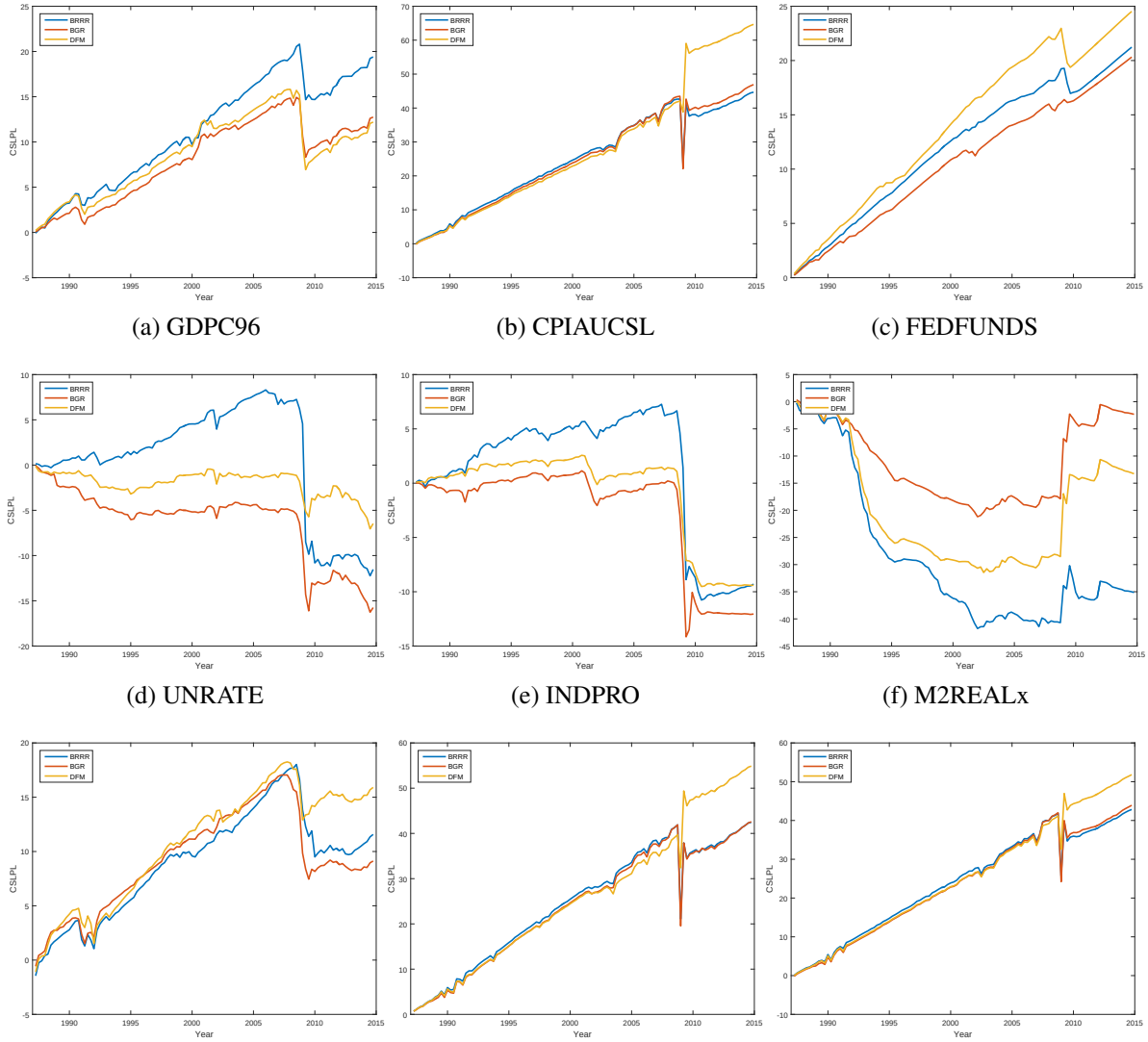
Variables/Models	**BRRR**	**BGR**	**DFM**	**BRRR**	**BGR**	**DFM**
WMSFE/MSFE		**h=1**			**h=4**	
Multivariate	0.532	0.522	**0.518**	0.733	0.840	**0.613**
GDPC96	**0.702**	0.802	0.881	1.115	1.691	**0.742**
CPIAUCSL	0.380	**0.370**	0.381	0.510	0.525	**0.416**
FEDFUNDS	1.206	**1.140**	1.291	2.837	3.429	**1.150**
UNRATE	1.070	1.275	**0.779**	**0.665**	0.725	1.087
INDPRO	1.317	1.413	**0.949**	**0.915**	1.349	1.191
M2REALx	1.341	**0.941**	1.120	0.747	0.651	**0.632**
PCECC96	0.846	0.882	**0.831**	1.381	1.860	**1.048**
PPIFGS	**0.387**	0.385	0.398	0.537	0.544	**0.404**
PCECTPI	0.395	**0.389**	0.404	0.600	0.615	**0.394**
		h=6			**h=8**	
Multivariate	0.844	1.040	**0.760**	0.708	0.914	**0.707**
GDPC96	0.785	1.551	**0.641**	**0.597**	1.417	0.609
CPIAUCSL	0.669	0.696	**0.558**	0.594	0.661	**0.535**
FEDFUNDS	2.768	2.844	**1.218**	2.362	2.630	**1.181**
UNRATE	**0.479**	0.682	0.965	**0.382**	0.501	0.807
INDPRO	**0.698**	1.323	1.162	**0.589**	1.325	1.309
M2REALx	0.946	0.771	**0.763**	0.937	**0.611**	0.680
PCECC96	1.043	1.985	**0.986**	**0.730**	1.694	0.888
PPIFGS	0.742	0.772	**0.577**	0.557	0.624	**0.469**
PCECTPI	0.928	0.956	**0.620**	0.725	0.799	**0.523**
		h=10			**h=12**	
Multivariate	**0.624**	0.823	0.682	**0.640**	0.847	0.758
GDPC96	**0.551**	1.447	0.664	**0.457**	1.188	0.588
CPIAUCSL	**0.478**	0.558	0.449	0.556	0.672	**0.546**
FEDFUNDS	1.877	2.507	**1.153**	1.431	2.373	1.176
UNRATE	**0.476**	0.463	0.864	0.617	**0.441**	0.956
INDPRO	**0.542**	1.375	1.603	**0.453**	1.132	1.578
M2REALx	1.177	**0.644**	0.779	1.092	**0.563**	0.718
PCECC96	**0.588**	1.550	0.911	**0.504**	1.409	1.009
PPIFGS	0.503	0.590	**0.441**	0.470	0.568	**0.429**
PCECTPI	0.542	0.633	**0.413**	0.687	0.849	**0.551**

The above forecast measures do not provide the relative forecast performance of the models to that of the benchmark at each time period. To shed a light on this issue, we consider the cumulative log of one-step ahead predictive likelihoods which are summarised in Figure 4.6.1. A pattern worth noting in Figure 4.6.1 is that the density forecast metric drops drastically in forecasting the variables during the financial crisis during the period 2007-2009. A sudden drop in CSLPL can be due to the fact that our model specification in this empirical study does not consider the volatility and structural break. Our model specification is in a static set-up. Incorporating the volatility in our model is beyond the scope of this chapter, but is worth to be

Table 4.6.2: Out of sample density forecast performance (MVALPL/ALPL) of BRRR, BGR, and DFM relative to a random walk model

Variables/Models	**BRRR**	**BGR**	**DFM**	**BRRR**	**BGR**	**DFM**
MVALPL/ALPL		**h=1**			**h=4**	
Multivariate	1.137	1.311	**1.402**	0.308	0.362	**0.625**
GDPC96	**0.175**	0.115	0.110	-0.105	-0.304	**0.142**
CPIAUCSL	0.402	0.423	**0.583**	0.115	0.081	**0.481**
FEDFUNDS	0.191	0.183	**0.221**	-0.153	-0.264	**0.106**
UNRATE	-0.104	-0.142	**-0.058**	0.060	**0.193**	-0.329
INDPRO	-0.084	-0.109	**-0.085**	**-0.159**	-0.304	-0.418
M2REALx	-0.317	**-0.021**	-0.120	0.161	**0.340**	0.252
PCECC96	0.104	0.082	**0.143**	-0.284	-0.414	**-0.103**
PPIFGS	0.382	0.384	**0.494**	0.056	0.035	**0.477**
PCECTPI	0.386	0.396	**0.467**	0.003	-0.024	**0.461**
		h=6			**h=8**	
Multivariate	0.382	0.344	**0.451**	**0.442**	-0.705	-0.841
GDPC96	0.054	-0.300	**0.190**	0.158	-0.267	**0.185**
CPIAUCSL	-0.079	-0.111	**0.304**	-0.022	-0.119	**0.309**
FEDFUNDS	-0.250	-0.250	**0.016**	-0.229	-0.264	**-0.052**
UNRATE	**0.306**	0.280	-0.312	0.544	**0.610**	-0.274
INDPRO	**0.050**	-0.312	-0.498	**0.132**	-0.320	-0.801
M2REALx	-0.071	**0.179**	0.011	-0.096	**0.379**	0.105
PCECC96	**-0.071**	-0.449	-0.091	**0.082**	-0.392	-0.126
PPIFGS	-0.164	-0.202	**0.239**	0.008	-0.107	**0.350**
PCECTPI	-0.257	-0.278	**0.210**	-0.136	-0.227	**0.287**
		h=10			**h=12**	
Multivariate	**0.723**	-0.356	-2.175	**0.744**	-0.139	-4.008
GDPC96	**0.187**	-0.276	0.131	**0.264**	-0.199	0.171
CPIAUCSL	0.121	-0.025	**0.415**	0.031	-0.152	**0.270**
FEDFUNDS	-0.131	-0.262	**-0.124**	**-0.017**	-0.250	-0.200
UNRATE	0.377	**0.722**	-0.451	0.084	**0.806**	-0.607
INDPRO	**0.176**	-0.341	-1.134	**0.314**	-0.194	-1.252
M2REALx	-0.333	0.347	**-0.035**	-0.256	**0.482**	0.041
PCECC96	**0.166**	-0.364	-0.237	**0.227**	-0.330	-0.382
PPIFGS	0.100	-0.066	**0.393**	0.157	-0.037	**0.411**
PCECTPI	0.062	-0.090	**0.429**	-0.059	-0.265	**0.246**

investigated. We leave it for a future research. For the variables (GDP and personal income consumption, real personal consumption, producer price index, personal consumption expenditure), although cumulative sum LPLs (CSLPL) drop after the financial crisis, the measures are still positive, suggesting all the models forecast better than the benchmark model (AR(1) with a random walk). For the industrial production index and unemployment rate, RRR and DFM provide better forecasts than the benchmark before the crisis, but the benchmark outperforms the others after 2007. As for money stock M2, the benchmark model is the best in forecasting the variable in a one-step forecast ahead.



(a) GDPC96 (b) CPIAUCSL (c) FEDFUNDS

(d) UNRATE (e) INDPRO (f) M2REALx

(g) PCECC (h) PPIFGS (i) PCECTPI

Figure 4.6.1: Cumulative sum of log predictive likelihood of one-step forecast ahead for the nine variables of interest: real gross domestic product (GDPC96), consumer price index (CPIAUCSL), federal funds rate (FEDFUNDS), civilian unemployment rate (UNRATE), industrial production growth (INDPRO), money stock M2 (M2REALx), real personal consumption (PCECC), producer price index (PPIFGS), and personal consumption expenditure (PCECTPI)

4.6.4 Impulse Responses

We now turn to the structural analysis of GDP and some macroeconomic variables to a monetary policy shock. To identify the monetary shock, we use a recursive identification scheme. We follow Bernanke *et al.* (2005) and Banbura *et al.* (2010), divide the variables of the VAR sys-

tem into two categories: slow-moving and fast-moving. The slow-moving group contains the real variables and prices while the fast groups includes financial variables. The slow-moving variables do not contemporaneously respond to a monetary policy shock while the fast-moving variables respond do. We order the variables as $Y_t = [Y_t^{sl}, r_t, Y_t^{fa}]'$, where Y_t^{sl} contains n_1 slow-moving variables, r_t is the monetary policy instrument (federal interest rate), and Y_t^{fa} include n_2 fast-moving variables. The monetary policy shock is orthogonal to all other shocks driving the economy.

To facilitate further discussion we rewrite VAR(s) in the form of VAR(1) such as

$$\mathbf{Y}_t = \mathbf{\Pi}_0 + \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \mathbf{e}_t, \tag{4.6.4}$$

$$\text{where} \mathbf{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-s-1} \end{bmatrix}, \mathbf{\Pi}_0 = \begin{bmatrix} \Phi_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{\Pi}_1 = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{s-1} & \Phi_s \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{bmatrix}, \mathbf{e}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Let $\mathbf{\Theta}$ be a $n \times n$ lower diagonal Cholesky matrix of the covariance matrix of the residuals $\mathbf{e}_t$ of the VAR system, i.e. $\mathbf{\Omega} = \mathbf{\Theta}\mathbf{\Theta}'$. The Wold representation of the reduced form VAR(s) is

$$\begin{aligned} \mathbf{Y}_t &= (\mathbf{I} - \mathbf{\Pi}(L))^{-1}\mathbf{\Pi}_0 + (\mathbf{I} - \mathbf{\Pi}(L))^{-1}\mathbf{\Theta}\mathbf{\Theta}^{-1}\mathbf{e}_t, \\ \mathbf{Y}_t &= (\mathbf{I} - \mathbf{\Pi}(L))^{-1}\mathbf{\Pi}_0 + (\mathbf{I} - \mathbf{\Pi}(L))^{-1}\mathbf{\Theta}\mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \mathbf{I}) \end{aligned}$$

where $(\mathbf{I} - \mathbf{\Pi}(L))^{-1} = \sum_{s=0}^{\infty} \mathbf{\Pi}_s L^s, \mathbf{\Pi}_0 = \mathbf{I}, \mathbf{\Pi}_s = \mathbf{\Pi}_1^s$. $\mathbf{u}_t$ is the linear transformation of the VAR residuals $\mathbf{u}_t = (u_{1t}, \dots, u_{nt})' = \mathbf{\Theta}^{-1}\mathbf{e}_t$. The monetary shock in the row of $\mathbf{u}_t$ corresponds to the variable r_t, which is $\mathbf{u}_{n_1+1,t}$. The coefficients $(\mathbf{I} - \mathbf{\Pi}(L))^{-1}\mathbf{\Theta}$ reflect the response functions of the variables to a structural shock.

At the outset of our discussion we note that our aim here is not to compare the results obtained from a large VAR model to a small/medium one. Here we wish to investigate whether our model provides sound results.[12]

Figure 4.6.2 displays the impulse response functions of a selection of macroeconomics of interest to a monetary shocks. The results provide some insight into how monetary policy may influence the real economy. We find that the "price puzzle" issue - an counterintuitive increase

[12]A number of studies (Banbura *et al.* (2010), Koop (2013), Pettenuzzo *et al.* (2016)) in the literature of large VARs investigate whether a large VAR model provide additional information in comparison to a small/ medium VARs.

of inflation by a contractionary monetary policy which often occurs in a small VAR model - did not appear here. Also, we find that responses of some macro economic variables have expected signs. A contractionary monetary policy is followed by a decline in GDP, which is in line with our expectation and the economic theory (Bernanke and Blinder (1988), Romer and Romer (1989)). The responses of other variables such as (civilian unemployment rate, industrial production growth, and civilian employment) have expected sign. That is, one would expect an increase in unemployment rate and a decrease in industrial production growth if there is a contractionary monetary policy innovation. A contractionary monetary policy tends to depress economic activity. For instance, an increase in interest rate often leads to an increase in the cost of capital, which lead to a fall in industrial production (see Barth III and Ramey (2000) for further discussion). Further an increase in the cost of capital affects capital accumulation, and capital accumulation influences the demand for labour, and the demand of labour affects unemployment rate (Galí (2010)).

4.7 Conclusions

In this chapter, we consider a reduced rank regression in large Bayesian VARs with an invariant specification using singular value decompositions. Different from Bayesian RRR specifications in a large VAR, our RRR specification does not depend on the ordering of the variables.

We carry out an extensive Monte Carlo simulation to examine the performance of cross entropy, predictive likelihood, and Laplace approximation used in a rank selection of VAR coefficients. We learn from the Monte Carlo simulation results that these approaches underestimate the number of rank when the singular values of the VAR matrices are small (close to zero) or the dimensions of VAR systems grow. We then go further to examine whether the forecast performance of the model is affected by a misspecified rank based on some metrics of point forecast (e.g. mean squared forecast errors, weighted mean squared forecast errors) and density forecast (e.g. log predictive likelihood). Our results suggest that the models with lower rank perform worse than the benchmark for short forecast horizons, however, they perform as well as or even better than the benchmark for long forecast horizons. These patterns are more evident when the magnitudes of the singular values of the VAR coefficient are small.

We apply the model to the data of the FED_QD database of quarterly U.S. variables from 1960Q1 to 2014Q3. Our results suggest no single model dominates in forecasting the variables. In comparison to other popular approaches used in large VAR models (shrinkage priors Banbura *et al.* (2010) and factor models Bernanke *et al.* (2005)), our proposed model provides

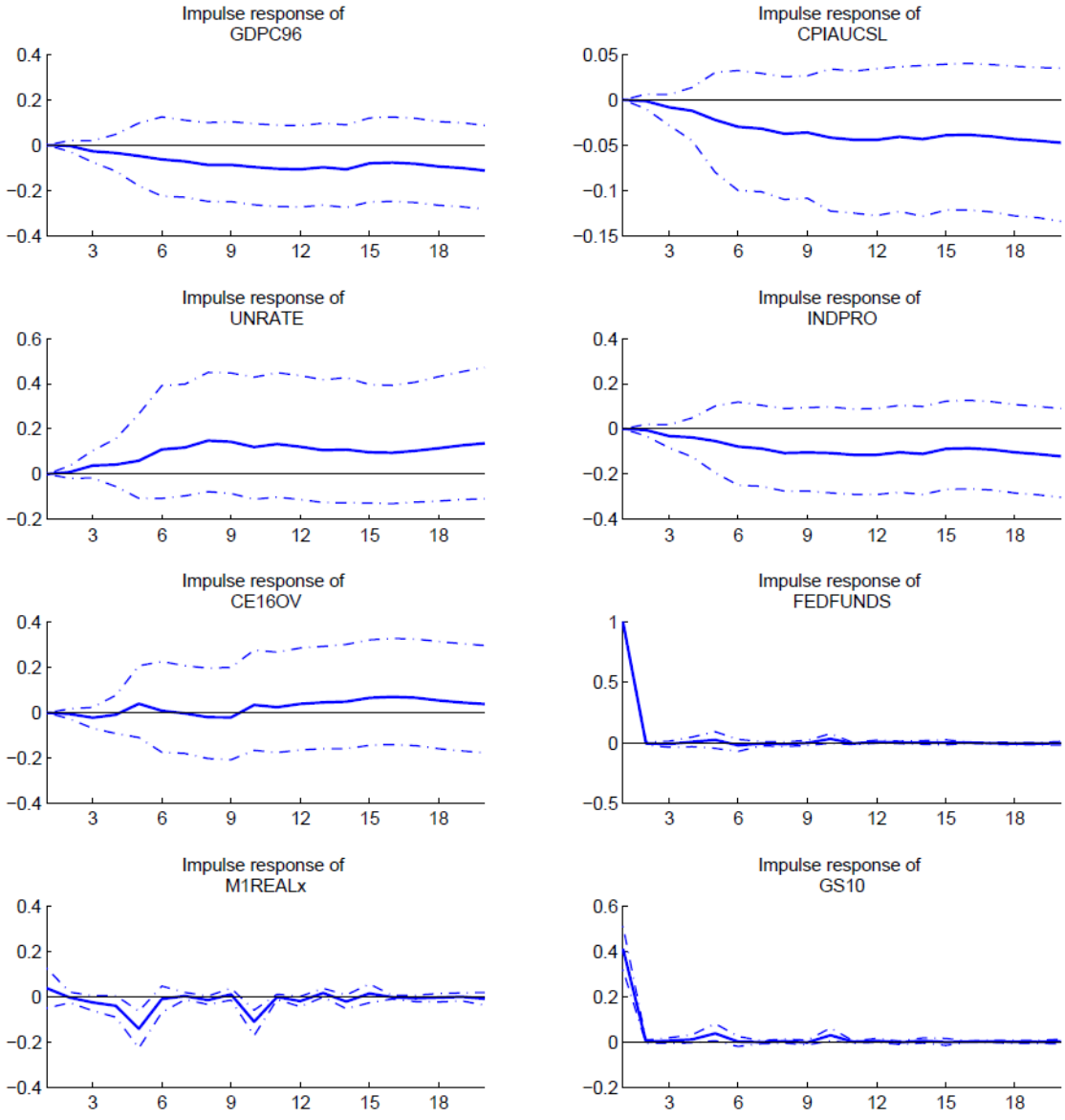


Figure 4.6.2: The impulse response functions to a monetary policy shock and the corresponding posterior intervals at 0.95% level for real gross domestic product (GDPC96), consumer price index (CPIAUCSL), civilian unemployment rate (UNRATE), industrial production growth (INDPRO), Civilian Employment (CE16OV), federal funds rate (FEDFUNDS), money supply M1 (M1REALx), 10-year treasury constant maturity rate (GS10).

better forecasts of gross domestic product, consumer price index, and producer price index in term of some point and density forecast measures. Also the impulse response functions of a selection of macroeconomic variables are sensible and provide some support to economic theories.

4.A Appendix A: Priors, Conditional Posterior Distributions and Joint Densities

4.A.1 Priors for (A, B)

Here we provide the proof for Theorem 4.3.1, which is an adaptation of Chan *et al.* (2017). For convenience, we recall the transformation of (U, Λ, V) to (A, B)

$$\begin{aligned} U\Lambda V &= (UC')(C\Lambda V), \quad \text{where } C \in \mathcal{O}(k) \text{ (i.e. } C'C = I_k) \\ &= B^*A^* = (B^*\kappa)(\kappa^{-1}A) = BA, \end{aligned}$$

where κ is a upper triangular positive definite.

The first parameter expansion, the orthogonal matrix $C \in \mathcal{O}(k)$ (i.e. $C'C = I_k$), is introduced. The roles of the parameter expansion are i) to map the parameters (U, Λ, V) to the expanded parameters (B^*, A^*), and ii) to obtain standard form for the priors of B^*, A^*. That is,

$$\begin{aligned} \Phi &= U\Lambda V' = UC'C\Lambda V' = B^*A^*, \\ B^* &= BC', \\ A^* &= C\Lambda V'. \end{aligned} \tag{4.A.1}$$

To attain the priors for B^* and A^*, the Jacobian transformations of the transformation from (U, Λ, V) to the expanded parameters (B^*, A^*) is required. The Jacobian transformations are taken from James (1954, p.71):

$$\begin{aligned} (B^{*'}dB^*) &= (U'dU), \\ (dA^*) &= (\prod_{i=1}^{k} \lambda_i)^{n-k} \prod_{i<j}^{k} (\lambda_i^2 - \lambda_j^2)(d\Lambda)(V'dV)(C'dC), \end{aligned} \tag{4.A.2}$$

Substituting (4.A.1) and (4.A.2) to (4.3.1)-(4.3.5), we obtain the equivalent priors

$$\frac{p(\Lambda, V|c_\lambda, \Sigma)(d\Lambda)(V'dV)(U'dU)(C'dC)}{c_\Lambda c_U} = \frac{p(A^*|c_\lambda, \Sigma)(dA^*)(B^{*'}dB^*)}{c_\Lambda c_U}, \tag{4.A.3}$$

$$p(A^*|c_\lambda, \Sigma) = \exp\{-\frac{c_\lambda}{2}\text{tr}(\Sigma^{-1}A^{*'}A^*)\}, \tag{4.A.4}$$

$$c_\Lambda = \int p(A^*|c_\lambda, \Sigma)(dA^*) = \left(\frac{2\pi}{c_\lambda}\right)^{nk/2}|\Sigma|^{k/2}. \tag{4.A.5}$$

The prior $p(A^*|c_\lambda, \Sigma)$ in (4.A.4) suggests the prior for A^* is a normal distribution with means of zero and the variance matrix $\frac{1}{c_\lambda}\Sigma$. The prior of B^* is the same as that of U, which is a uniform distribution over the Stiefel manifold $\mathcal{V}_{k,ns}$. For any matrix H in the Stiefel manifold, $\mathcal{V}_{k,ns}$, its differential form is invariant under the left and right transformations. For instance, under the left transformation, $H \to CH$, where $C \in \mathcal{O}_{ns}$, then $(H'dH) \to (H'dH)$. The prior, posterior distributions and measures are also preserved under the transformation. Unfortunately, the uniform prior of B^* does not lead to a standard form for its posterior.

To achieve a conjugate prior for B^*, the second parameter expansion κ is introduced. The parameter expansion is an upper triangular of $k \times k$ positive definite matrix. It maps (B^*, A^*) to (B, A):

$$B^*A^* = B^*\kappa\kappa^{-1}A^* = BA, \text{where } B = B^*\kappa, A = \kappa^{-1}A^*.$$

The Jacobian transformation of the bijective transformation from (A^*, B^*, κ) to (B, A) is based on Theorem 2.1.4, Theorem 2.1.13, and Theorem 2.1.14 in Muirhead (2005, Chapter 2). For convenience, we define $F = B'B = \kappa'\kappa$:

$$\begin{aligned}
B = B^*\kappa \Rightarrow (dB) &= 2^{-k}|F|^{\frac{np-k-1}{2}}(dF)(B^{*'}dB^*),\\
A = \kappa^{-1}A^* \Rightarrow (dA) &= |\kappa^{-1}|^n(dA^*) = |F|^{\frac{-n}{2}}(dA^*),\\
\Rightarrow (dA)(dB) &= 2^{-k}|F|^{\frac{np-n-k-1}{2}}(dF)(B^{*'}dB^*)(dA^*),\\
(dF)(B^{*'}dB^*)(dA^*) &= 2^k|F|^{-\frac{np-n-k-1}{2}}(dA)(dB),\\
&= 2^k|B'B|^{-\frac{np-n-k-1}{2}}(dA)(dB).
\end{aligned} \tag{4.A.6}$$

The priors in (4.A.3) are now equivalent to

$$\begin{aligned}\frac{p(A^*|c_\lambda,\Sigma)(dA^*)(B^{*'}dB^*)d(F)}{c_\Lambda c_U} &= \frac{p(A,B|c_\lambda,\Sigma)J(B)(dA)(dB)}{c_\Lambda c_U}, \\ J(B) &= 2^k|B'B|^{-(ns-n-k-1)/2}, \\ p(A,B|c_\lambda,\Sigma) &= \exp -\frac{c_\lambda}{2}\text{tr}(\Sigma^{-1}A'B'BA),\end{aligned} \tag{4.A.7}$$

where $J(.)$ denotes the Jacobian transformation. Chan *et al.* (2017) point out that the presence of the determinant $|B'B|$ in the priors of A and B complicates computation, i.e. the posterior distribution for B is not in the standard form. This can be overcome by exploiting the prior of F. In our case, if the number of lags $s = 1$, then a diffuse prior, $|F|^{-(ns-n-k-1)/2}$, is chosen so that conjugate priors for A and B are achieved. In the case $s > 1$, which is the most common case in empirical studies, a Wishart distribution for F with degrees of freedom of $ns - n$ results in the standard priors for A and B. For the sake of space, we present only the derivations of the priors for A and B in the most common case ($s > 1$). The priors in the case $s = 1$ are straightforward and are attained following the same procedure.

Incorporating the Wishart probability density prior for F into (4.A.7), we have

$$\begin{aligned}&\frac{|F|^{(ns-n-k-1)/2}\Big(\exp -\frac{1}{2}\text{tr}F\Big)p(A^*)(dA^*)(B^{*'}dB^*)d(F)}{c_\Lambda c_U c_F} \\ &= c|B'B|^{(ns-n-k-1)/2}\bigg(\exp -\frac{1}{2}\text{tr}B'B\bigg)\exp -\frac{c_\lambda}{2}\text{tr}\Big(\Sigma^{-1}A'B'BA\Big)(dA)(dB), \\ &= c\bigg(\exp -\frac{1}{2}\text{tr}(B'B)\bigg)\exp -\frac{c_\lambda}{2}\text{tr}\Big(\Sigma^{-1}A'B'BA\Big)(dA)(dB),\end{aligned} \tag{4.A.8}$$

where $c = \frac{2^k}{c_\Lambda c_U c_F}$ and $c_F = \int |F|^{(ns-n-k-1)/2}\bigg(\exp -\frac{1}{2}\text{tr}(F)\bigg) = 2^{(ns-n)k/2}\Gamma_k\bigg(\frac{ns-n}{2}\bigg)$.

The resulting joint prior distribution for A and B is given by

$$p(A,B|c_\lambda,\Sigma)(dA)(dB) = c\bigg(\exp -\frac{1}{2}\text{tr}(B'B)\bigg)\bigg(\exp -\frac{c_\lambda}{2}\text{tr}(\Sigma^{-1}A'B'BA)\bigg)(dA)(dB). \tag{4.A.9}$$

Given the priors in (4.A.9), it is clear that the prior of A conditional on B and Σ is a matrix variate normal distribution with means of zero and covariance matrix $\frac{1}{c_\lambda}\Sigma \otimes (B'B)^{-1}$, denoted as $N(\mathbf{0}, \frac{1}{c_\lambda}\Sigma \otimes (B'B)^{-1})$ (see Gupta and Nagar (2000, pp. 55-56) for the definition of the matrix variate normal distribution). The prior of B conditional on A and Σ is also a matrix variate normal distribution, $N(\mathbf{0}, (I_k + c_\lambda A\Sigma^{-1}A')^{-1} \otimes I_{ns})$. For the case $p = 1$, the prior of A is $N(\mathbf{0}, \frac{1}{c_\lambda}\Sigma \otimes (B'B)^{-1})$ while the prior for B is $N(\mathbf{0}, (c_\lambda A\Sigma^{-1}A')^{-1} \otimes I_{ns})$.

4.A.2 Joint Densities

Here we provide the derivation of the joint density function of observed Y and unknown parameter B in Theorem 4.4.1.

The joint density function of the likelihood function and the priors of A, B, and Σ^{-1} is

$$p(A, B, \Sigma^{-1}, Y) = c_3|\Sigma^{-1}|^{\frac{T+v_0+k-n-1}{2}} \exp\Big\{ -\frac{c_\lambda}{2}\text{tr}\Big[\Sigma^{-1}A'B'BA\Big]\Big\} \exp\Big\{ -\frac{1}{2}B'B\Big\} \times \exp\Big\{ -\frac{1}{2}\Sigma^{-1}S_0^{-1}\Big\} \exp\Big\{ -\frac{1}{2}\text{tr}\Big[\Sigma^{-1}(Y-XBA)'(Y-XBA)\Big]\Big\}, \tag{4.A.10}$$

where $c_3 = (c_\lambda)^{\frac{nk}{2}} 2^{-\frac{n(sk+v_0+T)}{2}} \pi^{-\frac{n(ks+k+T)}{2}} |S_0^{-1}|^{\frac{v_0}{2}} \frac{\Gamma_k(\frac{ns}{2})}{\Gamma_k(\frac{ns-n}{2})}$.

The joint density of B, Σ^{-1} and Y

Integrating A out of the joint density $p(A, B, |\Sigma^{-1}, Y)$, we have

$$\begin{aligned}
p(B, \Sigma^{-1}, Y) &= \int p(A, B, \Sigma^{-1}, Y)(dA), \\
&= c_3|\Sigma^{-1}|^{\frac{T+v_0+k-n-1}{2}} \exp\Big\{ -\frac{1}{2}B'B\Big\} \exp\Big\{ -\frac{1}{2}\Sigma^{-1}S_0^{-1}\Big\} \\
&\times \int \exp\Big\{ -\frac{c_\lambda}{2}\text{tr}[\Sigma^{-1}A'B'BA]\Big\} \exp\Big\{ -\frac{1}{2}\text{tr}[\Sigma^{-1}(Y-XBA)'(Y-XBA)]\Big\}(dA), \\
&= c_3|\Sigma^{-1}|^{\frac{T+v_0+k-n-1}{2}} \exp\Big\{ -\frac{1}{2}B'B\Big\} \exp\Big\{ -\frac{1}{2}\Sigma^{-1}S_0^{-1}\Big\} \exp\Big\{ -\frac{1}{2}\text{tr}[\Sigma^{-1}Y'Y]\Big\} \\
&\times \int \exp\Big\{ -\frac{1}{2}\text{tr}\Sigma^{-1}[A'(B'X'XB + c_\lambda B'B)A - 2A'B'X'Y]\Big\}(dA), \\
&= c_3(2\pi)^{\frac{nk}{2}}|\Sigma^{-1}|^{\frac{T+v_0-n-1}{2}}|V_A|^{\frac{n}{2}} \exp\Big\{ -\frac{1}{2}B'B\Big\} \exp\Big\{ -\frac{1}{2}\Sigma^{-1}S_0^{-1}\Big\} \\
&\times \exp\Big\{ -\frac{1}{2}\text{tr}[\Sigma^{-1}Y'Y]\Big\} \exp\Big\{ -\frac{1}{2}\text{tr}[\Sigma^{-1}\mu_A'V_A^{-1}\mu_A],\Big\}
\end{aligned}$$

where $V_A = [B'X'XB + c_\lambda B'B]^{-1}, \mu_A = V_A(B'X'Y)$.

The joint density of B and Y

Integrating Σ^{-1} out of $p(B, \Sigma^{-1}, Y)$, we have

$$\begin{aligned}
p(B,Y) &= \int p(B,\Sigma^{-1},Y)d\Sigma^{-1} \\
&= c_3(2\pi)^{\frac{nk}{2}}|V_A|^{\frac{n}{2}}\exp\Big\{-\frac{1}{2}B'B\Big\} \\
&\times \int |\Sigma^{-1}|^{\frac{T+v_0-n-1}{2}}\exp\Big\{-\frac{1}{2}\Sigma^{-1}S_0^{-1}\Big\}\exp\Big\{-\frac{1}{2}\text{tr}[\Sigma^{-1}Y'Y]\Big\} \\
&\times \exp\Big\{-\frac{1}{2}\text{tr}[\Sigma^{-1}\mu_A'V_A^{-1}\mu_A],\Big\}d(\Sigma^{-1}) \\
&= c_3(2\pi)^{\frac{nk}{2}}2^{\frac{(T+v0)n}{2}}\Gamma_n\Big(\frac{T+v_0}{2}\Big)|V_A|^{\frac{n}{2}}\exp\Big\{-\frac{1}{2}B'B\Big\}|S_0^{-1}+Y'Y-\mu_AV_A^{-1}\mu_A|^{-\frac{T+v_0}{2}} \\
&= c_3(2\pi)^{\frac{nk}{2}}2^{\frac{(T+v0)n}{2}}\Gamma_n\Big(\frac{T+v_0}{2}\Big)|V_A|^{\frac{n}{2}}\exp\Big\{-\frac{1}{2}B'B\Big\} \\
&\times |S_0^{-1}+Y'Y-Y'XB(B'X'XB+c_\lambda B'B)^{-1}B'X'Y|^{-\frac{T+v_0}{2}}|B'X'XB+c_\lambda B'B|^{\frac{-n}{2}} \qquad (4.A.11) \\
&= c_3(2\pi)^{\frac{nk}{2}}2^{\frac{(T+v0)n}{2}}\Gamma_n\Big(\frac{T+v_0}{2}\Big)\exp\Big\{-\frac{1}{2}B'B\Big\}|B'PB|^{\frac{T+v0-n}{2}}|B'QB|^{-\frac{T+v_0}{2}}, \qquad (4.A.12)
\end{aligned}$$

where $P = c_\lambda I_{np} + X'X$, $Q = P - X'Y(S_0^{-1}+Y'Y)^{-1}Y'X$. To obtain equation (4.A.12) from (4.A.11) we use some algebra results regarding the determinant of a matrix in Harville (1997, Chapter 13). That is

$$\begin{aligned}
& |S_0^{-1}+Y'Y-Y'XB(B'X'XB+c_\lambda B'B)^{-1}B'X'Y|^{-\frac{T+v_0}{2}} \\
= & \left|\begin{array}{cc} S_0^{-1}+Y'Y & Y'XB \\ B'X'Y & B'X'XB+c_\lambda B'B \end{array}\right|^{-\frac{T+v_0}{2}} \times |B'(X'X+c_\lambda I_{np})B|^{\frac{T+v_0}{2}}, \\
= & |B'(X'X+c_\lambda I_{np})B|^{\frac{T+v_0}{2}}|S_0^{-1}+Y'Y|^{-\frac{T+v_0}{2}}|B'(X'X+c_\lambda I_{np})B \\
& -B'X'Y(S_0^{-1}+Y'Y)^{-1}Y'XB|^{-\frac{T+v_0}{2}}, \\
= & |B'PB|^{\frac{T+v_0}{2}}|S_0^{-1}+Y'Y|^{-\frac{T+v_0}{2}}|B'PB-B'X'Y(S_0^{-1}+Y'Y)^{-1}Y'XB|^{-\frac{T+v_0}{2}}.
\end{aligned}$$

4.B Appendix B: Laplace Approximation

In this section, we show how Laplace approximation is used to approximate the integral

$$L = \int_{\mathcal{G}_{k,ns}} f(\tilde{B})g(\tilde{B})dg_k^{ns} = \frac{\int_{\mathcal{V}_{k,ns}} f(\tilde{B})g(\tilde{B})dv_k^{ns}}{\int_{\mathcal{O}_k} dv_k^k}, \qquad (4.B.1)$$

where $f(\tilde{B}) = |\tilde{B}'P\tilde{B}|^{1/2}|\tilde{B}'Q\tilde{B}|^{-1/2}$, $g(\tilde{B}) = |\tilde{B}'P\tilde{B}|^{1/2}$.

Before proceeding further, we define a number of matrices

1. $P = c_\lambda I_{ns} + X'X$, $Q = P - X'Y(S_0^{-1} + Y'Y)^{-1}Y'X = P - R$, where $R = X'Y(S_0^{-1} + Y'Y)^{-1}Y'X$,
2. $Z = P^{-1/2}RP^{-1/2}$ and $Z = C\Lambda C'$, where $C \in \mathcal{O}_{ns}$, $\Lambda = \text{diag}(\lambda_1(Z), \ldots, \lambda_{ns}(Z))$ is the eigenvalues of Z,
3. $\tilde{B}_0 = C'\tilde{B}$, $\tilde{B} = [\tilde{b}_1, \ldots, \tilde{b}_k]$, $\tilde{B}_0 = [\tilde{b}_{01}, \ldots, \tilde{b}_{0k}]$, then $dv_k^{ns} = \Lambda_{w=1}^{ns}\Lambda_{j=i+1}^{ns}\tilde{b}_j d\tilde{b}_s = \Lambda_{w=1}^{ns}\Lambda_{j=i+1}^{ns}\tilde{b}_{0j}d\tilde{b}_{0w}$.

From (3), the numerator of (4.B.1) is expressed as

$$L = \int_{\mathcal{V}_{k,ns}} f(\tilde{B})g(\tilde{B})dv_k^{ns} = \int_{\mathcal{V}_{k,ns}} f(C\tilde{B}_0)g(C\tilde{B}_0)dv_k^{ns} \tag{4.B.2}$$

$f(C\tilde{B}_0) = |\tilde{B}_0'P_0\tilde{B}_0|^{1/2}|\tilde{B}_0'Q_0\tilde{B}_0|^{-1/2}$, $g(C\tilde{B}_0) = |\tilde{B}_0'P_0\tilde{B}_0|^{1/2}$ $P_0 = C'PC, Q_0 = C'QC$.

To approximate the integral with respect to the $\tilde{B}_0$, the mode of $f(.)$ and the negative Hessian matrix of $-\log f(.)$ at the mode are required. The mode of the $f(.)$ is found by maximising $f(.)$, which is equivalent to minimising $f^{-1} = |\tilde{B}_0'P_0\tilde{B}_0|^{-1}|\tilde{B}_0'Q_0\tilde{B}_0|$. Following Theorem 3.29 in (Schott, 2005, pp. 121-122), we have:

$$\begin{aligned}
\min f^{-2} &= \min|\tilde{B}'P_0^{-1}Q_0\tilde{B}|, \\
&= \prod_{j=1}^{k}\lambda_{ns-k+1}(P_0^{-1}Q_0) = \prod_{j=1}^{k}\lambda_{ns-k+1}((C'PC)^{-1}(C'QC)), \\
&= \prod_{j=1}^{k}\lambda_{ns-k+1}(P^{-1}Q).
\end{aligned} \tag{4.B.3}$$

Using Theorem 3.2 Schott (2005, p.89), we obtain

$$\lambda_{ns-k+j}(P^{-1}Q) = \lambda_{ns-k+j}(P^{1/2}P^{-1}QP^{-1/2}) = \lambda_{ns-k+j}(P^{-1/2}QP^{-1/2}). \tag{4.B.4}$$

Further, it is noted that

$$P^{-1/2}QP^{-1/2} = P^{-1/2}(P-R)P^{-1/2} = I_{ns} - P^{-1/2}RP^{-1/2} = I_{ns} - Z. \tag{4.B.5}$$

Given (4.B.4) and (4.B.5), (4.B.3) can be written as

$$\begin{aligned}\prod_{j=1}^{k}\lambda_{ns-k+1}(P^{-1}Q) &= \prod_{j=1}^{k}\lambda_{ns-k+j}(P^{-1/2}QP^{-1/2}) \\ &= \prod_{j=1}^{k}(1-\lambda_j(Z)) \qquad (4.B.6)\\ &= \min|\tilde{B}_0'(I-Z)\tilde{B}_0| \\ &= \min|\tilde{B}'C'(I-Z)C\tilde{B}| \\ &= \min|I-\tilde{B}'C'ZC\tilde{B}| \\ &= \min|I-\tilde{B}'\Lambda\tilde{B}| \qquad (4.B.7)\end{aligned}$$

The minimum occurs when $\tilde{B} = [\pm I_k \ 0_{ns-k,k}]$, where $\pm I_k$ implies that the main diagonal elements yield either 1 or -1 and the elements in the off-diagonals are zeros. In other words, the function $f(.)$ yields the maximum values at each of the 2^k matrices $\tilde{B} = [\pm I_k \ 0_{ns-k}]$. To evaluate the integral at the mode, we split $\mathcal{V}_{k,ns}$ up into 2^k disjoint pieces, each containing exactly one of the matrices $[\pm I_k \ 0_{ns-k}]$ (see Muirhead (2005, Chapter 9) for further discussion). Hence, at the mode $\hat{B}^*$, for a large T we have

$$L \sim 2^k \int_{N\begin{bmatrix} I_k \\ 0_{ns-k,k}\end{bmatrix}} f^{T+v_0}(\tilde{B})g(\tilde{B})dv_k^{ns}, \qquad (4.B.8)$$

where $N\begin{bmatrix} I_k \\ 0_{ns-k,k}\end{bmatrix}$ denotes a neighbourhood of $\begin{bmatrix} I_k \\ 0_{ns-k,k}\end{bmatrix}$.

The negative Hessian matrix in our case is not simply the second derivatives of $\log(f)$, i.e., $\Psi = \partial^2 \log f/(\partial(vec)(\tilde{B}))'(\partial(vec)(\tilde{B}))$ as the $\tilde{B}$ is in the Stiefel manifold, and has $nsk-k(k+1)/2$ free parameters. One can follow Strachan and Inder (2004) to derive the negative Hessian matrix. Here we use some results in Villani (2005) and Magnus and Neudecker (2007, Chapter 9-10) to derive the Hessian matrix.

We can re-parametrise the matrix $\tilde{B}$ as follows

$$\tilde{B} = \begin{bmatrix}\tilde{B}_1 \\ \tilde{B}_2\end{bmatrix} = \begin{bmatrix} I_k \\ \overline{B}_2\end{bmatrix}\tilde{B}_1 = \overline{B}\tilde{B}_1. \qquad (4.B.9)$$

The function $f(.)g(.)$ is now equivalent to

$$\begin{aligned} f(\tilde{B})g(\tilde{B}) &= f(\overline{B})g(\overline{B})|\tilde{B}_1'\tilde{B}_1|^{-n/2}, \\ &= f(\overline{B})g(\overline{B})|\tilde{B}_1'\tilde{B}_1|^{-n/2}, \\ &= |\overline{B}'P\overline{B}|^{1/2}|\overline{B}'Q\overline{B}|^{-1/2}|\overline{B}'P\overline{B}|^{1/2}|\tilde{B}_1'\tilde{B}_1|^{-n/2}. \end{aligned} \tag{4.B.10}$$

Further, as $\tilde{B} \in \mathcal{V}_{k,np}$, $\tilde{B}'\tilde{B} = I_k$

$$\begin{aligned} I_k &= \tilde{B}_1'\tilde{B}_1 + \tilde{B}_2'\tilde{B}_2 \\ \tilde{B}_1^{-1'}\tilde{B}_1^{-1} &= I_k + \tilde{B}_1^{-1'}\tilde{B}_2'\tilde{B}_2\tilde{B}_1^{-1}, \\ \tilde{B}_1^{-1'}\tilde{B}_1^{-1} &= I_k + \overline{B}_2'\overline{B}_2 \qquad \text{This is from (4.B.9).} \end{aligned} \tag{4.B.11}$$

Substitute (4.B.11) to (4.B.10) and use $\overline{B} = \begin{bmatrix} I_k \\ \overline{B}_2 \end{bmatrix}$, we rewrite (4.B.10) as the function of $\overline{B}_2$. The underlying idea here is that we aim to evaluate the integral with respect to $\overline{B}_2$ which does not belong to Stiefel manifold and all elements are free in the matrix. This helps us to obtain the negative Hessian matrix easily by taking the second-order differentials of $-\log f$ with respect to $\overline{B}_2$. Hence, the expression of (4.B.10) in term of $\overline{B}_2$ is

$$\begin{aligned} |\tilde{B}P\tilde{B}| &= |P_{11} + \overline{B}_2'P_{21} + P_{12}\overline{B}_2' + \overline{B}_2'P_{22}\overline{B}_2'||I + \overline{B}_2'\overline{B}_2|, \\ |\tilde{B}Q\tilde{B}| &= |Q_{11} + \overline{B}_2'Q_{21} + Q_{12}\overline{B}_2' + \overline{B}_2'Q_{22}\overline{B}_2'||I + \overline{B}_2'\overline{B}_2|, \\ f^{T+v_0} &= |P_{11} + \overline{B}_2'P_{21} + P_{12}\overline{B}_2' + \overline{B}_2'P_{22}\overline{B}_2'|^{(T+v_0)/2}|Q_{11} + \overline{B}_2'Q_{21} + Q_{12}\overline{B}_2' + \overline{B}_2'Q_{22}\overline{B}_2'|^{-(T+v_0)/2} \\ g^n &= |P_{11} + \overline{B}_2'P_{21} + P_{12}\overline{B}_2' + \overline{B}_2'P_{22}\overline{B}_2'|^{-n/2}|I + \overline{B}_2'\overline{B}_2|^{-n/2} \end{aligned} \tag{4.B.12}$$

where

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

.

We use Lemma 3.4 from Villani (2005, p.331) and some results in James (1954) to obtain the Jacobian transformation $dv_k^{ns} \to d\overline{B}_2$. More specifically, Villani (2005) provides a Jacobian transformation from the Grassmann manifold to the real-line space, i.e., $\tilde{B}'d(\tilde{B}) = dg_k^{ns} = |I + \overline{B}_2'\overline{B}_2|^{ns/2}d\overline{B}_2$. Meanwhile, James (1954) provides the relationship among the measures of Grassmann manifold, Stiefel manifold, and orthogonal manifold, i.e., $dg_k^{ns} = dv_k^{ns}dg_k^k$.

Given the two results and the transformation in (4.B.12), equation (4.B.8) is now equivalent to

$$L \sim 2^k \int |P_{11} + \overline{B}_2' P_{21} + P_{12}\overline{B}_2' + \overline{B}_2' P_{22}\overline{B}_2'|^{\frac{T+v_0-n}{2}} |Q_{11} + \overline{B}_2' Q_{21} + Q_{12}\overline{B}_2' + \overline{B}_2' Q_{22}\overline{B}_2'|^{-\frac{T+v_0}{2}} \times |I + \overline{B}_2'\overline{B}_2|^{\frac{n-np}{2}} d\overline{B}_2 \tag{4.B.13}$$

To this end, our task is to evaluate the integral using Laplace approximation, the mode of f is at $\overline{B} = 0$, where $f = |P_{11}+\overline{B}_2' P_{21}+P_{12}\overline{B}_2'+\overline{B}_2' P_{22}\overline{B}_2'|^{\frac{1}{2}} |Q_{11}+\overline{B}_2' Q_{21}+Q_{12}\overline{B}_2'+\overline{B}_2' Q_{22}\overline{B}_2'|^{-\frac{1}{2}} |I+\overline{B}_2'\overline{B}_2|^{-\frac{1}{2}}$ the negative Hessian matrix at the model is equivalent to

$$\begin{aligned}
\Psi &= -(\Psi_1 - \Psi_2 - \Psi_3), \\
\Psi_1 &= -(P_{11}^{-1}P_{12} \otimes P_{21}P_{11}^{-1})K_{np-k,k} - (P_{11}^{-1} \otimes P_{21}P_{11}^{-1}P_{12}) + (P_{11}^{-1} \otimes P_{22}), \\
\Psi_2 &= -(Q_{11}^{-1}Q_{12} \otimes Q_{21}Q_{11}^{-1})K_{np-k,k} - (Q_{11}^{-1} \otimes Q_{21}P_{11}^{-1}Q_{12}) + (Q_{11}^{-1} \otimes Q_{22}), \\
\Psi_3 &= -I_k
\end{aligned} \tag{4.B.14}$$

where $K_{ns-k,k}$ is a permutation matrix, e.g. for any $(ns-k)\times k$ matrix of E, then $K_{ns-k,k}\text{vec}(E) = \text{vec}(E')$.

4.C Appendix C: Data

Table 4.C.1: Data Description

id	Series id	Tcode	Group	Fred Mnemonic	Brief Description
1	1	5	G1	GDPC96	Real Gross Domestic Product (Billions of Chained 2009 Dollars)
2	2	5	G1	PCECC96	Real Personal Cons. Exp., (Billions of Chained 2009 Dollars)
3	3	5	G1	PCDGx	Real Personal Cons. Exp.,: Durable goods (Billions of Chained 2009 Dollars)
4	4	5	G1	PCESVx	Real Personal Cons. Exp.,: Services (Billions of 2009 dollars)
5	5	5	G1	PCNDx	Real Personal Cons. Exp.,: Nondurable goods (Billions of chanied dollars
6	6	5	G1	GPDIC96	Real Gross Private Domestic Investment (Billions of Chained 2009)
7	7	5	G1	FPIx	Real Private Fixed Investment (Billions of chained 2009 dollars) using PCE
8	8	5	G1	Y033RC1Q027SBEAx	Real GPD Investment: Fixed Investment Nonresidential Equipment, deflated using PCE
9	9	5	G1	PNFIx	Real Private Fixed Investment: Nonresidential (Billions of Chained 2009), deflated using PCE
10	10	5	G1	PRFIx	Real Private Fixed Investment:Residential (Billions of Chained 2009), deflated using PCE
11	12	5	G1	GCEC96	Real Govt Cons Exp. & Gross Investment (Billions of Chained 2009 Dollars)
12	13	1	G1	A823RL1Q225SBEA	Gov Cons. Exp. & Gross Investment: Federal (Percentage change from preceding period)
13	15	5	G1	SLCEx	Real Government State and Local Consumption Expend., deflated using PCE (Billions of Chained 2009 Dollars), deflated using PCE
14	16	5	G1	EXPGSC96	Real Exports of Goods and Services (Billions of Chained 2009 Dollars)
15	17	5	G1	IMPGSC96	Real Imports of Goods and Services (Billions of Chained 2009 Dollars)
16	22	5	G2	INDPRO	Industrial Production Index: Total (Index 2012=100)
17	25	5	G2	IPMAT	Industrial Production Index: Materials (Index 2012 =100)
18	27	5	G2	IPNMAT	Industrial Production Index: Nondurable Goods Materials (Index 2012=100)
19	28	5	G2	IPDCONGD	Industrial Production Index: Durable Consumer Goods (Index 2012=100)
20	29	5	G2	IPB51110SQ	Industrial Production Index: Durable Goods Automative Products (Index 2012=100)
21	30	5	G2	IPNCONGD	Industrial Production Index: Nondurable Consumer Goods (Index 2012=100)
22	31	5	G2	IPBUSEQ	Industrial Production Index: Business Equipment (Index 2012=100)
23	34	1	G2	CUMFNS	Capacity Utilization: Manufacturing (SIC)(Percent of Capacity)
24	198	5	G2	IPMANSICS	Industrial Production Index: Manufacturing (Index 2012=100)
25	199	5	G2	IPB512222S	Industrial Production Index: Residential Utilities (Index 2012=100)
26	200	5	G2	IPFUELS	Industrial Production Index: Fuels (Index 2012=100)
27	201	1	G2	NAPMPI	ISM Manufacturing: Production Index
28	205	1	G2	NAPM	ISM Manufacturing: PMI Composite Index
29	37	5	G3	MANEMP	All Employees: Manufacturing (Thousands of Persons)
30	38	5	G3	SRVPRD	All Employees: Service-Providing Industries (Thousands of Persons)
31	39	5	G3	USGOOD	All Employees: Goods-Producing Industries (Thousands of Persons)
32	40	5	G3	DMANEMP	All Employees: Durable Goods (Thousands of Persons)
33	41	5	G3	NDMANEMP	All Employees: Nondurable Goods (Thousands of Persons)
34	42	5	G3	USCONS	All Employees: Construction (Thousands of Persons)
35	43	5	G3	USEHS	All Employees: Education and Health Services (Thousands of Persons)
36	44	5	G3	USFIRE	All Employees: Financial Activities (Thousands of Persons)
37	49	5	G3	USMINE	All Employees: Mining and Loggins (Thousands of Persons)
38	50	5	G3	USTPU	All Employees: Trade, Transportation and Utilities (Thousands of Persons)
39	51	5	G3	USGOVT	All Employees: Government (Thousands of Persons)
40	52	5	G3	USTRADE	All Employees: Retail trade (Thousands of Persons)
41	53	5	G3	USWTRADE	All Employees: Wholesale trade (Thousands of Persons)

Table 4.C.1 Continued:

42	57	5	G3	CE16OV	Civilian Employment (Thousands of Persons)
43	59	2	G3	UNRATE	Civilian Unemployment Rate (percent)
44	65	5	G3	UEMPLT5	Number of civilians unemployed-less than 5 wks (Thousands of Persons)
45	66	5	G3	UEMP5TO14	Number of civilians unemployed-for 5 to 14 weeks (Thousands of Persons)
46	67	5	G3	UEMP15TO26	Number of civilians unemployed-for 15 to 26 weeks (Thousands of Persons)
47	68	5	G3	UEMP27OV	Number of civilians unemployed-for 27 weeks and over (Thousands of Persons)
48	76	5	G3	HOANBS	Nonfarm Business Sector: Hours of All Persons (Index 2009=100)
49	77	2	G3	AWHNONAG	Avg Weekly hours of Production and Nonsupervisory Employees: Manufacturing
50	79	5	G3	AWOTMAN	Avg Weekly Overtime Hrs of Production and Nonsupervisory Employees: Manufacturing (Hours)
51	80	1	G3	HWIx	Index of help-wanted
52	229	5	G3	HWIURATIO	Ratio: Help-wanted ads to number unemployed
53	81	5	G4	HOUST	Housing starts: Total New Privately Owned Housing Units Started (Thousands of Unit)
54	84	5	G4	HOUSTMW	Housing Starts: Midwest (Thousands of Unit)
55	85	5	G4	HOUSTNE	Housing Starts: Northeast (Thousands of Unit)
56	86	5	G4	HOUSTS	Housing Starts: South (Thousands of Unit)
57	87	5	G4	HOUSTW	Housing Starts: West (Thousands of Unit)
58	206	1	G5	NAPMNOI	NAPM New Orders Index (%)
59	207	1	G5	PMNV	NAPM Inventories Index (%)
60	96	6	G6	PCECTPI	Personal Consumption Exp.: Chain-type Price Index (Index 2009=100)
61	97	6	G6	PCEPILE	PCE Price Index Excluding Food and Energy (Index 2009=100)
62	98	6	G6	GDPCTPI	Gross Domestic Product, Chain-price index (Index 2009=100)
63	102	6	G6	DDURRG3Q086SBEA	Personal Cons Expenditures: Durable Goods (chain-type price index)
64	103	6	G6	DSERRG3Q086SBEA	Personal Consumption Expenditures: Services, Price Index
65	104	6	G6	DNDGRG3Q086SBEA	Personal Consumption Expenditures: Nondur Goods, Price Index
66	105	6	G6	DHCERG3Q086SBEA	Personal Consumption Expenditures: Services: Household Consumption Expenditures
67	106	6	G6	DMOTRG3Q086SBEA	Personal Consumption Expenditures: Motor vehicles and parts
68	107	6	G6	DFDHRG3Q086SBEA	Personal Consumption Expenditures: Furniture and Household Equipment
69	108	6	G6	DREQRG3Q086SBEA	Personal Consumption Expenditures: Recreational Goods and Vehicles
70	109	6	G6	DODGRG3Q086SBEA	Personal Consumption Expenditures: Other Durables
71	110	6	G6	DFXARG3Q086SBEA	Personal Consumption Expenditures: Food and Beverages Purchased
72	111	6	G6	DCLORG3Q086SBEA	Personal Consumption Expenditures: Clothing and Footwear
73	112	6	G6	DGOERG3Q086SBEA	Personal Consumption Expenditures: Gas, Fuel Oil, and other Energy Goods, Price Index
74	113	6	G6	DONGRG3Q086SBEA	Personal Consumption Expenditures: Other Nondurables
75	114	6	G6	DHUTRG3Q86SBEA	Personal Consumption Expenditures: Housing and Utilities
76	115	6	G10	DHLCRG3Q086SBEA	Personal Consumption Expenditures: Medical Care
77	116	6	G9	DTRSRG3Q086SBEA	Personal Consumption Expenditures: Transportations Services
78	117	6	G11	DRCARG3Q086SBEA	Personal Consumption Expenditures: Recreation Services
79	120	6	G12	DOTSSRG3Q086SBEA	Personal Consumption Expenditures Other Services
80	121	6	G6	CPIAUCSL	Consumer Price Index for all Urban Consumers: All items (Index 1982-84=100
81	123	6	G6	PPIFGS	Producer Price Index by Commodity for Finished Goods (index 1982=100)
82	124	6	G6	PPIACO	Producer Price Index for All Commodities (Index 1982=100)
83	125	6	G6	PPIFCG	Producer Price Index by Commodity for Finished Consumer Goods (Index 1982=100)
84	126	6	G6	PPIFCF	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)

Table 4.C.1 Continued:

85	128	6	G6	PPIITM	Producer Price Index by Commodity Intermediate Materials: Supplies and Components (Index 1982=100)
86	131	5	G6	WPU0561	Producer Price Index by Commodity for Commodity for Fuels and Related Products and Power: (Index 1982=100)
87	214	6	G6	PPICRM	Producer Price Index: Crude Materials for Further Processing (index 1982=100)
88	222	6	G6	CPIULFSL	CPI for All Urban Consumers: All Items Less Food (index 1982-84=100)
89	134	5	G7	CES2000000008x	Real Avg Hourly Earnings of Production and Nonsupervisory Employees: Construction (2009 Dollars per Hour), deflated by core PCE
90	135	5	G7	CES3000000008x	Real Avg Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2009 dollars per hour), deflated by core PCE
91	137	5	G7	COMPRNFB	Nonfarm Business Sector: Real Compensation per hour (Index 2009=100)
92	138	5	G7	RCPHBS	Business sector: Real Compensation per hour (Index 2009=100)
93	141	5	G7	OPHPBS	Business Sector: Real Output Per Hour of All Persons (Index 2009=100)
94	142	5	G7	ULCBS	Business Sector: Unit Labor Cost (Index 2009=100)
95	144	5	G7	ULCNFB	Nonfarm Business Sector: Unit Labor Cost (Index 2009=100)
96	146	2	G8	FEDFUNDS	Effective Federal Funds Rate (Percent)
97	147	2	G8	TB3MS	3-Month Treasury Bill : Secondary Market Rate (Percent)
98	148	2	G8	TB6MS	6-Month Treasury Bill : Secondary Market Rate (percent)
99	150	2	G8	TB-1YR	1-Year Trasury Constant Maturity Rate (percent)
100	151	2	G8	GS10	10-year treasury constant maturity rate
101	153	2	G8	AAA	Moody's Seadoned Aaa Corporate Bond Yield (Percent)
102	154	2	G8	BAA	Moody's Seadoned Baa Corporate Bond Yield (Percent)
103	155	1	G8	BAA10YM	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10 -Year Treasury Constant Maturity (Percent)
104	157	1	G8	TB6M3Mx	6-Month Treasury Bill minus 3-Month Treasury Bill, secondary market (Percent)
105	158	1	G8	GS1TB3Mx	1-Year Treasury Constant maturity Minus 3-Month Treasury Bill, secondary market (percent)
106	159	1	G8	GS10TB3Mx	Spread btwn 10 year and 3 month T-bill rates
107	210	2	G8	GS5	5-Year Trasury Constant Maturity Rate (percent)
108	162	5	G9	AMBSKREALx	St Louis Adjusted Monetary Base, deflated by CPI
109	164	5	G9	M1REALx	Money stock: M1 (Billions of 1982-84 Dollars), deflated by CPI
110	165	5	G9	M2REALx	Money stock: M2 Billions of 1982-84 Dolars, deflated by CPI
111	167	5	G9	BUSLOANSx	Real Commerical and Industrial Loans, All Commercial Banks (Billions of U.S. Dollars), deflated by core PCE
112	172	5	G9	TOTALSLx	Consumer Credit Outstanding: deflated by core PCE
113	208	6	G9	TOTRESNS	Total Reserves of Depository Institution (Billions of Dollars)
114	209	7	G9	NONBORRES	Reserves of Depository Institutions (Billions of Dollars)
115	188	5	G11	EXRSW	Switzerland /U.S. Foerign Exchange Rate
116	189	5	G11	EXRJAN	Japan /U.S. Foreign Exchange Rate
117	190	5	G11	EXRUK	U.S. /U.K Foreign Exchange Rate
118	191	5	G11	EXRCAN	Canada / U.S. Foreign Exchange Rate
119	192	1	G12	UMCSENTx	Univ of Mich index of consumer expectations
120	254	5	G13	S&P500	S&Ps Common Stock Price Index: Composite
121	255	5	G13	S & P indust	S&Ps Common Stock Price Index: Industrials

4.D Appendix D: Figures

Table 4.D.1: The means and standard errors (in parenthesis) of selected ranks for DGPs with $k = 2$

DGP	Methods	(i) n=25	(ii) n=35	(iii) n=50
DGP1	**CE**	2.01(0.08)	2.00(0.00)	2.00(0.04)
	Pred.Like	1.98(0.15)	1.94(0.23)	0.00(0.00)
	Laplace	2.00(0.00)	2.00(0.00)	2.00(0.00)
DGP2	**CE**	2.00(0.00)	2.00(0.04)	2.00(0.06)
	Pred.Like	1.91(0.28)	1.79(0.41)	1.68(0.50)
	Laplace	2.00(0.00)	1.98(0.13)	1.92(0.27)
DGP3	**CE**	2.00(0.00)	2.01(0.08)	2.01(0.08)
	Pred.Like	1.90(0.45)	1.62(0.50)	1.56(0.59)
	Laplace	2.00(0.00)	1.45(0.50)	1.08(0.26)
DGP4	**CE**	2.00(0.04)	2.01(0.10)	2.01(0.10)
	Pred.Like	1.83(0.54)	1.64(0.53)	1.52(0.67)
	Laplace	1.99(0.10)	1.23(0.42)	1.02(0.15)
DGP5	**CE**	2.00(0.04)	2.00(0.06)	2.00(0.06)
	Pred.Like	1.87(0.64)	1.58(0.62)	1.51(0.74)
	Laplace	1.98(0.15)	1.07(0.26)	1.00(0.04)
DGP6	**CE**	2.01(0.11)	2.00(0.13)	2.00(0.11)
	Pred.Like	1.89(0.85)	1.58(0.68)	1.59(0.89)
	Laplace	1.87(0.34)	1.02(0.13)	1.00(0.00)
DGP7	**CE**	2.02(0.13)	1.99(0.16)	2.00(0.11)
	Pred.Like	1.88(0.80)	1.52(0.71)	1.74(1.12)
	Laplace	1.65(0.48)	1.00(0.00)	1.00(0.00)
DGP8	**CE**	2.01(0.15)	1.99(0.23)	1.99(0.16)
	Pred.Like	1.98(1.09)	1.51(0.84)	1.75(1.17)
	Laplace	1.35(0.48)	1.00(0.00)	1.00(0.00)
DGP9	**CE**	1.00(0.00)	1.01(0.08)	1.02(0.15)
	Pred.Like	1.76(1.29)	1.53(1.15)	1.79(1.32)
	Laplace	1.00(0.00)	1.00 (0.00)	1.00(0.04)

DGP1: $\lambda_1 = 0.97$, $\lambda_2 = 0.96$;
DGP2: $\lambda_1 = 0.97$, $\lambda_2 = 0.87$;
DGP3: $\lambda_1 = 0.97$, $\lambda_2 = 0.78$;
DGP4: $\lambda_1 = 0.97$, $\lambda_2 = 0.76$;
DGP5: $\lambda_1 = 0.97$, $\lambda_2 = 0.73$;
DGP6: $\lambda_1 = 0.97$, $\lambda_2 = 0.70$;
DGP7: $\lambda_1 = 0.97$, $\lambda_2 = 0.67$;
DGP8: $\lambda_1 = 0.97$, $\lambda_2 = 0.64$;
DGP9: $\lambda_1 = 0.97$, $\lambda_2 = 0.14$.

Table 4.D.2: The means and standard errors (in parenthesis) of selected ranks for DGPs with $k = 3$

DGP	Methods	(i) n=25	(ii) n=35	(iii) n=50
DGP1	**CE**	3.00(0.00)	3.00(0.00)	3.00(0.00)
	Pred.Like	2.92(0.29)	2.83(0.41)	2.57(0.62)
	Laplace	3.00(0.00)	3.00(0.00)	3.00(0.04)
DGP2	**CE**	3.00(0.06)	3.00(0.04)	3.00(0.00)
	Pred.Like	2.75(0.49)	2.60(0.58)	2.12(0.74)
	Laplace	3.00(0.00)	3.00(0.00)	2.86(0.37)
DGP3	**CE**	3.00(0.11)	2.97(0.18)	2.92(0.27)
	Pred.Like	2.50(0.62)	2.20(0.67)	1.83(0.66)
	Laplace	3.00(0.00)	2.71(0.45)	1.75(0.51)
DGP4	**CE**	2.81(0.39)	2.72(0.45)	2.45(0.50)
	Pred.Like	2.21(0.66)	1.94(0.66)	1.61(0.58)
	Laplace	2.48(0.50)	1.96(0.27)	1.20(0.40)
DGP5	**CE**	2.35(0.48)	2.32(0.46)	2.09(0.29)
	Pred.Like	2.03(0.66)	1.82(0.67)	1.54(0.72)
	Laplace	2.01(0.13)	1.55(0.50)	1.01(0.09)
DGP6	**CE**	2.11(0.32)	2.10(0.31)	2.04(0.21)
	Pred.Like	1.76(0.59)	1.83(0.95)	1.74(1.14)
	Laplace	1.92(0.27)	1.12(0.33)	1.00(0.00)
DGP7	**CE**	2.04(0.19)	2.04(0.23)	1.99(0.21)
	Pred.Like	1.75(0.77)	1.90 (1.17)	1.74(1.29)
	Laplace	1.59(0.49)	1.01(0.09)	1.00(0.00)
DGP8	**CE**	2.02(0.17)	1.98(0.33)	1.90(0.32)
	Pred.Like	1.78(0.94)	1.91(1.38)	1.86(1.51)
	Laplace	1.16(0.37)	1.00(0.00)	1.00(0.00)
DGP9	**CE**	1.99(0.20)	1.77(0.44)	1.74(0.46)
	Pred.Like	1.83(1.13)	2.05(1.50)	1.75(1.45)
	Laplace	1.01(0.09)	1.00(0.00)	1.00(0.00)

DGP1: $\lambda_1 = 0.97$, $\lambda_2 = 0.96$, $\lambda_3 = 0.95$;
DGP2: $\lambda_1 = 0.97$, $\lambda_2 = 0.92$, $\lambda_3 = 0.87$;
DGP3: $\lambda_1 = 0.97$, $\lambda_2 = 0.87$, $\lambda_3 = 0.77$;
DGP4: $\lambda_1 = 0.97$, $\lambda_2 = 0.82$, $\lambda_3 = 0.67$;
DGP5: $\lambda_1 = 0.97$, $\lambda_2 = 0.77$, $\lambda_3 = 0.57$;
DGP6: $\lambda_1 = 0.97$, $\lambda_2 = 0.72$, $\lambda_3 = 0.47$;
DGP7: $\lambda_1 = 0.97$, $\lambda_2 = 0.67$, $\lambda_3 = 0.37$;
DGP8: $\lambda_1 = 0.97$, $\lambda_2 = 0.62$, $\lambda_3 = 0.27$;
DGP9: $\lambda_1 = 0.97$, $\lambda_2 = 0.57$, $\lambda_3 = 0.17$.

Table 4.D.3: The means and standard errors (in parenthesis) of selected ranks for DGPs with $k = 4$

DGP	Methods	(i) n=25	(ii) n=35	(iii) n=50
DGP1	**CE**	4.00(0.00)	4.00(0.06)	4.00(0.10)
	Pred.Like	3.84(0.46)	3.79(0.52)	3.27(0.90)
	Laplace	4.00(0.00)	4.00(0.00)	3.99(0.09)
DGP2	**CE**	4.00(0.12)	3.92(0.26)	3.83(0.38)
	Pred.Like	3.44(0.81)	3.19(0.88)	2.51(0.99)
	Laplace	4.00(0.00)	3.97(0.17)	3.59(0.56)
DGP3	**CE**	3.88(0.32)	3.60(0.49)	3.42(0.49)
	Pred.Like	3.07(0.91)	1.99(1.14)	1.98(0.77)
	Laplace	3.99(0.12)	3.37(0.60)	2.37(0.53)
DGP4	**CE**	3.52(0.50)	3.26(0.46)	3.04(0.37)
	Pred.Like	2.58(0.94)	2.40(0.86)	1.71(0.60)
	Laplace	3.55(0.51)	2.60(0.50)	1.91(0.33)
DGP5	**CE**	3.26(0.44)	2.86(0.49)	2.73(0.47)
	Pred.Like	2.23(0.84)	2.06(0.71)	1.64(0.54)
	Laplace	2.90(0.37)	2.06(0.33)	1.63(0.48)
DGP6	**CE**	3.14(0.39)	2.65(0.51)	2.51(0.50)
	Pred.Like	2.21(0.82)	1.93(0.67)	1.60(0.59)
	Laplace	2.60(0.49)	1.95(0.25)	1.43(0.50)
DGP7	**CE**	2.85(0.44)	2.32(0.46)	2.20(0.40)
	Pred.Like	1.94(0.72)	1.87(0.64)	1.56(0.57)
	Laplace	2.06(0.39)	1.83(0.38)	1.16(0.37)
DGP8	**CE**	2.59(0.49)	2.13(0.33)	2.05(0.23)
	Pred.Like	1.80(0.63)	1.89(0.85)	1.57(0.77)
	Laplace	1.88(0.33)	1.56(0.50)	1.03(0.18)
DGP9	**CE**	2.13(0.33)	2.01(0.11)	1.98(0.15)
	Pred.Like	1.64(0.66)	1.84(1.07)	1.02(0.98)
	Laplace	1.40(0.49)	1.09(0.28)	1.00(0.00)

DGP1: $\lambda_1 = 0.97$, $\lambda_2 = 0.96$, $\lambda_3 = 0.95$, $\lambda_4 = 0.94$;
DGP2: $\lambda_1 = 0.97$, $\lambda_2 = 0.93$, $\lambda_3 = 0.89$, $\lambda_4 = 0.85$;
DGP3: $\lambda_1 = 0.97$, $\lambda_2 = 0.90$, $\lambda_3 = 0.83$, $\lambda_4 = 0.77$;
DGP4: $\lambda_1 = 0.97$, $\lambda_2 = 0.87$, $\lambda_3 = 0.78$, $\lambda_4 = 0.68$;
DGP5: $\lambda_1 = 0.97$, $\lambda_2 = 0.84$, $\lambda_3 = 0.72$, $\lambda_4 = 0.59$;
DGP6: $\lambda_1 = 0.97$, $\lambda_2 = 0.82$, $\lambda_3 = 0.66$, $\lambda_4 = 0.51$;
DGP7: $\lambda_1 = 0.97$, $\lambda_2 = 0.79$, $\lambda_3 = 0.60$, $\lambda_4 = 0.42$;
DGP8: $\lambda_1 = 0.97$, $\lambda_2 = 0.76$, $\lambda_3 = 0.55$, $\lambda_4 = 0.33$;
DGP9: $\lambda_1 = 0.97$, $\lambda_2 = 0.70$, $\lambda_3 = 0.43$, $\lambda_4 = 0.16$.

Table 4.D.4: The means and standard errors (in parenthesis) of selected ranks for DGPs with $k = 6$

DGP	Methods	(i) n=25	(ii) n=35	(iii) n=50
DGP1	**CE**	5.99(0.02)	5.98(0.08)	5.95(0.12)
	Pred.Like	5.84(0.42)	5.79(0.52)	5.27(0.95)
	Laplace	6.00(0.00)	6.00(0.00)	5.99(0.1)
DGP2	**CE**	5.98(0.12)	5.92(0.26)	5.83(0.38)
	Pred.Like	5.44(0.81)	5.19(0.88)	5.51(0.97)
	Laplace	6.00(0.00)	5.97(0.17)	5.59(0.56)
DGP3	**CE**	5.88(0.32)	5.60(0.49)	5.42(0.49)
	Pred.Like	5.07(0.91)	4.99(1.14)	4.98(0.77)
	Laplace	5.99(0.12)	4.7(0.60)	4.37(0.53)
DGP4	**CE**	5.52(0.50)	5.26(0.46)	5.04(0.37)
	Pred.Like	4.58(0.94)	4.40(0.86)	3.71(0.60)
	Laplace	4.55(0.51)	3.60(0.50)	3.91(0.33)
DGP5	**CE**	5.26(0.44)	4.86(0.49)	4.73(0.47)
	Pred.Like	4.23(0.84)	4.06(0.71)	3.64(0.54)
	Laplace	4.90(0.37)	4.06(0.33)	3.63(0.48)
DGP6	**CE**	5.14(0.39)	4.65(0.51)	4.51(0.50)
	Pred.Like	4.21(0.82)	3.93(0.67)	3.60(0.59)
	Laplace	4.60(0.49)	3.95(0.25)	3.43(0.50)
DGP7	**CE**	4.85(0.44)	4.32(0.46)	4.20(0.40)
	Pred.Like	3.94(0.72)	3.87(0.64)	3.56(0.57)
	Laplace	4.06(0.39)	3.83(0.38)	3.16(0.37)
DGP8	**CE**	4.59(0.49)	4.13(0.33)	4.05(0.23)
	Pred.Like	3.80(0.63)	3.89(0.85)	3.57(0.77)
	Laplace	3.88(0.33)	3.56(0.50)	3.03(0.18)
DGP9	**CE**	4.13(0.33)	4.01(0.11)	4.01(0.11)
	Pred.Like	3.64(0.66)	3.84(1.07)	3.00(0.98)
	Laplace	3.40(0.49)	3.09(0.28)	3.00(0.95)

DGP1: $\lambda_1 = 0.97$, $\lambda_2 = 0.95$, $\lambda_3 = 0.93$, $\lambda_4 = 0.91$, $\lambda_5 = 0.90$, $\lambda_6 = 0.88$;
DGP2: $\lambda_1 = 0.97$, $\lambda_2 = 0.93$, $\lambda_3 = 0.90$, $\lambda_4 = 0.86$, $\lambda_5 = 0.83$, $\lambda_6 = 0.79$;
DGP3: $\lambda_1 = 0.97$, $\lambda_2 = 0.92$, $\lambda_3 = 0.86$, $\lambda_4 = 0.81$, $\lambda_5 = 0.76$, $\lambda_6 = 0.70$;
DGP4: $\lambda_1 = 0.97$, $\lambda_2 = 0.90$, $\lambda_3 = 0.83$, $\lambda_4 = 0.76$, $\lambda_5 = 0.69$, $\lambda_6 = 0.61$;
DGP5: $\lambda_1 = 0.97$, $\lambda_2 = 0.88$, $\lambda_3 = 0.79$, $\lambda_4 = 0.70$, $\lambda_5 = 0.61$, $\lambda_6 = 0.53$;
DGP6: $\lambda_1 = 0.97$, $\lambda_2 = 0.86$, $\lambda_3 = 0.75$, $\lambda_4 = 0.65$, $\lambda_5 = 0.54$, $\lambda_6 = 0.44$;
DGP7: $\lambda_1 = 0.97$, $\lambda_2 = 0.85$, $\lambda_3 = 0.72$, $\lambda_4 = 0.60$, $\lambda_5 = 0.47$, $\lambda_6 = 0.35$;
DGP8: $\lambda_1 = 0.97$, $\lambda_2 = 0.83$, $\lambda_3 = 0.69$, $\lambda_4 = 0.54$, $\lambda_5 = 0.40$, $\lambda_6 = 0.26$;
DGP9: $\lambda_1 = 0.97$, $\lambda_2 = 0.81$, $\lambda_3 = 0.65$, $\lambda_4 = 0.49$, $\lambda_5 = 0.33$, $\lambda_6 = 0.17$.

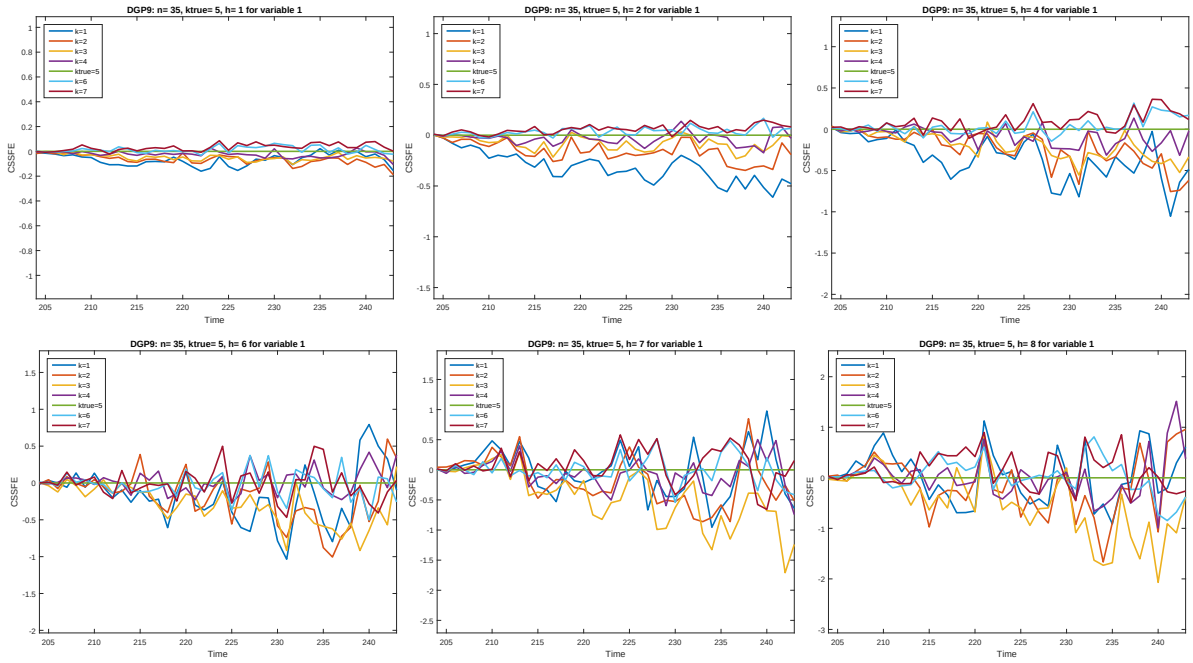


Figure 4.D.1: Cumulative squared forecast errors differentials for variable 1, $n = 35$, $k = 5$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$)

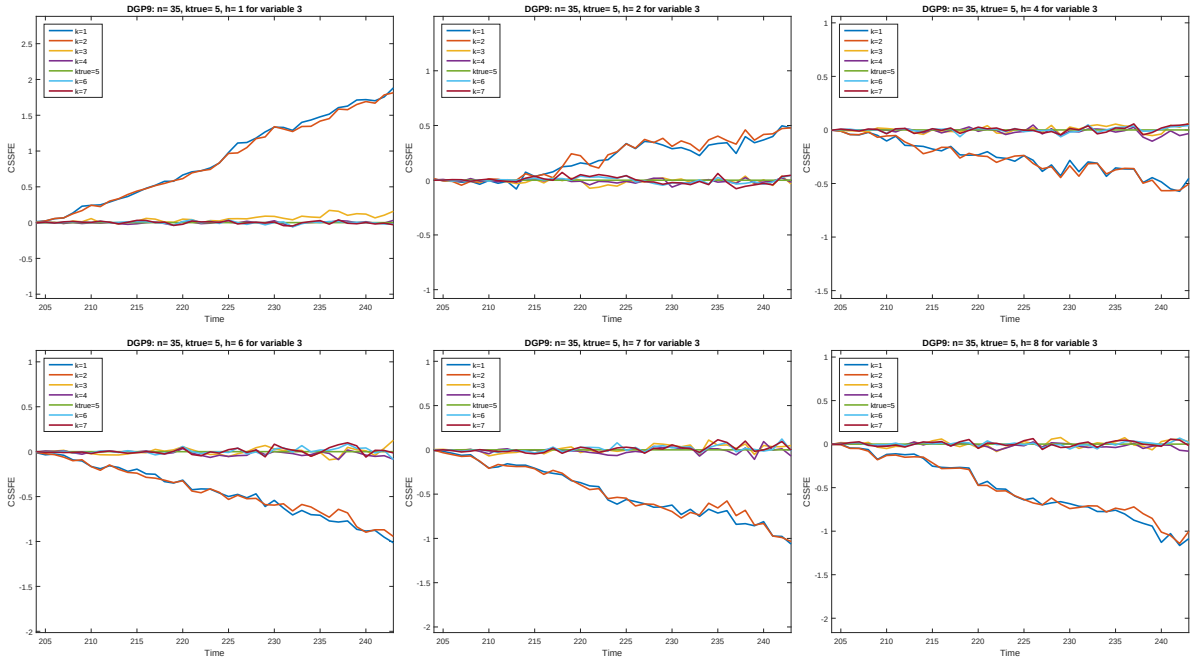


Figure 4.D.2: Cumulative squared forecast errors differentials for variable 3, $n = 35$, $k = 5$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$)

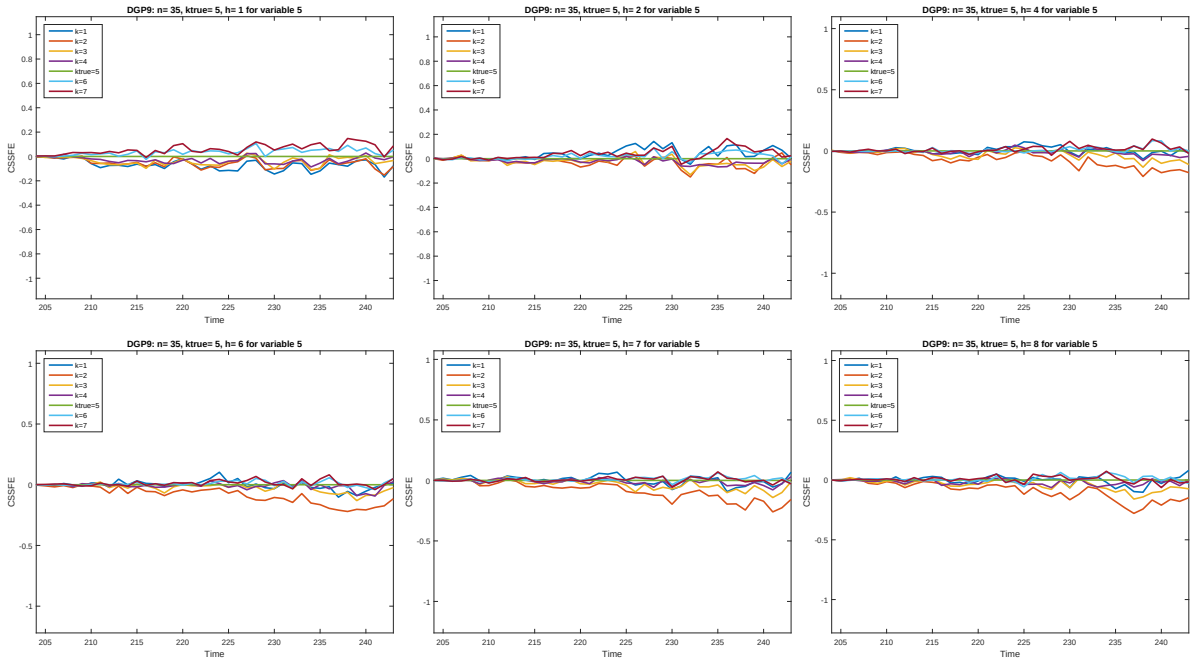


Figure 4.D.3: Cumulative squared forecast errors differentials for variable 5, $n = 35$, $k = 5$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$)

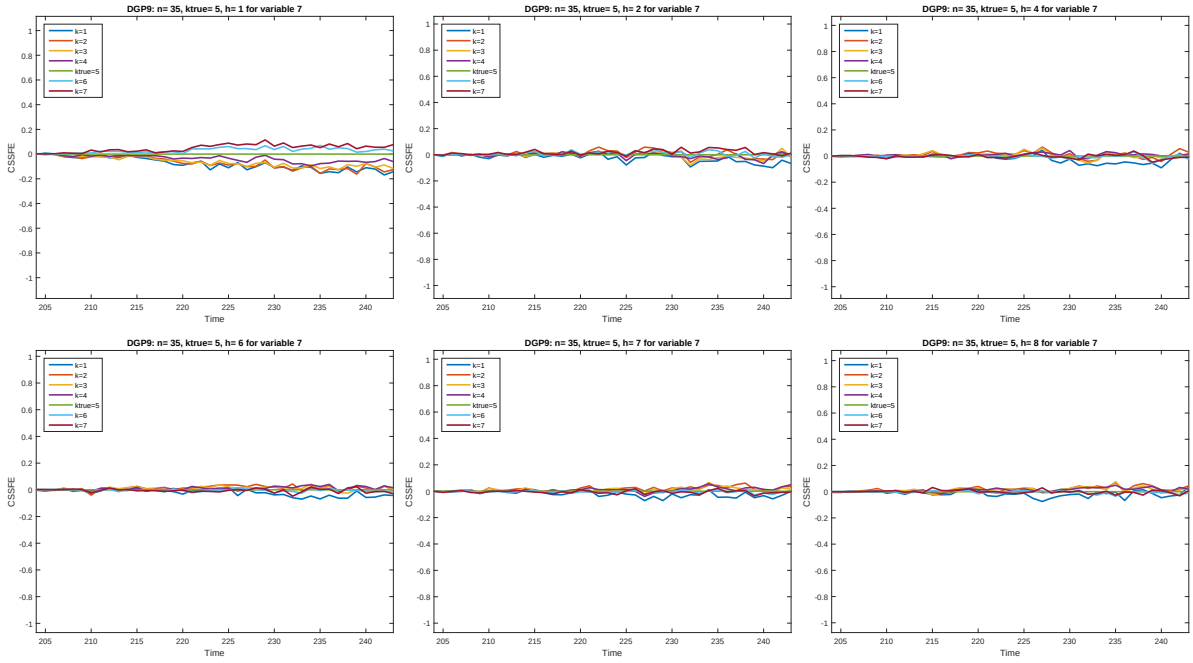


Figure 4.D.4: Cumulative squared forecast errors differentials for variable 7, $n = 35$, $k = 5$, DGP9 ($\lambda_1 = 0.97, \lambda_2 = 0.77, \lambda_3 = 0.57, \lambda_4 = 0.37, \lambda_5 = 0.17$)

CHAPTER 5

Future Research

The end of one journey is the beginning of another one...

In this thesis, a number of additional insights about the behaviours and the drivers of economic output (e.g. GDP, labour productivity growth) are gained by extending a range of econometric models available.

In **Chapter 2** we generalise the existing models to account for noise in the estimation of the production function and to model technological change in the context of two-period panel data. The new methods confirm some of the findings and also provide additional insights about the factors contributing to labour productivity of the countries. In particular, we confirm that capital deepening made the largest contribution to labour productivity growth from 1965 to 1990 and was also the main factor driving the transformation of the distribution of labour productivity from a unimodal to a bimodal distribution over this period. However, the new methods suggest that the magnitude of the contribution of capital deepening is underestimated when statistical noise is not accounted for (in standard DEA applied in the KR study). These methods do not confirm the hypothesis that capital deepening is a significant factor for the world convergence, which was found in the KR study. This implies that capital deepening might not have contributed to reducing the gap between the poor and the rich over this period to the extent previously suggested.

A number of research directions can be extended from this chapter. One can develop statistical inference (e.g. we can compute standard errors, confidence intervals) for efficiency scores or the decomposition indexes in the framework of LLE and SDEA. This provides an insight on

whether the contribution of the decomposition indexes to labour productivity are significant. One possible way of developing such statistical inferences is to develop a proper bootstrap algorithm. Another direction is to consider other factors such as human capital, foreign direct investment and investigate their impacts on productivity growth.

In **Chapter 3** we aim to capture the dynamic pattern of technical efficiencies and the effect of exogenous factors on a production process over time. For this, we extend the static model specification of Wang and Schmidt (2002), Alvarez *et al.* (2006) to a dynamic one. Compared to the existing models, the proposed model allows to i) capture the dynamic pattern of technical inefficiencies and the effects of exogenous variables on inefficiencies and ii) capture the heterogeneous effect of exogenous variables on inefficiencies across countries. We also provide a hypothesis test to examine whether the time-varying specification is needed.

In our application, FDI plays a more important role in influencing the shift of the production function than the movement towards the frontier (reductions in the technical inefficiencies). This suggests that FDI should be invested in innovation to improve the technological change for the countries. We also find that FDI makes a contribution to the growth of some of the developed economies such as Ireland, Norway, Finland and United Kingdom.

One possible extension of the work in this chapter is to consider developing countries and other exogenous variables such as human capital to investigate whether the same conclusion regarding the impacts of FDI on the production process for the developing countries hold.

In **Chapter 4** we focus on investigating the dynamic responses (from the impulse response functions) of output growth and other important macroeconomic variables (such as unemployment rate, inflation) to structural shocks (e.g. monetary policy), and the forecasts of the variables in a large VAR model with 119 variables. Our proposed model specification helps to improve the precision of estimating such a large VAR model by using a reduced rank regression which has an invariant specification to the ordering of the variables. Such a specification significantly mitigate the computation burden. That is, using the non-invariant model in our application one would need to compute over 118 million orderings for a rank of 5, and for a rank of 10 the number of orderings exceed over 100 trillion. Whereas using the specification in this thesis, only one model need be estimated for each rank.

In the empirical study, we find that the "price puzzle" issue - an counterintuitive increase of inflation with a contractionary monetary policy which often occurs in a small VAR model - did not appear here. Also, we find that a contractionary monetary policy was followed by a decrease in GDP, price level and an increase in unemployment rate, supporting the conventional channel of the effects of a tight monetary policy on a real economy. In terms of forecasting, we find that the forecast of GDP growth, consumer price index, and producer price index are improved in comparison to a random walk AR model in terms of a point forecast measure (e.g. MSFE) and a density forecast measure (e.g. ALPL).

In our extensive Monte Carlo simulation we investigate some existing approaches in rank selection for a large VAR model. We find that cross entropy, predictive likelihood and Laplace approximation tend to underestimate the rank of VAR matrices, and that underestimated rank models appear to provide better forecasts than the benchmark for long forecast horizons in terms of point forecast and density forecast measures.

One can extend the study to the context of a panel VAR to capture the interdependencies across the economies or potential spillover effects. Some issues relating model specifications are worthwhile to be investigated. Should the rank vary across countries? Should the rank reduction be applied to full lag matrix for all countries?

Thank you for reading my thesis!

Bibliography

Ahn, S. C. and Sickles, R. C. (2000). Estimation of long-run inefficiency levels: a dynamic frontier approach. *Econometric Reviews*, 19(4):461–492.

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Aitchison, J. and Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3):413–420.

Alvarez, A., Amsler, C., Orea, L., and Schmidt, P. (2006). Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis*, 25(3):201–212.

Amini, S. M. and Parmeter, C. F. (2012). Comparison of model averaging techniques: Assessing growth determinants. *Journal of Applied Econometrics*, 27(5):870–876.

Badunenko, O., Henderson, D. J., and Russell, R. R. (2013). Polarization of the worldwide distribution of productivity. *Journal of Productivity Analysis*, 40(2):153–171.

Badunenko, O., Henderson, D. J., and Zelenyuk, V. (2008). Technological change and transition: Relative contributions to worldwide growth during the 1990s. *Oxford Bulletin of Economics and Statistics*, 70(4):461–492.

Badunenko, O., Henderson, D. J., and Zelenyuk, V. (2017). The Productivity of Nations. In Emili Grifell-Tatjé, C. K. L. and C.Sickles, R., editors, *The Oxford Handbook of Productivity Analysis*, chapter 24.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Banbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Barro, R. J. (1999). Notes on growth accounting. *Journal of Economic Growth*, 4(2):119–137.

Barro, R. J. and Sala-i Martin, X. (1992). Convergence. *Journal of Political Economy*, 100(2):223–251.

Barth III, M. J. and Ramey, V. A. (2000). The cost channel of monetary transmission. Working Paper 7675, National Bureau of Economic Research.

Battese, G. and Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, 3(1-2):153–169.

Battese, G. and Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2):325–332.

Bauwens, Luc, M. L. and Richard, J.-F. (2000). *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Bernanke, B. S. and Blinder, A. S. (1988). Credit, money, and aggregate demand. *The American Economic Review*, 78(2):435–439.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525 – 546.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 79(4):655–673.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82(1):81–91.

Bruno, M. and Sachs, J. (1985). *The Economics of Worlwide Stagflation*. Oxford:Basil Blackwell.

Camba-Mendez, G., Kapetanios, G., Smith, R. J., and Weale, M. R. (2003). Tests of rank in reduced rank regression models. *Journal of Business and Economic Statistics*, 21(1):145–155.

Carriero, A., Clark, T. E., and Marcellino, M. (2016a). Common drifting volatility in large Bayesian VARs. *Journal of Business and Economic Statistics*, 34(3):375–390.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761.

Carriero, A., Kapetanios, G., and Marcellino, M. (2016b). Structural analysis with multivariate autoregressive index models. *Journal of Econometrics*, 192(2):332 – 348.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):pp. 541–553.

Caselli, F. (2005). Accounting for cross-country income differences. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 9, pages 679–741. Elsevier.

Caudill, S. B. and Ford, J. M. (1993). Biases in frontier estimation due to heteroscedasticity. *Economics Letters*, 41(1):17 – 20.

Caudill, S. B., Ford, J. M., and Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business and Economic Statistics*, 13(1):105–111.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6):1393–1414.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1–25.

Chan, J. and Strachan, R. (2014). The zero lower bound: Implications for modelling the interest rate. MPRA Paper 39360, Research School of Economics, Australian National University.

Chan, J. C., Leon-Gonzalez, R., and Strachan, R. W. (2017). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association, forthcoming*.

Chan, J. C. C. (2016). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, 0(0):1–17.

Chan, J. C. C. and Eisenstat, E. (2015). Marginal likelihood estimation with the cross-entropy method. *Econometric Reviews*, 34(3):256–285.

Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.

Chickering, D. M. and Heckerman, D. (1996). Efficient approximations for the marginal likelihood of incomplete data given a Bayesian network. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, pages 158–168, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Chikuse, Y. (2003). *Statistics on special manifolds*. Springer.

Christine, D. M., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318 – 328.

Christoffersen, P. F. and Diebold, F. X. (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics*, 16(4):450–458.

Cochrane, J. H. (1998). What do the VARs mean? Measuring the output effects of monetary policy . *Journal of Monetary Economics*, 41(2):277 – 300.

Cornwell, C., Schmidt, P., and Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46(1-2):185–200.

Cragg, J. G. and Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, 76(12):223 – 250.

Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24(1):93–121.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.

Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: fifteen years later. In *Oxford Handbook of Nonlinear Filtering*, pages 656–704.

Doucet, Arnaud, D. F. A., Gordon, N., and Johansen, A. M. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Du, P., Parmeter, C., and Racine, J. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, 23(3):1347–1371.

Duin, R. P. W. (1976). On the choice of smoothing parameters for parzen estimators of probability density functions. *IEEE Transactions on Computers*, C-25(11):1175–1179.

Durbin, J. and Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3):669–684.

Emvalomatis, G., Stefanou, S. E., and Lansink, A. O. (2011). A reduced-form model for dynamic efficiency measurement: Application to dairy farms in Germany and the Netherlands. *American Journal of Agricultural Economics*, 93(1):161–174.

Engle, R. F. and Kozicki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics*, 11(4):369–380.

Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the Euro area? *Journal of Monetary Economics*, 50(6):1243 – 1255.

Frühwirth-Schnatter, S. and Wagner, H. (2008). Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling. *Computational Statistics and Data Analysis*, 52(10):4608 – 4624.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85 – 100.

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3):397–416.

Galí, J. (2010). Monetary policy and unemployment. Working Paper 15871, National Bureau of Economic Research.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553 – 580.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121 – 146.

Geweke, J. (2001). Bayesian econometrics and forecasting. *Journal of Econometrics*, 100(1):11–15.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns . *International Journal of Forecasting*, 26(2):216 – 230.

Ghosh, J. and Dunson, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.

Gill, L. and Lewbel, A. (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models. *Journal of the American Statistical Association*, 87(419):766–776.

Glass, A. J. and Saggi, K. (1998). International technology transfer and the technology gap. *Journal of Development Economics*, 55(2):369 – 398.

Grubb, D., Jackman, R., and Layard, R. (1983). Wage rigidity and unemployment in OECD countries. *European Economic Review*, 21(1):11 – 39.

Gu, W. and Russell, R. R. (2010). Foreign direct investment and convergence: A nonparametric production frontier approach. Unpublished manuscript, University of California, Riverside.

Guhaniyogi, R. and Dunson, D. B. (2015). Bayesian compressed regression. *Journal of the American Statistical Association*, 110(512):1500–1514.

Gupta, A. and Nagar, D. (2000). *Matrix variate distribution*. Chapman & Hall.

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.*, 11(4):1156–1174.

Hall, P. and Marron, J. S. (1991). Local minima in cross-validation functions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(1):245–252.

Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, 75(4):1175–1189.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281 – 291.

Harville, D. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer.

Hayashi, F. and Prescott, E. C. (2002). The 1990s in Japan: A lost decade. *Review of Economic Dynamics*, 5(1):206 – 235.

Helpman, E. (1998). *General Purpose Technologies and Economic Growth*. The MIT Press, Elhanan Helpman (ed). Cambridge, MA.

Helpman, E. and Rangel, A. (1999). Adjusting to a new technology: Experience and training. *Journal of Economic Growth*, 4(4):359–383.

Henderson, D. J. and Parmeter, C. F. (2009). Imposing economic constraints in nonparametric regression: survey, implementation, and extension. In *Advances in Econometrics: Nonparametric*. Q. Li and J. S. Racine (Eds.), 25, Bingley, UK: Emerald, 433-469.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*. Cambridge University Press.

Henderson, D. J. and Russell, R. R. (2005). Human capital and convergence: A production-frontier approach. *International Economic Review*, 46(4):1167–1205.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Huang, C. and Liu, J.-T. (1994). Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5(2):171–180.

Hultberg, P. T., Nadiri, M. I., Sickles, R. C., and Hultberg, P. T. (1999). An international comparison of technology adoption and efficiency: A dynamic panel model. *Annales d'conomie et de Statistique*, (55/56):449–474.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688.

Iyer, K. G., Rambaldi, A. N., and Tang, K. K. (2008). Efficiency externalities of trade and alternative forms of foreign investment in oecd countries. *Journal of Applied Econometrics*, 23(6):749–766.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics*, 25(1):40–75.

Jin, H. and Jorgenson, D. W. (2010). Econometric modeling of technical change. *Journal of Econometrics*, 157(2):205 – 219.

Jondrow, J., Lovell, C. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19:233 – 238.

Kaufmann, S. and Schumacher, C. (2013). Bayesian estimation of sparse dynamic factor models with order-independent identification. Working Papers 13.04, Swiss National Bank, Study Center Gerzensee.

Kneller, R. and Stevens, P. A. (2006). Frontier technology and absorptive capacity: Evidence from oecd manufacturing industries. *Oxford Bulletin of Economics and Statistics*, 68(1):1–21.

Koop, G., Leon-Gonzalez, R. R., and Strachan, R. W. (2010). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29(2):224–242.

Koop, G., Osiewalski, J., and Steel, M. F. J. (1999). The components of output growth: A stochastic frontier analysis. *Oxford Bulletin of Economics and Statistics*, 61(4):455–487.

Koop, G., Osiewalski, J., and Steel, M. F. J. (2000). Modeling the sources of output growth in a panel of countries. *Journal of Business and Economic Statistics*, 18(3):284–299.

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARS. *Journal of Applied Econometrics*, 28(2):177–203.

Kroese, D. P. and Chan, J. (2014). *Statistical Modeling and Computation*. Springer, New York.

Kroese, D. P., Taimre, T., and Botev, Z. I. (2011). *Handbook of Monte Carlo*. John Wiley & Sons Ltd.

Kumar, S. and Russell, R. R. (2002). Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence. *The American Economic Review*, 92(3):527–548.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(12):201 – 211.

Kumbhakar, S. C., Ghosh, S., and McGuckin, J. T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business and Economic Statistics*, 9(3):279–286.

Kumbhakar, S. C. and Lovell, C. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.

Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics*, 137(1):1–27.

Kumbhakar, S. C. and Wang, H.-J. (2005). Estimation of growth convergence using a stochastic production frontier approach. *Economics Letters*, 88(3):300 – 305.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, 15(3):261–274.

Li, Q. (1999). Nonparametric testing the similarity of two unknown density functions: local power and bootstrap analysis. *Journal of Nonparametric Statistics*, 11(1-3):189–213.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics*, 4(1):25–38.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer Publishing Company, Incorporated.

Maddison, A. (1987). Growth and slowdown in advanced capitalist economies: Techniques of quantitative assessment. *Journal of Economic Literature*, 25(2):649–698.

Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2007 edition.

Magnus, J. R., Powell, O., and Prfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154(2):139 – 153.

Marron, J. S. (1988). Automatic smoothing parameter selection: A survey. *Empirical Economics*, 13(3):187–208.

Mastromarco, C. and Ghosh, S. (2009). Foreign capital, human capital, and efficiency: A stochastic frontier analysis for developing countries. *World Development*, 37(2):489 – 502.

Mastromarco, C. and Simar, L. (2015). Effect of FDI and time on catching up: New insights from a conditional nonparametric frontier analysis. *Journal of Applied Econometrics*, 30(5):826–847.

McCracken, M. W. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. Technical report, Federal Reserve Bank of St. Louis, working paper 2015-012A.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–44.

Muirhead, R. (2005). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.

O'Donnell, C. (2015). Using information about technologies, markets and firm behaviour to estimate and decompose a proper productivity index. *Journal of Econometrics*.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409):66–72.

Park, B. U., Simar, L., and Zelenyuk, V. (2008). Local likelihood estimation of truncated regression and its partial derivatives: Theory and application. *Journal of Econometrics*, 146(1):185 – 198.

Park, B. U., Simar, L., and Zelenyuk, V. (2015). Categorical data in local maximum likelihood: theory and applications to productivity analysis. *Journal of Productivity Analysis*, 43(2):199–214.

Park, B. U., Simar, L., and Zelenyuk, V. (2017). Nonparametric estimation of dynamic discrete choice models for time series data. *Computational Statistics and Data Analysis*, 108:97 – 120.

Parmeter, C. F. and Kumbhakar, S. C. (2014). Efficiency analysis: A primer on recent advances. *Foundations and Trends in Econometrics*, 7(3-4):191–385.

Parmeter, C. F., Sun, K., Henderson, D. J., and Kumbhakar, S. C. (2014). Regression and inference under smoothness restrictions. *Journal of Productivity Analysis*, (41):111–129.

Pettenuzzo, D., Koop, G., and Korobilis, D. (2016). Bayesian compressed vector autoregressions. Working Papers 103, Brandeis University, Department of Economics and International Businesss School.

Peyrache, A. and Rambaldi, A. N. (2012). A state-space stochastic frontier panel data model. CEPA Working Papers Series WP012012, School of Economics, University of Queensland, Australia.

Quah, D. (1996). Twin peaks: Growth and convergence in models of distribution dynamics. *The Economic Journal*, 106(437):1045–1055.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99 – 130.

Rasti, B., Sveinsson, J. R., and Ulfarsson, M. O. (2014). Wavelet-based sparse reduced-rank regression for hyperspectral image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6688–6698.

Ray, S. C. and Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *The American Economic Review*, 87(5):1033–1039.

Reifschneider, D. and Stevenson, R. (1991). Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review*, 32(3):715–723.

Reinsel, G. (1983). Some results on multivariate autoregressive index models. *Biometrika*, 70(1):145–156.

Reinsel, G. C. and Velu, R. (1988). *Multivariate Reduced-Rank Regression: Theory and Applications*. Springer.

Robin, J.-M. and Smith, R. J. (2000). Tests of rank. *Econometric Theory*, 16:151–175.

Romer, C. D. and Romer, D. H. (1989). Does monetary policy matter? a new test in the spirit of Friedman and Schwartz. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pages 121–184. National Bureau of Economic Research, Inc.

Rubinstein, R. Y. and Kroese, D. (2004). *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. New York: Springer-Verlag.

Sala-i Martin, X. X. (1996). The classical approach to convergence analysis. *The Economic Journal*, 106(437):1019–1036.

Schott, J. (2005). *Matrix Analysis for Statistics*. Wiley, New York.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton University Press.

Sickles, R. C., Hao, J., and Shang, C. (2015). *Panel Data and Productivity Measurement*. In The Oxford Handbook of Panel Data: Oxford University Press.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London:Chapman and Hall.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis*, 28(3):183–201.

Simar, L., Keilegom, I. V., and Zelenyuk, V. (2017). Nonparametric least squares methods for stochastic frontier models. *Journal of Productivity Analysis*, 47(3):189–204.

Simar, L., Lovel, A., and Vanden Eeckaut, P. (1994). Stochastic frontiers incorporating exogenous influences on efficiency. Technical report, Discussion Paper No 9403, Institut de Statistique, Universite Catholique de Louvain, Lovain-la-Neuve, Belgium.

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.

Simar, L. and Wilson, P. W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36(1):33–53.

Simar, L. and Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, 36(1):1–20.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts. *European Economic Review*, 36(5):975 – 1000.

Smolny, W. (2000). Post-war growth, productivity convergence and reconstruction. *Oxford Bulletin of Economics and Statistics*, 62(5):589–606.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, 12(4):1285–1297.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):307 – 325.

Ter Braak, C. J. F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, 55(3):519–531.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Tsionas, E. G. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics*, 21(5):669–676.

Vahid, F. and Issler, J. V. (2002). The importance of common cyclical features in var analysis: a Monte-Carlo study. *Journal of Econometrics*, 109(2):341 – 363.

van Es, B. (1991). Likelihood cross-validation bandwidth selection for nonparametric kernel density estimators. *Journal of Nonparametric Statistics*, 1(1-2):83–110.

van Pottelsberghe de la Potterie, B. and Lichtenberg, F. (2001). Does foreign direct investment transfer technology across borders? *The Review of Economics and Statistics*, 83(3):490–497.

Velu, R. P., Reinsel, G. C., and Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika*, 73(1):105–118.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, 21(2):326–357.

Vounou, M., Nichols, T. E., and Montana, G. (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage*, 53(3):1147 – 1159.

Wang, H.-j. and Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, 18(2):129–144.

Wang, M. and Wong, M. C. S. (2012). International R&D transfer and technical efficiency: Evidence from panel study using stochastic frontier analysis. *World Development*, 40(10):1982–1998.

Warne, A., Coenen, G., and Christoffel, K. (2017). Marginalized predictive likelihood comparisons of linear gaussian state-space models with application to DSGE, DSGE-VAR and VAR models. *Journal of Applied Econometrics*, 32(1):103–119.

Winston, G. C. (1974). The theory of capital utilization and idleness. *Journal of Economic Literature*, 12(4):1301–1320.

Young, A. (1995). The tyranny of numbers: Confronting the statistical realities of the East Asian growth experience. *The Quarterly Journal of Economics*, 110(3):641–680.

Young, A. T., Higgins, M. J., and Levy, D. (2008). Sigma convergence versus beta convergence: Evidence from U.S. county-level data. *Journal of Money, Credit and Banking*, 40(5):1083–1093.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Hoor of Bruno de Finetti (North-Holland, Amsterdam) 233-243*.