COMPUTATIONAL ECONOMETRICS WITH APPLICATIONS TO HOUSING MARKETS

by

CHARLES RAHAL

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Birmingham Business School
College of Social Sciences
The University of Birmingham
February 11, 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis develops a variety of econometric approaches in order to examine model specification and weighting choices in an era when computational power and efficiency is becoming less of a binding constraint and 'big data' becomes widely available. We apply our methodological contributions to housing markets, given their recently verified importance with regard to international financial stability. In particular, we develop two original forecasting routines based on high dimensional datasets, take simulated and empirical approaches to specifying large sets of spatial weighting matrices and we contrast structural (panel and single country based) vector autoregressions which are identified and analyzed in a number of ways. The appendices are reserved for introducing published econometric software and replication codes developed as a by-product of the main body of work.

Declaration

The material contained within this thesis has not previously been submitted for a degree at the University of Birmingham or any other university. The research reported within this thesis has been conducted by the author unless indicated otherwise.

Signed .....

ACKNOWLEDGMENTS

I would firstly like to thank Professor John Fender and Dr. Marco Barassi who provided unerring supervision. Professor Colin Lizieri and Professor Peter Byrne managed to instill a curiosity for research during my undergraduate studies at the University of Reading. Importantly, they allowed me to appreciate the necessity of a more rigorous, computational approach to real estate economics. Following this, I would like to formally acknowledge the large number of academics at the University of Birmingham who guided my transition from surveying to economics, and for enhancing my economics education more generally during Graduate Diploma, Masters and PhD level taught modules. This list might include, but is not limited to, lectures given by Ralph Bailey, Jackie Smith, Joanne Ercolani, Martin Jensen, Anindya Banerjee, John Fender, Peter Sinclair and William Pouliot. I would also like to thank Professor Sheri Markose of the University of Essex who invited me to attend a series of modules and events at the University as a visiting student in the field of Agent Based Modelling. Gareth Thomas contributed to several useful discussions on computational efficiency.

Financial support from the ESRC is very gratefully acknowledged. Gratitude is extended to several close personal friends, some of whom were part of the Economics PhD program at the University of Birmingham and otherwise, for their everlasting patience: Maria Koslovskaya, Graeme Pearce, Jon Tucker, Daniel Hayes, Emma Blyth, Mary Dawood, Marco Giansoldati, Neil Owen, Kemal Bagzhibagli, Naoki Funai, Wira Kusuma and Alexandros Rigos. Comments on individual chapters were received from Christophe Görtz, Afrasiab Mirza and Jane Binner, for which I am thankful. Special thanks are given to Boris Hoffman, Igor Masten, Rob Hyndman and James LeSage for making their replication files available which were hugely informative. Third party editing was undertaken by Timothy Davies (*clarityandprecision*).

DEDICATION

Most importantly, I would like to dedicate this short collection of works to my parents.

CONTENTS

1 Introduction **1**

1.1 Motivating the computational approach 2

1.2 Motivating the application to housing markets 3

1.3 Programming and replication files 4

1.4 The structure of the thesis 5

2 Computational Forecasting with FRED **9**

2.1 Introduction 10

2.2 FRED and Dallas FED 13

2.3 Models and metrics 15

2.3.1 Benchmarks 15

2.3.2 Iterated Multi-Step (IMS) 16

2.3.3 Direct Multi-Step (DMS) 17

2.3.4 Model selection and weighting 17

2.4 Evaluation and analysis 22

2.4.1 How does the AIC perform against the BIC? 22

2.4.2 How does the IMS perform against the DMS? 24

2.4.3 How do rolling windows perform against recursive? 26

2.4.4 Forecast errors 27

2.4.5 Forecast stability 29

2.4.6 Beating the benchmark 31

2.4.7 Forecast error and variance 33

2.5 Conclusion 35

3 Wait Before you Weight: A computational approach to spatio-temporal misspecification **38**

3.1 Introduction . . . 39
3.2 Literature . . . 42
3.3 Simulated analysis based on Dubin (1998) . . . 47
3.4 Empirical applications . . . 53
3.4.1 The SEM model . . . 53
3.4.2 The datasets . . . 54
3.4.3 Variation over delta . . . 57
3.4.4 Variation across alpha space . . . 66
3.4.5 Direct comparison between specifications . . . 70
3.5 Conclusion . . . 73

4 Housing Markets and Unconventional Monetary Policy **75**

4.1 Introduction and policy background . . . 76
4.2 Existing literature . . . 79
4.2.1 Housing in general equilibrium models . . . 79
4.2.2 Unconventional monetary policy shocks . . . 80
4.2.3 Single country VAR models of housing . . . 81
4.2.4 Panel VAR models of housing . . . 82
4.3 Data . . . 83
4.3.1 Quarterly dataset . . . 84
4.3.2 Monthly dataset . . . 84
4.4 Methodology . . . 86
4.4.1 Model specifications . . . 87
4.4.2 Estimation and identification strategies . . . 88
4.5 Results . . . 92
4.5.1 Monthly data . . . 92
4.5.2 Quarterly data . . . 97
4.6 Conclusion . . . 102

5 Housing Market Forecasts with Factor Combinations **105**

5.1 Introduction . . . 106

5.2 Existing literature . . . 109

5.3 Data . . . 114

5.3.1 Number of factors . . . 115

5.4 Methodology . . . 116

5.4.1 Factor extraction . . . 116

5.4.2 Classical models . . . 122

5.4.3 Factor augmented models . . . 123

5.4.4 Candidate models . . . 125

5.4.5 Estimation techniques . . . 126

5.4.6 Forecasting approach . . . 127

5.5 Results . . . 130

5.5.1 Forecast error . . . 130

5.5.2 The best performing model . . . 135

5.5.3 Beating the ARIMA benchmark . . . 136

5.5.4 Forecast error as a ratio against the benchmark . . . 137

5.5.5 Forecast error and underlying variance . . . 137

5.6 Conclusion . . . 139

6 Conclusion **141**

Appendix **144**

A A replication of Dubin (1998) **144**

A.1 Introduction . . . 145

A.2 Distance matrix . . . 145

A.3 The model and matrices used . . . 146

A.4 Six outputs which are identical to Dubin (1998) . . . 146

A.5 The replication code . . . 148

B A guide to the StatFact EViews Add-in **149**

B.1 Introduction . . . 150

B.2 Econometric procedure . . . 150

B.3 Use of the software . . . 151

B.4 Illustrative examples . . . 153

B.4.1 Stock and Watson (2005) . . . 153

B.4.2 Hurlin dataset . . . 154

B.5 Conclusion and further work . . . 154

C A guide to the MacroTrans EViews Add-in **156**

C.1 Introduction . . . 157

C.2 The command line and GUI . . . 158

C.3 An example using the Stock and Watson (2005) . . . 160

C.4 Conclusion and further work . . . 160

Bibliography **162**

LIST OF FIGURES

2.1 Descriptive plots of the 50 state-level series 10
2.2 Average quarterly pre-crisis house price growth 11
2.3 How does the ARIMA model behave? 16
2.4 Lags selected by the top 250 models 18
2.5 Variables selected by the top 250 models 20
2.6 How often can the BIC outperform the AIC at single point forecasts? 23
2.7 How often can DMS outperform the IMS at single point forecasts? 25
2.8 How often can recursive outperform rolling windows at single point forecasts? . . 27
2.9 Average MAPE for IMS: Recursive - BIC models 28
2.10 Forecast stability 30
2.11 How frequently can models beat the benchmark? 32
2.12 Benchmark ratios: IMS forecasts, BIC selection and recursive estimation 33
2.13 Forecast variance 34
2.14 Average MAPE to standard deviation ratio 35

3.1 $W_A^{\alpha,\delta}$ Specifications 50
3.2 $W_B^{\alpha,\delta}$ Specifications 51
3.3 $W_C^{\alpha,\delta}$ Specifications 52
3.4 Spatial distribution of the datasets 55
3.5 Temporal distribution of the datasets 55
3.6 The four main variables: San Francisco 56
3.7 The four main variables: Ohio 56
3.8 Hausmans p-value: delta variation 59
3.9 Moran's I: delta variation 60

3.10 MAPE: delta variation . . . 61
3.11 $\bar{R}^2$: delta variation . . . 63
3.12 $\lambda_{d,m}^{T,S,\alpha}$: delta variation . . . 65
3.13 Ohio coefficient variation: $W_A^{\alpha,\delta}$ . . . 67
3.14 Ohio coefficient variation: $W_B^{\alpha,\delta}$ . . . 67
3.15 Ohio coefficient variation: $W_C^{\alpha,\delta}$ . . . 67
3.16 San Francisco coefficient variation: $W_A^{\alpha,\delta}$ . . . 67
3.17 San Francisco coefficient variation: $W_B^{\alpha,\delta}$ . . . 67
3.18 San Francisco coefficient variation: $W_C^{\alpha,\delta}$ . . . 67
3.19 Movement of spatial statistics in α space . . . 68
3.20 Movement of coefficients in α space . . . 70
3.21 Movement of evaluation metrics in α space . . . 71

4.1 Economic indicators during the 'Great Recession' . . . 76
4.2 Central bank responses to the crisis . . . 78
4.3 Quarterly and monthly house prices . . . 85
4.4 Quarterly and monthly residential investment . . . 85
4.5 Quarterly and monthly mortgage rates . . . 85
4.6 Lag length selection . . . 88
4.7 Monthly house prices . . . 93
4.8 Monthly residential investment . . . 94
4.9 Monthly mortgage rates . . . 95
4.10 Monthly mean group responses for Model One . . . 96
4.11 Quarterly house prices . . . 98
4.12 Quarterly residential investment . . . 99
4.13 Quarterly mortgage rates . . . 100
4.14 Quarterly mean group responses for Model One . . . 101

5.1 Time plots of the four variables to be forecast . . . 107
5.2 Stationary factors (recursively extracted) . . . 118
5.3 Non-stationary and cyclical factors (recursively extracted) . . . 119

5.4 Percent of variance explained by each factor 120
5.5 Scree plots 121
5.6 Average forecast errors 131
5.7 Errors across horizons 132
5.8 Average forecast errors 133
5.9 Forecast error and variance 138

A.1 Replication of Figures A1-B3 on page 314-315 of Dubin (1998) 147

B.1 The GUI representation of the command line syntax 152
B.2 Graphical output from Stock and Watson (2005) data 153

C.1 The schematic of transformations 157
C.2 The GUI representation of the command line syntax 159
C.3 An example using the Stock and Watson (2005) dataset 160

LIST OF TABLES

2.1 Mean squared forecast error ratio of AIC vs BIC 24

2.2 Mean squared forecast error ratio of IMS vs DMS forecast generation 25

2.3 Mean squared forecast error ratio of rolling vs recursive windows 26

3.1 Variation across key coefficients in δ space 66

3.2 Comparative evaluation of the $\bar{R}^2$ 72

3.3 Comparative evaluation of the MAPE 72

4.1 Overview of the five model specifications 87

4.2 Identification of an unconventional monetary policy shock 91

4.3 The magnitiude and peak of mean group responses: Monthly data 92

4.4 Monthly data: Maximum absolute response of each element 97

4.5 The magnitiude and peak of mean group responses: Quarterly data 101

4.6 Quarterly data: Maximum absolute response of each element 102

5.1 Twelve models for evaluation 134

5.2 The percent of times a model performs best 136

5.3 Outperforming the Hyndman and Khandakar (2008) algorithm 137

5.4 A ratio to the Hyndman and Khandakar (2008) algorithm 138

B.1 A summary of the input options 152

B.2 Bai and Ng (2002) output for the Hurlin dataset 154

C.1 A summary of the input options 159

List of Abbreviations

Abreviations	Meaning
AIC	Akaike Information Criteria
API	Application Program Interface
ARIMA	AutoRegressive Integrated Moving Average
ALFRED	ArchivaL Federal Reserve Economic Data
BIC	Bayesian Information Criteria
BMA	Bayesian Model Averaging
BVAR	Bayesian Vector AutoRegression
DFM	Dynamic Factor Model
DMS	Dynamic Multi-Step
DSGE	Dynamic Stochastic General
EA	Euro Area
ECM	Error Correction Model
FAVAR	Factor Augmented Vector Autoregression
FECM	Factor Error Correction Model
FHFA	Federal Housing Finance Association
FGLS	Feasible Generalize Least Squares
FRED	Federal Reserve Economic Data
GUI	Graphical User Interface
IC	Information Criteria
IMS	Iterated Multi-Step
LSTAR	Local Spatio-Temporal AutoRegression
MAPE	Mean Absolute Percentage Error
MSA	Metropolitan Statistical Area
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PVAR	Panel Vector AutoRegression
RMSE	Root Mean Squared Error
SAIC	Smooth Akaike Information Criteria
SAR	Spatial AutoRegression
SEM	Spatial Error Model
STAR	Spatio-Temporal AutoRegression
SVAR	Structural Vector AutoRegression
TVP	Time-Varying Parameter
VAR	Vector AutoRegression
VECM	Vector Error Correction Model
ZLB	Zero Lower Bound

CHAPTER 1

INTRODUCTION

1.1 Motivating the computational approach

In the editorial introducing the first issue of *Econometrica* in 1933, Ragnar Frisch surmises that 'quantitative economic theory and statistical observation is the essence of econometrics'. To achieve such a suitable level of quantitative rigour, econometricians have used computational techniques since the development of early software at the University of Cambridge in the 1950s and the creation of the first operative stored-program electronic computer there. Throughout the 1950s, the use of computers in economics remained sparse, and in Adelman (2007), the author discusses her programming of the Klein-Goldberger model on an IBM 650 in 1956-57, and how machine time was charged at $600 per hour when her monthly salary at Berkeley was a mere $200. Ted Anderson later claimed that early research by the Cowles Foundation was curtailed due to 'the limitations of computational ability' (as discussed in the 'ET Interview' with Peter Phillips, documented in Phillips (1986)), rather than econometric methodology. However, progress is no better measured than by comparing Bean (1929), which claims that 'a single four variable regression analysis for 30 observations would take about 8 hours of work', with Hendry and Doornik (1999), which claims that 'at the end of 1996, our computer does about 30,000 per second (and much more accurately)'. At the time of the writing, a cursory couple of lines of code in a popular 'off the shelf' econometric package (i.e. without parallelization) displays an ability to generate a million random variables and regress each of them on a single independent variable in under two minutes using an average personal desktop computer.

While the primary applications of this thesis are focused on property markets, the overarching econometric theme is that of model selection and model combination in a heavily computational environment. In a study written by some of the leading authorities in the field, Doornik *et al.* (2002) identify four main sources of econometric demand on computational power: 1.) the high dimensional nature of the data, 2.) the non-stationary nature of macroeconometric series, 3.) the computation of likelihoods in non-linear models, 4.) the complexity of the relationships between variables and the need to select from a larger set of candidate variables. The ensuing work to be discussed addresses methods and approaches involving all of these econometric features. Two chapters focus on high dimensional forecasting, one uses a computational likelihood based spatial estimator, and several feature nonstationary data. Two of the appendices discuss

the production of, and guidance on using econometric software focused on dimension reduction or series transformation, and one focuses on the code to replicate a well cited paper from the relevant literature. There is a strong focus throughout on the comparison of the econometric methodologies which we utilize and develop further, with a 'classical' type benchmark model or approach.

1.2 Motivating the application to housing markets

Housing markets provide a rich environment within which to examine the importance of the econometric issues at hand.[1] From a policy perspective, their corresponding volatility is now more important than ever considering the global recession of 2008-2012. Given the dramatic increases in financial innovation, housing markets are not only scrutinized by active market participants, but those involved in the securitization of the financial products underlying related mortgages, as well as central bankers and regulatory agencies monitoring economic stability and mortgage pricing and underwriting. Pre-crisis literature such as Foster and Van Order (1985) and Crawford and Rosenblatt (1995) shows that the volatility of house prices is a major determinant of not only mortgage default, but also prepayment risk. Non-recourse loans, prevalent in some developed economies, facilitated the advent of situations where the optimal strategy for home-buyers was to exercise their option to default in the scenario of falling home values, further depressing the market (the frictionless option-theoretic model).

Models which forecast real house price inflation are able to give an impression of the overall macroeconomy, and aid in the formulation of policies regarding mortgage defaults, property taxes, and a range of consumption, investment and other household decisions. A significant body of research focuses on the housing wealth effect (including an unpublished Master's dissertation of this author). In particular, two influential papers (Case *et al.* (2005) and Case *et al.* (2013)) co-authored by one of the 2013 Nobel Memorial Prize laureates, Robert Shiller, claim that a ten percent increase in housing wealth increases consumption somewhere in the region of 1-1.5%,

[1]The rationale for the small four or five dimensional systems used in this thesis are based upon a core macroeconometric model of housing which lies at the heart of almost every macroeconometric paper which analyzes similar questions (two recent, frequently cited papers in the thesis are Gattini and Hiebert (2010) and Goodhart and Hofmann (2008)). They contain fundamental variables related to house prices such as income, residential investment, and a mortgage market variable as a proxy for cost of financing.

but not at all for financial (stock) wealth.[2] Renowned economist William Buiter discusses the conditions required for a change in house prices to have no aggregate effect in an over-lapping generations model (Buiter (2010)).

House prices are important for the creation of new residential investment (construction starts) which is able to stimulate the supply side of an economy and create real activity. In addition to this, higher house prices lead to an increase in wealth for homeowners, and, with the ever increasing opportunity for mortgage equity withdrawal, this provides an additional channel for housing to affect consumption. The importance of housing markets in the business cycle, given the global recession of 2008-2012, will not be understated for a substantial period of time. Influential econometrician Edward Leamer emphasizes the importance of housing, entitling his paper (Leamer (2007)) on the predictive power of housing starts 'Housing IS the business cycle', in response to an earlier paper entitled 'Housing and the business cycle' (Davis and Heathcote (2001)). On similar subject matter, albeit with a different approach (a dynamic stochastic partial equilibrium model), Ren and Yuan (2014) also explain 'Why the housing sector leads the whole economy', and an influential policy-orientated paper (Iacoviello and Neri (2010)) uses a DSGE model to show that 'housing market spillovers are non-negligible, concentrated on consumption rather than business investment, and have become more important over time'. In a more time-series based macro-forecasting exercise, Stock and Watson (2003) specifically cite the role of house prices in forecasting inflation.

1.3 Programming and replication files

The research in the subsequent chapters is programmed in a variety of statistical programming languages, most notably EViews (Chapters 2 and 5), MATLAB (Chapters 2, 3 and 5) and RATS (Chapter 4). Appendices B and C explicitly document the construction of and guide the use of two (now official) EViews add-ins. However, R and Gauss were also used, predominantly to understand existing functions which were then translated into other languages as required. All functions written by others are acknowledged throughout the text where appropriate, although

[2]Professor Shiller also writes extensively about the importance of housing markets and their importance in the global economy in Irrational Exuberance (Shiller (2005)) and The Subprime Solution (Shiller (2008)).

the majority of code in each chapter was either significantly modified or re-written from first principles. There was a strong desire to make all programs executable on a local machine, as opposed to a distributed or high performance system, and therefore, where applicable, the replication files make use of parallel processing tools and vectorization in order to achieve all possible efficiency gains.

Given the recent high profile controversies in economics[3] and political science[4], it has never been more important to develop a transparent strategy to provide replication files. An example of the code for a replication is explicitly provided in Appendix A, and the two *.aipz* which accompany Appendix B and C are available directly via Eviews at `http://www.eviews.com/Addins/StatFact.aipz` and `http://www.eviews.com/Addins/MacroTrans.aipz`. All replication files and functions are available on request from the author at `charlierahal@gmail.com`.

1.4 The structure of the thesis

The thesis is structured as follows. Chapter 2 takes a computational approach to forecasting 50 state level (Federal Housing Finance Association) house price indexes using a new computational combinatorial forecasting algorithm. This chapter is ordered first because it best represents the intentions of the thesis, in that it takes a rigorous approach to questions such as the size of model set, the method in which models are selected and weighted across, the dimensionality of the data and the complexity of calculation involved. Chapter 3 follows in a similar vein in that it specifically examines the effect of estimating an almost continuous model set across multiple dimensions, with both simulated and empirical approaches to some common spatial econometrics questions (albeit almost infinitely more computational in design). Chapter 4 changes focus to a small macroeconometric model, taking a structural VAR based approach to consider how housing markets have responded to unconventional monetary policy shocks. We consider several different issues, such as the effect of using different aggregations and representations of the same data, the difference between estimation strategies (panel vs non-panel), different identification

[3]Such as the 'Reinhart and Rogoff' scandal.

[4]Such as the alleged falsification of data and funding sources in, and retraction of "When contact changes minds: An experiment on transmission of support for gay equality" in the journal *Science* by Michael Lacour and Donald Green.

strategies (zero and sign restrictions), various model specifications, and method of impulse response analysis (mean, median and 'median target' estimators) across eight OECD countries. Chapter 5, as in Chapter 3, considers a series of small macroeconometric models for housing markets in a cross-country setting, this time taking a reduced form forecasting approach (more of a classical 'horse race' scenario between competing model classes), selecting and weighting across large sets of factor models.

Appendix A features a slice of code and some details on the replication of part of an influential paper in the spatial literature, and, as mentioned previously, Appendices B and C relate to EViews add-ins which specifically relate to high dimension dataset reduction and transformation building on the underlying code required for Chapters 2 and 5. Chapter 6 briefly concludes and suggests an overarching avenue for future research. While the thesis can be considered as a series of inter-related yet independent pieces of work, each chapter is prefaced by a short prelude which introduces the work and describes its relationship with other chapters.

PREFACE TO CHAPTER 2

In this chapter,[5] we undertake the work defined as the purest form of econometrics according to Paul Samuelson and William Nordhaus: "sifting (sic) through mountains of data to extract simple relationships." (Samuelson and Nordhaus, 2009, p.5). We develop and explain a new dynamic 'big data' forecasting algorithm which uses 128,732 series from the Federal Reserve Economic Data (FRED) database with the intention of generating the smallest possible forecast errors. We evaluate it through a pseudo real-time forecasting exercise involving 50 all-transactions house price indexes at US state level, with an out-of-sample period across the recent Great Recession. We estimate approximately 482 million bivariate models drawn from a cross-section of dimension sparsely seen in the econometric literature, selecting and weighting variables across different horizons, forecast approaches, model set sizes, rolling and recursive windows and combination metrics. Our approach performs extremely favourably against each of the three commonly used benchmarks which we evaluate it against, including a comparable big data macroeconometric (a Factor Augmented Vector Autoregression) model which is outperformed in approximately 80% of instances. We aim to provide several stylized results to guide big-data forecasting in macroeconometrics and related fields. The algorithm shows a general preference for Bayesian Model Averaging (BMA hereafter) over Smooth Akaike Information Criteria (SAIC hereafter) weightings, iterated multi-step (IMS hereafter) over dynamic multi-step (DMS hereafter) forecast generation and for recursive over rolling estimation windows. There is further strong evidence of forecast error increasing in the volatility of underlying series, and we show that our forecast errors are relatively stable, except for during the global recession of 2008-2012.

The majority of macroeconometric forecasting exercises which utilize high dimensional data (more often than not using some variant of the Stock and Watson (2002a) dataset) have centred

[5]This work is co-authored with my second supervisor, Dr. Marco Barassi, who contributed primarily to the theoretical expositions.

on three competing methods. The first two of the three - starting with dynamic factor models in the form of Stock and Watson (2002a) and large Bayesian shrinkage methods as per Banbura *et al.* (2010) - have dominated the recent applied literature. However, neither of these methods are as yet truly able to fully utilize real 'big data', instead focusing on macro-panels with a cross-sectional dimension of between approximately one hundred and four hundred variables. Combined with the increasing importance and availability of economics related APIs (application program interfaces) such as FRED, the first substantive chapter of this thesis attempts to provide evidence in favour of a combinatorial approach more akin to the machine learning literature, rarely studied in econometrics (other than a couple of exceptions discussed in Section 2.1). However, such an approach is increasingly viable as the cross-sectional dimension of macro-panels increases along with computational power.

CHAPTER 2

COMPUTATIONAL FORECASTING WITH FRED

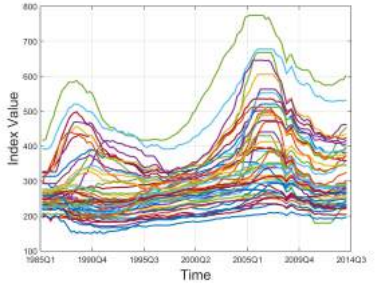
2.1 Introduction

If we consider the increased computational power and the emerging availability of large 'macro-panel' datasets over recent years, the small volume of academic literature which focuses on forecasting key macroeconomic series with either computationally or structurally detailed models may be considered surprising. Historically, the majority of work on this topic has focused on low dimensional models: either univariate, single equations, or vector autoregressions with a handful of variables. More recent work which involves a factor-based or Bayesian VAR approach typically only utilizes macro-aggregates within a cross-sectional framework of between 100-400 variables. However, this chapter provides a way of incorporating all potentially useful information which these approaches may omit. The problem of forecasting state-level house prices is especially relevant to policy-makers and is made inherently more difficult by regional disparities: a problem our approach intends to overcome.

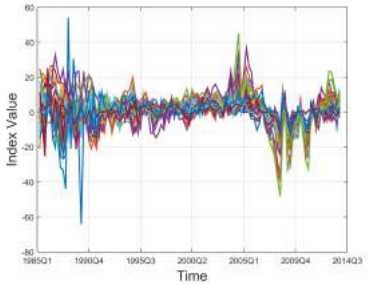
Specifically, we aim to contribute to the literature by detailing a pseudo out-of-sample experimental design which uses a large subset of FRED (depending on series availability and frequency) to forecast state-level house price indexes using bivariate single equation dynamic multi-step (DMS) forecasts and vector auto-regression based iterated multi-step (IMS) forecasts. We generate over 482 million models spread across each of 50 indexes, for 28 recursive periods, six lag lengths, two different forecasting strategies, six horizons and 14,363 (filtered from 128,732) series from FRED. Forecasts are generated from these models which are then compared across two different combination weight metrics using rankings of a subset of the model set based on two different information criteria (IC). Descriptive plots of the house price series to be fore-

Figure 2.1: Descriptive plots of the 50 state-level series

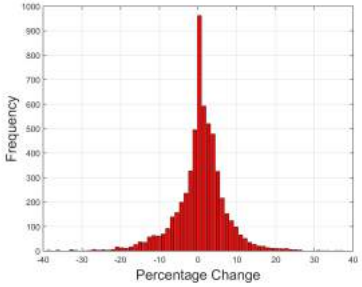
2.1.1 Levels



2.1.2 Annualized returns

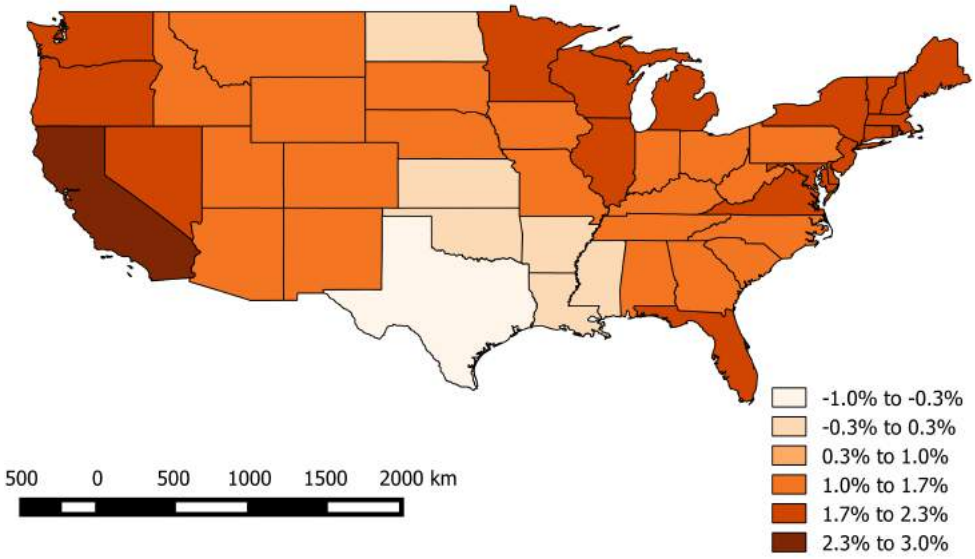


2.1.3 Annualized returns



Source: Federal Housing Finance Association (FHFA).

Figure 2.2: Average quarterly pre-crisis house price growth



cast (and a histogram of annualized returns) can be seen in Figure 2.1, and a map of average annualized pre-crisis returns can be seen in a state by state choropleth map in Figure 2.2.

The macroeconometric literature traditionally focuses upon a small monetary system consisting of one real, one nominal and one financial variable (e.g. Rudebusch and Svensson (1999)), potentially augmented with factors (in a Factor Augmented Vector Autoregression - Marcellino *et al.* (2005), or Factor Error Correction Model - Banerjee *et al.* (2014) setting). However, considering events of the previous decade, the importance of house prices in these contexts can no longer be under-stated. The primary reason for this is the ever-increasing proportion of housing as an asset on the household balance sheet. This is complemented by financial innovations which allow various types of securitization and equity withdrawal, potentially creating important housing wealth effects. Furthermore, the aforementioned securitization markets have been shown to have huge effects on global financial flows and are inherently dependent on accurate assessments of future house price movements for the efficient functioning of a range of related derivative instruments. House prices also form a vital component in the decision-making processes relating to 'housing starts', an increasingly important variable in the economy responsible for huge spillovers and the creation of demand for natural resources and employment (see, for example Leamer (2007)).

Our context of forecasting a multitude of different state indexes is a suitably interesting

environment within which to apply our computational routine, as opposed to the majority of house price forecasting research which often focuses on regional sub-markets (e.g. five U.S. states as per Crawford and Fratantoni (2003) in a regime-switching context) or a broader aggregate (such as the euro area as per Gattini and Hiebert (2010)). A particular paper of interest which uses a large macro-panel dataset is that of Bork and Møller (2012), who take a computational combination approach to show that a simple three factor augmented autoregressive model can consistently outperform autoregressive and combination benchmarks which forecast percentage returns of the real FHFA index. In addition to this, Gupta and Kabundi (2010) use a range of different models (FAVAR, VAR, DSGE) and estimators (OLS, Bayesian) to forecast the deflated Census Bureau House Price Index, showing the ability of the DSGE model of Iacoviello and Neri (2010) to forecast the downturn in 2006:Q2.

Within the econometric and forecasting literature, model averaging is not a recent innovation, with plenty of papers supporting the notion that improvements can be made in terms of forecast accuracy relative to model selection. In research which explicitly compares the performance of IMS and DMS forecasts, Marcellino *et al.* (2006)(building on the work by Chevillon and Hendry (2005)), claim that IMS forecasts are more efficient when correctly specified, and that direct forecasts are more robustly resistant to model misspecification. They empirically show (in a context similar to the current chapter) that IMS forecasts have lower mean squared forecast errors than their DMS counterparts, especially as the horizon lengthens, albeit with a significantly smaller cross-section.

Our choice of information criteria is inspired by their dominance in the model selection literature such as in Granger and Jeon (2004), or in Stock and Watson (1998), who show that models selected by the Akaike Information criteria (AIC hereafter) and Bayesian Information Criteria (BIC hereafter) do not forecast better than an AR(4). The initial Akaike (1973) paper suggests using the exponentiated AIC as model weights - an idea expanded upon in the biometric literature by Buckland *et al.* (1997) - a technique now known as SAIC. Using BMA for forecast combinations was pioneered by Zellner (1962), with approximations using the BIC occurring in several papers since. One of the most relevant papers to this chapter is that of Banternghansa and McCracken (2011), who undertake a real-time forecasting exercise with ALFRED. The authors show that model averaging does not always improve forecast accuracy relative to the

more standard BIC-based approach to model selection, with significantly different results for real and nominal variables.

However, the paper which has the closest resemblance to the chapter at hand is undoubtedly Bork and Møller (2015). Using the same set of 50 state-level dependent variables, the authors examine the performance of recently proposed, and closely related Dynamic Model Averaging and Dynamic Model Selection techniques of Raftery *et al.* (2010). The main advantage of such an approach is that it allows the set of predictors (and weights assigned to them) to change at each period, 'showing substantial forecasting improvements over simple benchmark regressions', as in this chapter. While the sample and forecasting strategy in Bork and Møller (2015) is generally different to those in this chapter, the results encourage a more sophisticated approach than the historical one of using a small number of observed variables, such as price to income ratios, or population growth. In particular, our algorithm has the potential to include regional data, in addition to aggregate data at a state specific level, should such a predictor exhibit a good in-sample fit. For this reason, we feel that a panel of US state-level house price indexes has a cross-sectional dimension rich enough to assess the universal applicability of our methods.

2.2 FRED and Dallas FED

Our routine uses a list of 128,732 tickers from the FRED database which includes all quarterly and monthly series. The list is available as part of the supplementary material related to this chapter, which also includes a program to download and filter the series from FRED API, a program which undertakes our transformation algorithms as described below, a program to run all of the model estimations and forecast combinations, and a program to analyze and automatically generate all of the figures used in this chapter.[1] We then filter down and obtain a dataset of 14,363 series which not only have no missing observations between 1985:Q1 and 2014:Q2, but also contain no zero values (out of our desire to construct percentage changes) and those that are nothing more than a vector of constants (to avoid singularity as we include a constant in exactly half of our models). We also ensure that there are no duplicate series,

[1] Also available on request from the corresponding author is a library of all downloaded and filtered series.

potentially due to repetition of a series through different FRED sources. This leaves us with 9,582 quarterly and 4,781 monthly series. We combine this with 50 of the 320 house price housing indexes made available on FRED by the FHFA, which broadly measure the movement of single-family residential house prices. The indexes are a weighted repeat sales index, created from mortgage transactions on single-family properties purchased or securitized by Fannie Mae or Freddie Mac. In particular, we focus on the 'all-transactions' indexes, which, unlike the headline 'purchase only' indexes, contain appraisal data obtained from the Enterprises. The indexes are not seasonally adjusted.

At this point, it is important to note that we follow the mainstream forecasting literature, transforming our dependent variable of interest ($y_{s,t+j}$) - which represents house price indexes across states (s) over time at each recursion (where t denotes the initial in-sample and j each subsequent recursion) for each forecast horizon (h), to the annualized log real growth rate as:

$$y^h_{s,t+j} = \frac{400}{h} ln(\frac{y^h_{s,t+j}}{y_{s,t-h+j}}) \tag{2.1}$$

From this, the forecasts in levels are reconstructed for evaluation. For the approaches in Section 2.3.1 and 2.3.2, we set $h = 1$ (by design) and 'roll' the models forward, but in Section 2.3.3, it is dependent on how many steps ahead we are forecasting. All house prices which are not being forecast (of which there are naturally 49 for a specific index) are added back into the FRED set of conditioning variables on a state by state basis. For all of the FRED variables which pass the filter (14,363), we transform them in a variety of ways depending on their properties in an automated procedure grounded firmly within the macroeconometric literature.[2] If the series contains all elements greater than zero, logarithms are taken. We then run augmented Dickey Fuller tests at the 5% level, and if the series is non-stationary and has been logged, we take first differences, and if it has not been logged, we take one period percentage changes. If the series is stationary, it remains in either logged levels or levels, depending on their properties. In Section 2.3 we discuss the competing methods of forecasting, variable selection and weighting, in addition to the estimation of benchmark models.

[2]This procedure is described in Appendix C.

2.3 Models and metrics

2.3.1 Benchmarks

In order to evaluate our combinatorial based methodology, we estimate three main benchmark models. The first of these is a univariate model which is estimated based on the algorithm of Hyndman and Khandakar (2008). While not the exact same step-wise approach, it estimates models in the possible parameter space of an ARIMA(p,d,q) of the form:

$$\Phi(B)(1-B^d)y^h_{s,t+j} = c + \theta(B)\varepsilon_{t+j} \tag{2.2}$$

where $\{\varepsilon_{t+j}\}$ is a white noise process with mean zero and variance σ^2, B is a backshift operator and Φ and θ are polynomials of order p and q. We omit any models which yield non-stationary or non-invertible ARMA roots. We search the model space where $p = 0, \ldots, 6$, $d = 0, 1, 2$ and $q = 0, \ldots, 6$, first running a KPSS test, and then estimating the ARMA(p,q) model on the suitably transformed data (at the 5% level[3]). The resulting model is the one which satisfies the above condition and minimizes *either* the AIC or the BIC, depending on which set of models we are comparing it with, and, like all of the following models in Section 2.3, the specifications are determined and estimated at each of 28 recursive steps for each country. The frequencies at which various values of p and q are chosen are shown in Figure 2.3, and indicate the well known parsimonious properties of the BIC in comparison to the AIC. As Chapter 5 will show, this is an extremely competitive benchmark in general. We also estimate a second univariate 'mean' benchmark: a model which uses a constant only, as in Bork and Møller (2015). In addition, we employ an extremely commonly used big-data benchmark in the form of a FAVAR(1,k) model: a factor augmented vector autoregression with one (to be comparable to the bivariate models discussed below) static factor ($f_{1,t+j}$):[4]

$$\Phi(L)z_{t+j} = c + \varepsilon_{t+j} \tag{2.3}$$

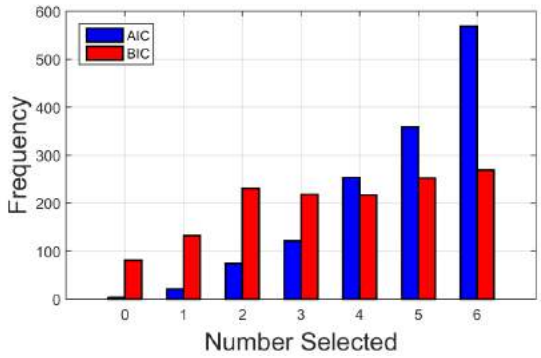
[3]By considering only integer values of d, we omit the possibility of fractionally integrated ARMA processes.

[4]The factor is extracted recursively using the principle components method of Stock and Watson (2002a) using the software described in Appendix A.

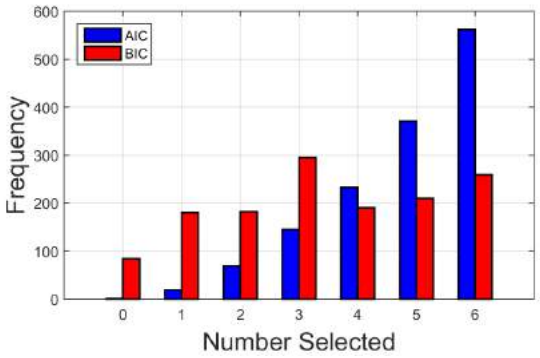
where $z_{t+j} = [y^h_{s,t+j}, f_{1,t+j}]'$. Where Φ(L) denotes a 2×2 matrix polynomial in the lag operator: $\Phi(L) = (I_2 - \Phi_1 L - \ldots - \Phi_p L^k)$, it has a dimension which is determined by the AIC and BIC, depending on which weighting scheme we are evaluating against (to be discussed in Section 2.3.4).

Figure 2.3: How does the ARIMA model behave?

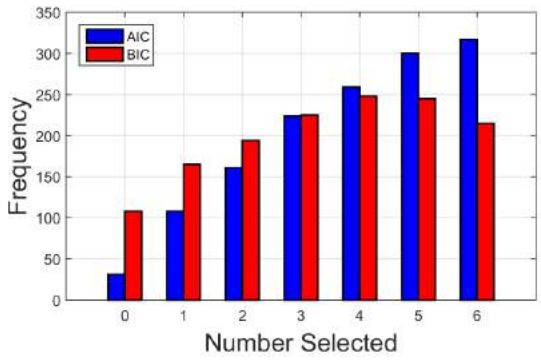
2.3.1 Recursive (number AR parameters chosen)



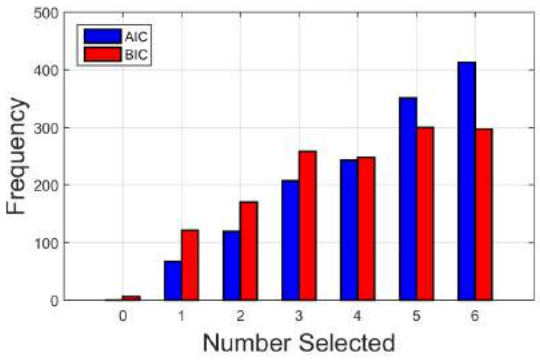
2.3.2 Recursive (number MA parameters chosen)



2.3.3 Rolling (number AR parameters chosen)



2.3.4 Rolling (number MA parameters chosen)



2.3.2 Iterated Multi-Step (IMS)

Considering first our IMS approach, at each recursive period, we re-estimate our bi-variate vector autoregressive model (again using ordinary least squares) with the same representation as Equation 2.3, but now where $z_{t+j} = [y^h_{s,t+j}, x_{i,t+j}]'$. Allow $x_{i,t+j}$ to represent the conditioning variables which pass the aforementioned filtering restrictions where i=1,. . .,14,363, for k=1,. . .,6. Allow c to represent a vector of deterministic terms and ε_{t+j} be a vector of innovations such that $\varepsilon_{t+j} \sim N(\mu, \Omega)$ where μ is a vector of zeros and Ω is a variance-covariance matrix. We recursively estimate the k different Φ matrices of coefficients and generate our forecasts which are h steps

ahead for our eight horizons ($h = 1, \ldots, 8(H)$) to obtain the future values of our dependent variable via an iterated sequence of 1 through H step ahead forecasts: $\hat{y}^1_{s,t+j}, \ldots, \hat{y}^8_{s,t+j}$. Unlike the DMS approach, the same parameter estimates are used regardless of horizon, but for each recursive iteration in our sample, we re-estimate Φ once. One important point to note is that we estimate all variants of each type of model with and without a constant, doubling the size of our model set to select from.

2.3.3 Direct Multi-Step (DMS)

For our DMS forecasts, conversely, a separate model is re-estimated each period for each horizon. For each value of h, the model takes the form:

$$y^h_{s,t+h+j} = \alpha_y + \Phi_y(L)y^h_{s,t+j} + \Phi_x(L)x_{i,t+j} + \varepsilon_{y^h_{s,t+j+h}} \tag{2.4}$$

where Φ_y and Φ_x denote coefficient vectors of the appropriate dimension for $y_{s,t+j}$ and $x_{i,t+j}$ respectively. In this approach, ('direct') forecasts are generated as:

$$\hat{y}^h_{s,t+h+j} = \hat{\alpha}_y + \hat{\Phi}_y(L)y^h_{s,t+j} + \hat{\Phi}_x(L)x_{i,t+j}. \tag{2.5}$$

For clarification; the DMS and IMS strategies contain the exact same model specification in terms of lag dynamics and variable composition, but differ only in the way that the forecasts themselves are generated, and are identical at the $h = 1$ horizon. The single equation models used for DMS models are naturally nested within a VAR model, but are estimated as single equations as there is likely a computational gain in doing so. This is due to the fact that we do not need to estimate half of the coefficients in the 'system', and re-estimates of ϕ are required for each horizon which becomes significant when estimated hundreds of millions of times.

2.3.4 Model selection and weighting

The two information criteria which we use to rank our models are the AIC and the BIC, which are then used to allocate the necessary weights via the SAIC and approximate BMA procedures.

For model $m^{k,i}$ with k lags for variable i, the information criterion are calculated as:

$$\text{BIC}_{m^{k,i}_{s,t+j}} = \frac{-2 \times ln(\mathcal{L}_{m^{k,i}_{s,t+j}})}{t+j} + \frac{g \times ln(t+j)}{t+j} \tag{2.6}$$

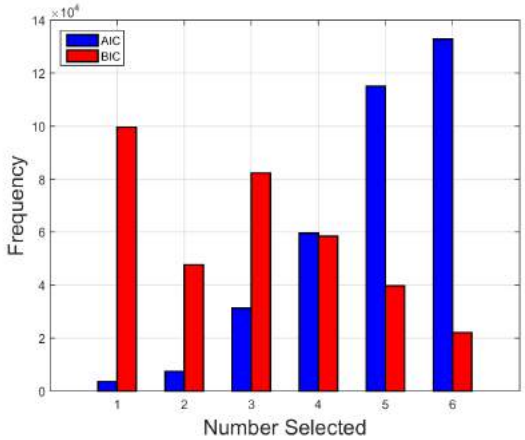
$$\text{AIC}_{m^{k,i}_{s,t+j}} = \frac{-2 \times ln(\mathcal{L}_{m^{k,i}_{s,t+j}})}{t+j} + 2\frac{g}{t+j} \tag{2.7}$$

where g represents the number of estimated coefficients, $t+j$ is the sample size; the initial in-sample (t); and the number of iterations completed in the algorithm (j). $\mathcal{L}_{m^{k,i}_{s,t+j}}$ is the log likelihood of model $m^{k,i}_{s,t+j}$. All models are estimated using ordinary least squares. This is not only in the interest of computational simplicity and efficiency, but also the fact that with moderate samples, Bayesian estimation of the model shows no improvement over OLS in papers such as Baumeister and Kilian (2011). At this point, it may be appropriate to note that $t+j$ is a simplified notation which relates to the recursive framework only, and does not adequately capture the sampling scheme in operation for our rolling procedure (which also updates the sample starting point by j periods also, resulting in a time dimension of fixed length).

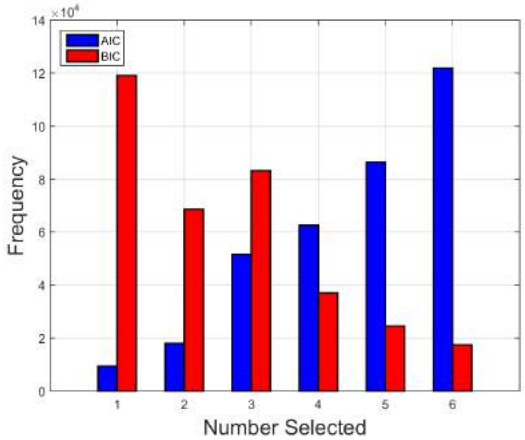
Eight forecasts for evaluation (four single-equation and four VAR-based DMS and IMS forecasts for both rolling and recursive windows, for each of the two information criteria,) are models from the 'full' model set ($\mathbb{M}$) which minimize information criteria (two for the AIC and two for the BIC). These models, represented here as $m^{k,i}_{s,t+j} \in \mathbb{M}$, may be thought of as equivalent to a model selection routine which corresponds to the arg min of a vector of AIC/BIC values. We

Figure 2.4: Lags selected by the top 250 models

2.4.1 Recursive (number of lags selected)



2.4.2 Rolling (number of lags selected)



estimate $\mathbb{M}$ across each index, for each recursion using rolling and recursive samples (discussed below), with and without a constant, and across each recursion and for each of six lags, meaning we generate $482,596,800$ models in total. The frequency of the number of lags which is being selected over these models is shown in Figure 2.4. The frequency at which a specific variable is selected as one of the best 250 models per recursion per state is shown in Figure 2.5. While we analyze a total of 50 states over 28 recursions, we can see that some variables enter into the largest model set more frequently than a mere 1,400 times. This is because variables can enter in the form of models which do or do not contain a constant, or as part of the same deterministic and variable specification across a multitude of different lag lengths. The panels of Figure 2.5 are indicative of the fact that this occurs frequently, with the same variable entering into the same model set repeatedly as part of different model specifications.

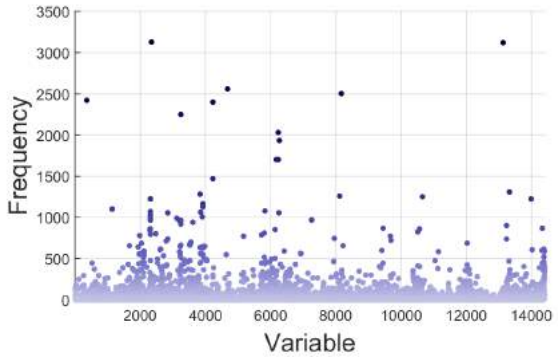
We follow the approach of papers such as Timmermann (2006) and Banternghansa and McCracken (2011), 'trimming' $\mathbb{M}$ to calculate weights. However, our approach differs in that we take sequential ordering of in-sample performance to the weighting problem. Specifically, we order each of the estimated models in the full set according to their AIC/BIC value so as to obtain an ordinal ranking. This ranking is used to sequentially extrapolate models to create new model sets. Out of the $|\mathbb{M}|$ total model sets, we obtain new model sets $(M^n_{s,t+j})$ of size $|M^n_{s,t+j}|$ based on the aforementioned ranking, where $n = 1, \ldots, 250$ are drawn from the full model set such that $m^{k,i}_{s,t+j} \in M^n_{s,t+j} \subset \mathbb{M}$.

To clarify: for each of the two ICs, our approach first constructs the model set $(M^1_{s,t+j})$ which has the lowest AIC/BIC value and therefore only contains one model (with weight $w^{m^{k,i}_{s,t+j} \in M^1_{s,t+j}}_{s,t+j} = 1$). The second model set $(M^2_{s,t+j})$ is a weighted combination of the two models with the lowest two AIC or BIC values, where the first model $(M^1_{s,t+j})$ set enters into the second model set $(M^2_{s,t+j})$, with weights determined as per Equation 2.8 below. Our algorithms continue in this fashion until we have a set of model sets which incorporate information from up to the top 250 of all models.[5] If any two models (to twelve significant figures) contain exactly the same in-sample IC value and straddle the cutoff for inclusion in our trimmed model sets, one of the two models is chosen randomly. A key feature of our approach is the fact that

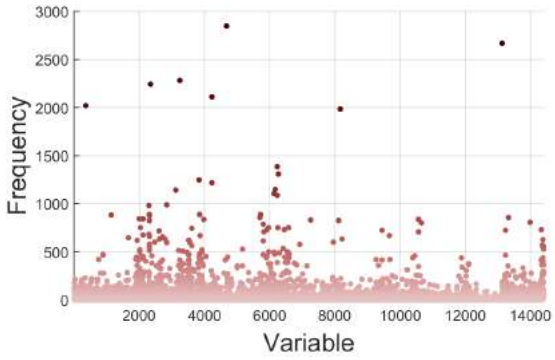
[5] While a discretionary choice, this number represents a value which allows us to report all interesting dynamics of the model set size with sufficient clarity while allowing the set to remain tractable. Other values do not dramatically alter any reported results.

Figure 2.5: Variables selected by the top 250 models

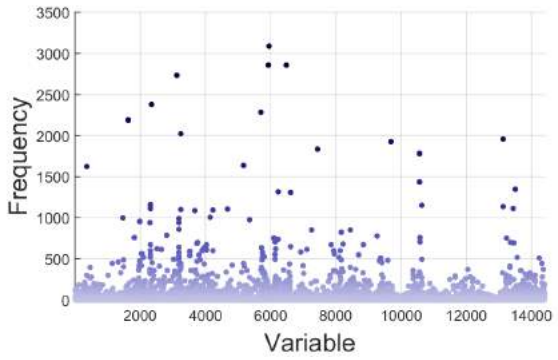
2.5.1 Recursive: AIC



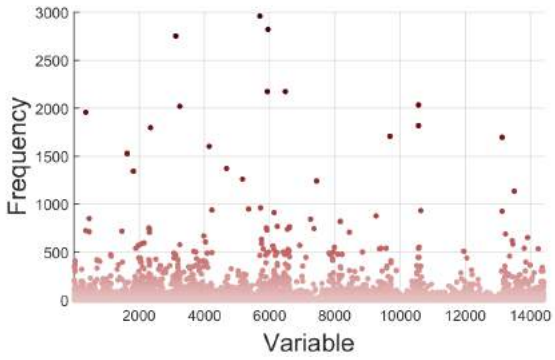
2.5.2 Recursive: BIC



2.5.3 Rolling: AIC



2.5.4 Rolling: BIC



our set of total estimable IMS models has an exactly equivalent DMS representation which will produce identical forecasts in the one-step horizon ($h = 1$), allowing us to calculate a single set of AIC/BIC values for sorting and weighting purposes, reducing the computational burden involved. Specificaly – each of the single equation models are estimated for the h=1 horizon, and then, if the AIC/BIC is in the top 250, further estimates are made for all single equation DMS forecasts, and all VAR IMS forecasts.

Figure 2.5 outlines which variables are selected by the Rolling (AIC), Recursive (AIC), Rolling (BIC) and Recursive(BIC) schemes. The variables with the highest frequency are (denoted in the list of predictors as): 5945, 2353, 5706 and variable 4667 respectively, which are FRED tickers `hnotmsq027s`, `arnptl`, `mdothfrafhatp1t4fr` and `bpfapi02mxq637n`. These are: Total mortgages to households, Non-performing loans (90+ days past), debt outstanding by Federal and related agencies, and portfolio investment assets for Mexico. The first three, at least, are especially relevant to forecasting U.S. housing markets over the recent financial period, and are able to provide a strong indication of the fact that our algorithm is selecting economically

meaningful variables to weight across.

To derive the weights, following Garratt *et al.* (2008), we approximate the marginal likelihood of a given model using its AIC/BIC values in order to create SAIC/approximate BMA weights. For each recursion for each country, we take our set of 172,356 (14,363×6×2) AIC/BIC values (number of series multiplied by lag length for a constant and no constant) and assign a weight to each model $m^{k,i}_{s,t+j}$, in each model set ($M^n_{s,t+j}$) at each recursion using the formula for each of the two ICs:

$$w^{m^{k,i}_{s,t+j} \in M^n_{s,t+j}}_{s,t+j} = \frac{exp(-\frac{1}{2} IC^{m^{k,i}_{s,t+j}}_{t+j})}{\sum_{m^{k,i}_{s,t+j} \in M^n_{s,t+j}} exp(-\frac{1}{2} IC^{m^{k,i}_{s,t+j}}_{s,t+j})} \tag{2.8}$$

It is important to again note the equivalence of $M^1_{s,t+j}$ to a model selection routine with $w^{m^{k,i}_{s,t+j} \in M^1_{s,t+j}}_{s,t+j} = 1$, and $w^{m^{a,b} \notin M^1_{s,t+j}}_{s,t+j} = 0, \forall a \neq k$ and $\forall b \neq i$. For each of the subsequent model sets, the model from $M^1_{s,t+j}$ will still be weighted higher, but with an increasingly diminishing weight as it represents a lower fraction of the denominator in Equation 2.8. Due to the equivalence of IMS and DMS for the one-step horizon, the model selection and weights chosen are identical for both of our forecasting approaches. Having derived weights, forecast combinations ($\hat{Y}^{M^n_{s,t+j}}_{s,t+h+j}$) are calculated from individual model forecasts combined across the trimmed sets ($\hat{Y}^{m^{k,i}_{s,t+j}}_{s,t+h+j}$), for recursion j, as:

$$\hat{Y}^{M^n_{s,t+j}}_{s,t+h+j} = \sum_{m^{k,i}_{s,t+j} \in M^n_{s,t+j}} w^{m^{k,i}_{s,t+j} \in M^n_{s,t+j}}_{t+h+j} \times \hat{Y}^{m^{k,i}_{s,t+j}}_{s,t+h+j} \tag{2.9}$$

It is these forecast combinations ($\hat{Y}^{M^n_{s,t+j}}_{t+h+j}$) for each of the different model set sizes which we then evaluate in Section 2.4 below. Specifically, we generate ('pseudo real-time') out-of-sample forecasts, updating the information going into the forecasts on a recursive or rolling scheme as detailed above, but we use data which is not real-time (and has been subjected to revisions by the provider). As an example (weights in brackets), for a model set size of 5, house prices in Washington State in 2006Q1 might be best forecast by Washington State unemployment (0.39), a Federal Funds Rate type variable (0.21), a West Coast proxy for residential investment or housing starts (0.15), house prices in a neighboring state such as Oregon (0.13), and some other miscellaneous variable related to housing (0.12). Alternatively, it could be five neighboring house

price indexes with an approximately equal weight each.

2.4 Evaluation and analysis

We systematically analyze the forecast performance of all of our models in Section 2.4.1 to 2.4.7, discussing the evaluation metrics more specifically in each subsection as applicable. In Sections 2.4.1 to 2.4.3, we directly compare specific competing aspects of forecast performance: SAIC and BMA, IMS and DMS and rolling and recursive windows. Following this, we take the best performing combination of each of the above modelling approaches (BMA, IMS, recursive), and consider the average forecast error across each state, averaged over recursion, in order to consider the effect of changing the model set size ($n = 1, \ldots, 250$ for $M^n_{s,t+j}$). We then measure the stability of our forecasts over one of the most volatile periods for U.S. housing markets for three specific model set sizes ($M^1_{s,t+j}, M^{125}_{s,t+j}, M^{250}_{s,t+j}$). In the penultimate subsection (Section 2.4.6) we show the dominance of our models over the aforementioned univariate benchmarks, and finally we show how forecast error increases as a function of the volatility of the underlying housing markets (Section 2.4.7).

2.4.1 How does the AIC perform against the BIC?

In Sections 2.4.1 to 2.4.3, each corresponding figure represents the number of times a specific method, information criterion or sampling approach is able to generate a forecast closer in absolute distance ($|\hat{Y}^{M^n_{s,t+j}}_{s,t+h+j} - y^h_{s,t+h+j}|$) to the realized value of the target state index. Figures 2.6 to 2.8 detail the averaged binary metric at specific (point) horizons for all model sets. Tables 2.1 to 2.3 present a more conventional evaluation in the form of a ratio of forecast errors between the two metrics which are aggregated across state, recursion and horizon.

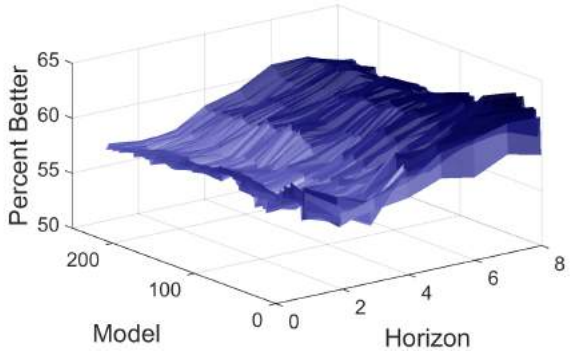
The first comparison which we discuss is displayed in Figure 2.6 which reports the percentage of times the BMA outperforms the SAIC in terms of providing a lower forecast error. Each of the four panels in Figure 2.6 represent one of the four broader modelling strategies which we employ, and shows a direct comparison of the information criteria for each strategy across a total of 1,400 possible comparisons (from all 50 states across 28 recursions) for each of the

isolated eight point-forecast horizons. The graphs can be interpreted as follows: the elements on the y-axis represent model sets of various sizes, from $M^1_{s,t+j}$ to $M^{250}_{s,t+j}$ where the first element represents the first model set (the 'model selection' scenario), the second is a weighting over two models, and so forth. The x-axis represents the horizon in question, and the z-axis refers to the percentage of times one modelling strategy outperforms the other (e.g. in this case: a BIC based forecast generating lower errors than a corresponding AIC forecast using the same forecasting methodology and sampling technique). Considering a direct comparison between the AIC and BIC, an interesting unreported revelation is that for the smallest model set/model selection scenario, both ICs frequently select the same one model (approximately 30% of instances) and these scenarios are omitted from the ensuing analysis.

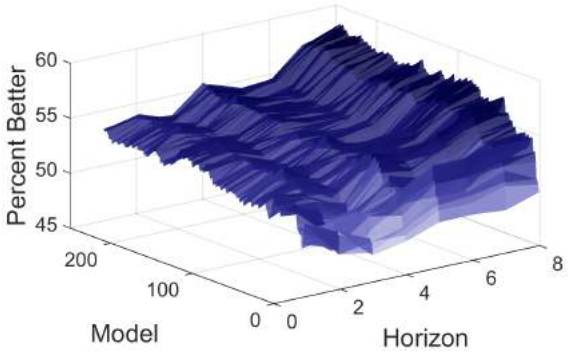
The general conclusion from this comparison is that the BIC generates superior forecasts for both VAR based IMS and single equation DMS forecasts across both rolling and recursive horizons. This result is wholly in accordance with the earlier literature such as Lütkepohl (1985) at the one period horizon (using VAR models), and papers such as Medel and Salgado (2012)

Figure 2.6: How often can the BIC outperform the AIC at single point forecasts?

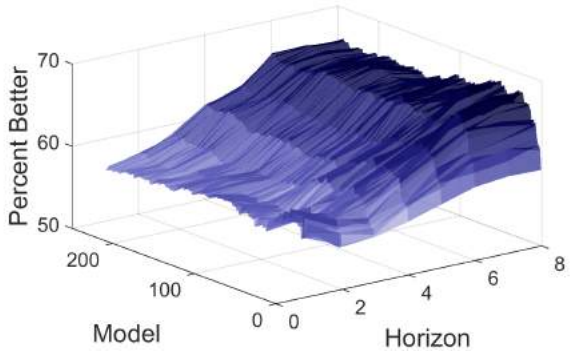
2.6.1 Rolling (IMS)



2.6.2 Recursive (DMS)



2.6.3 Rolling (DMS)



2.6.4 Recursive (IMS)

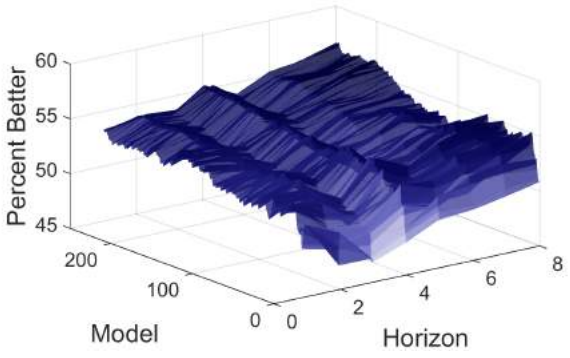


Table 2.1: Mean squared forecast error ratio of AIC vs BIC

$M^n_{s,t+j}$	Rolling (IMS)	Recursive (IMS)	Rolling (DMS)	Recursive (DMS)
n=1	0.8501	0.9830	0.8458	0.9963
n=30	0.8737	0.9635	0.8252	0.9530
n=50	0.8401	0.9435	0.8466	0.9460
n=125	0.8456	0.9438	0.8511	0.9484
n=250	0.8829	0.9195	0.8665	0.9519

in the univariate context (who both however only consider the model selection properties of the information criteria). Pesaran and Timmermann (2005) further corroborate our finding (in conjunction with Figure 2.4), claiming that 'parsimonious models tend to produce better out-of-sample forecasting performance than models with a large number of variables and parameters'.

The two information criteria are more comparable for forecasts generated from recursive windows, almost reaching parity with some of the smaller models. As we will discuss further in Section 2.4.4, an interesting dynamic which becomes apparent (and can also be seen in the subsection above, and to a lesser extent in the analysis below), is that small model sets (in the region of $M^1_{s,t+j}$ to approximately $M^{50}_{s,t+j}$) perform slightly differently to their larger counterparts. Table 2.1 shows the ability of the BMA method to not only outperform the SAIC method more frequently, but in doing so generates lower forecast errors in magnitude (as opposed to frequency) on average as well. The aforementioned near-equivalence between the SAIC and BMA for smaller model sets with recursive sampling is again apparent (with a ratio ~ 1), although for the rolling window method, the approximate BMA method produces mean squared forecast errors which are about 15% lower than the SAIC method on average.

2.4.2 How does the IMS perform against the DMS?

In the four panels of Figure 2.7 we display the percentage of times that a forecast produced by the IMS methodology can outperform the corresponding DMS counterpart. An important point to make is that there are only seven horizons to evaluate, given that at $h = 1$, we have the case that the IMS and DMS forecasts are exactly equivalent. Regarding our frequency of out-performance metric, our results concur with the primary finding of Marcellino *et al.* (2006) that

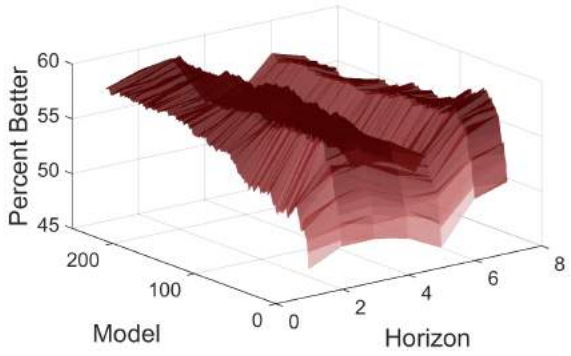
Table 2.2: Mean squared forecast error ratio of IMS vs DMS forecast generation

$M^n_{s,t+j}$	Rolling (AIC)	Rolling (BIC)	Recursive (AIC)	Recursive (BIC)
n=1	1.0217	1.0270	1.0089	0.9955
n=30	0.9906	1.0488	1.0097	1.0208
n=50	1.0481	1.0401	1.0130	1.0103
n=125	1.0895	1.0825	1.0166	1.0117
n=250	1.1404	1.1620	1.0628	1.0267

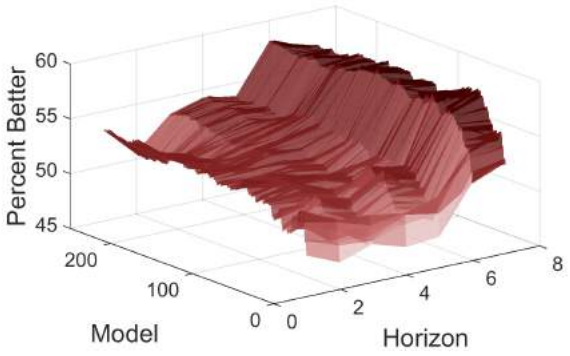
'the iterated forecasts typically outperform the direct forecasts' (although we do not further test whether this is a function of lag length (k), as in the aforementioned paper). As in Marcellino *et al.* (2006), we claim that the reason for the higher accuracy of this forecasting method is that in general, IMS are more efficient when the model is correctly specified (which we hope to achieve by specifying our models according to the ranking of the information criteria), thereby producing lower forecast errors.

Figure 2.7: How often can DMS outperform the IMS at single point forecasts?

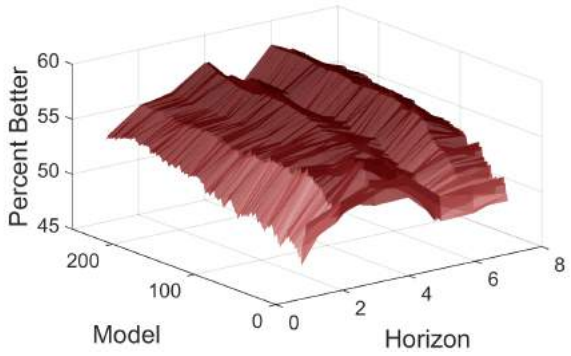
2.7.1 Rolling (AIC)



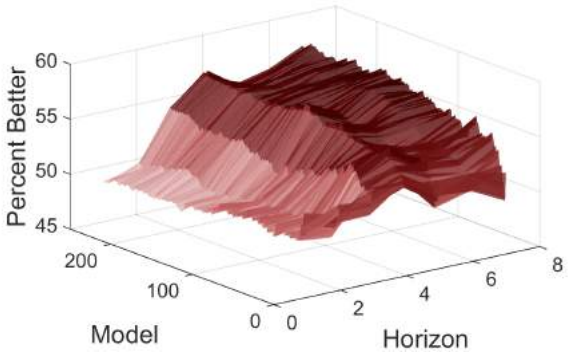
2.7.2 Rolling (BIC)



2.7.3 Recursive (AIC)



2.7.4 Recursive (BIC)



This result is apparent across both of the sample window variations and also across both of the two model weighting schemes. The results are almost uniformly consistent with outperformance of the IMS method above 50% for all model set sizes apart from some of the smaller model sets (with more frequent outperformance occurring with larger model sets). An extremely interesting observation comes through in Table 2.2. The table shows that while smaller model sets show equivalence between IMS and DMS methods, at the longer horizons, the average DMS based forecast can be substantially lower, especially for the rolling window estimation procedures.

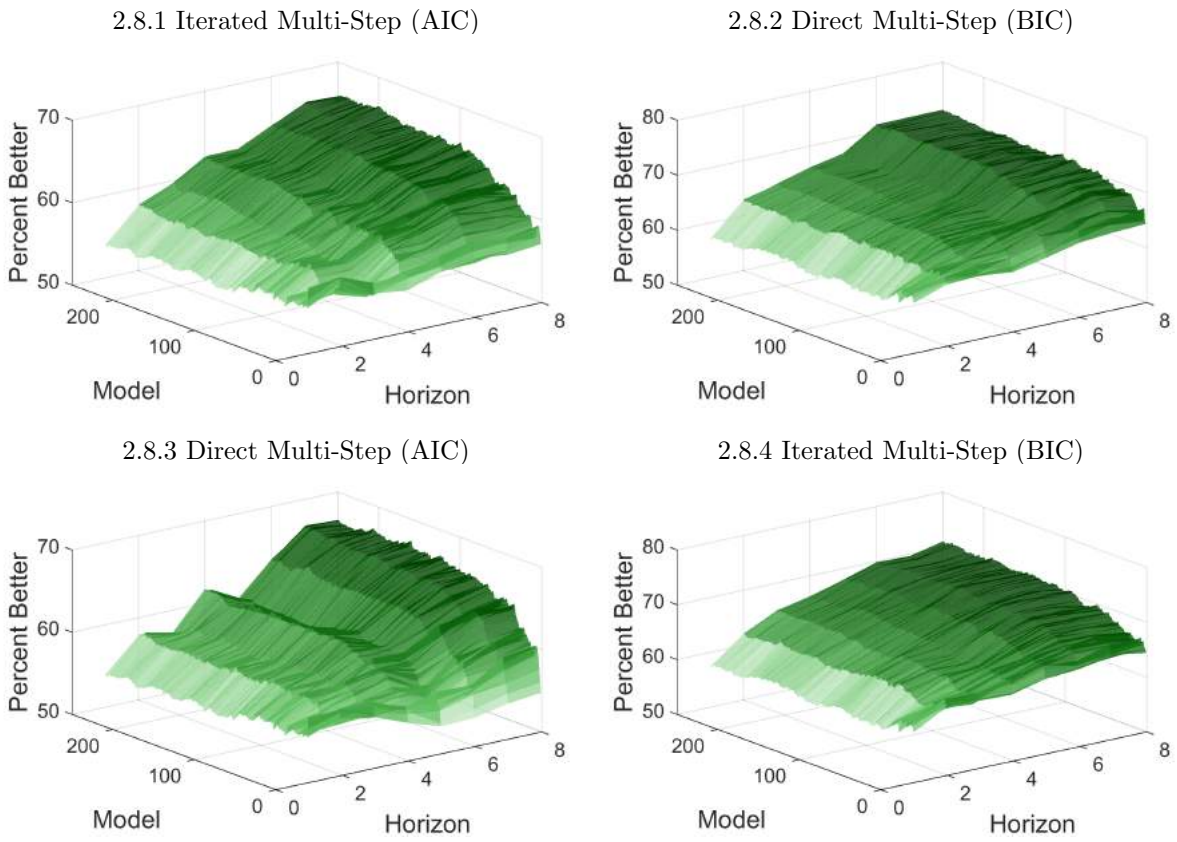
2.4.3 How do rolling windows perform against recursive?

The third and final modelling choice to explicitly compare is the performance of rolling and recursive windows. Figure 2.8 shows that it is the recursive sampling method which is preferred in each of the four panels (across IMS and DMS forecast generation and also SAIC and BMA weights also), generating lower forecast errors upwards of 60% of the time, strongly increasing in the horizon for each of the four variants. This is potentially due to the fact that we are dealing with a limited number of degrees of freedom in each model due to our desire to maximize the number of included series (the cross-section of our panel). By restricting the rolling window to a smaller sample (40 observations over ten years), we assert that the information criteria become unable to adequately select the best models from the total model set. Another potential reason for this under-performance of rolling window estimation is the fact that the models are likely to be well specified and not subject to a significant number of turning points over the period of the 'Great Moderation' of U.S. macro-aggregates (which spans more or less the entire in-sample

Table 2.3: Mean squared forecast error ratio of rolling vs recursive windows

$M^n_{s,t+j}$	IMS (AIC)	IMS (BIC)	DMS (AIC)	DMS (BIC)
n=1	0.6998	0.8092	0.7087	0.8348
n=30	0.7215	0.7957	0.7079	0.8176
n=50	0.6808	0.7646	0.7044	0.7871
n=125	0.6645	0.7417	0.7122	0.7936
n=250	0.6869	0.7154	0.7370	0.8096

Figure 2.8: How often can recursive outperform rolling windows at single point forecasts?



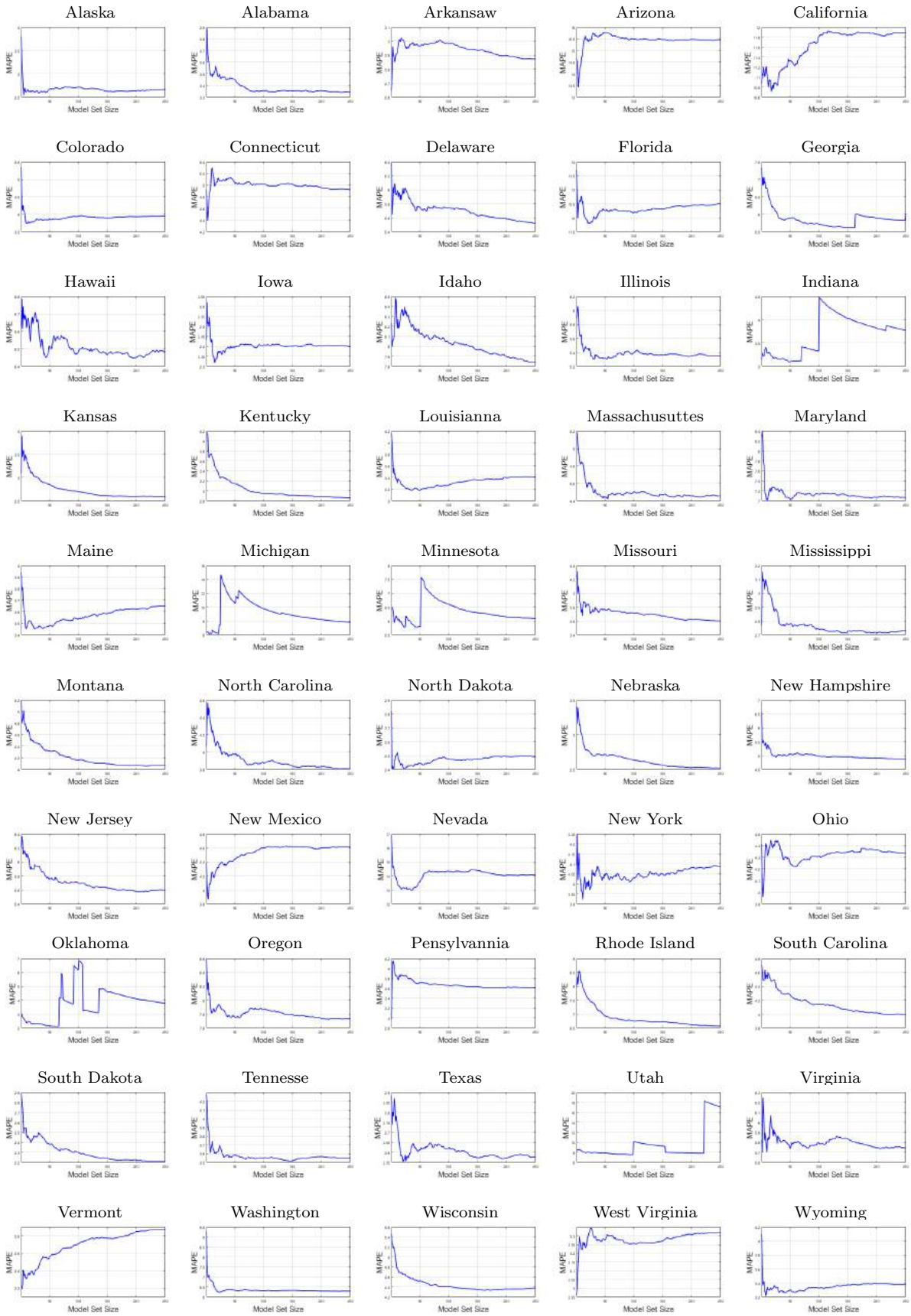
period of our FRED dataset). Theory would expect that a rolling estimation across the 'Great Recession' period would be favourable, although this appears to not be the case. Table 2.3 confirms this unambiguous result, showing that forecast errors can range from between 16.52% (n=1 for the DMS (BIC)) to as much as 33.55% lower (n=125 for the IMS (AIC)).

2.4.4 Forecast errors

In Figure 2.9, we consider the mean absolute percentage error for each eight-step forecast for each of the individual states, averaged across each of the twenty-eight recursions ($J = 28$). We choose to display only results from the IMS forecasts which are selected and weighted using the BIC, with a recursive window technique, given the results of the analysis above.[6] The MAPE is calculated as:

[6] All results for the seven other combinations of forecast generation, information criteria and sampling technique are available on request.

Figure 2.9: Average MAPE for IMS: Recursive - BIC models



$$MAPE_n = \sum_{j=1}^{J} \frac{\sum_{h=1}^{H} \frac{|\hat{Y}^{M^n_{s,t+j}}_{s,t+h+j} - y_{s,t+h+j}|}{y_{s,t+h+j}}}{J \times H} \tag{2.10}$$

The 50 panels represent each of the individual states, as indicated by the panel title. On these plots, the 250 elements of the x-axis represent a different $M^n_{s,t+j}$, where $n = 1, \ldots, 250$, increasing in the model set size as before. The first result to discuss in this context is that in approximately 65% of the states, we can see a steep reduction in forecast error when we expand our model set from $M^1_{s,t+j}$ which is the model selection equivalent set, to the region between $M^{30}_{s,t+j}$ and $M^{50}_{s,t+j}$. In a significant number of states, the forecast error appears to stabilize as the model sets become larger and larger.

However, in a small number of states, such as Michigan, Minnesota, Oklahoma and Utah, there is a large jump in forecast error at a specific model set size. This indicates that for these house price indexes, a model is being selected which may have a suitable in-sample fit, but at a specific recursion, the value which filters into the lags of the model may feature either a sharp increase or decrease in value, causing the generated forecast to be unfeasible. A further argument in favour of avoiding the larger model sets is that of computational efficiency. While all models are estimated, generating a large number of multi-step forecasts may prove computationally difficult (especially if also calculating standard errors), dependent on the statistical programming language being used. In addition to this, in two of the panels (Michigan and Vermont), it is actually the smallest model set which performs best.

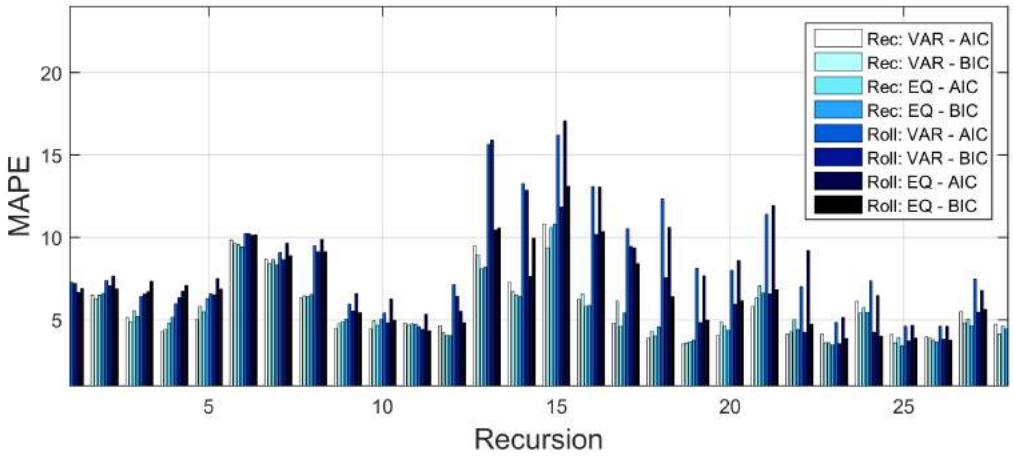
2.4.5 Forecast stability

In Figure 2.10 we can consider the stability of the forecasts across one of the most volatile periods of U.S. housing market history. The x-axis represents the time-dimension (bunched in groups of eight, corresponding to the combinations of IMS/DMS, AIC/BIC and recursive/rolling methods discussed above), and each bar represents the *average* (across $h = 1, \ldots, 8(H)$) mean absolute percentage error calculated across all fifty states ($S = 50$) generated at each horizon (j):

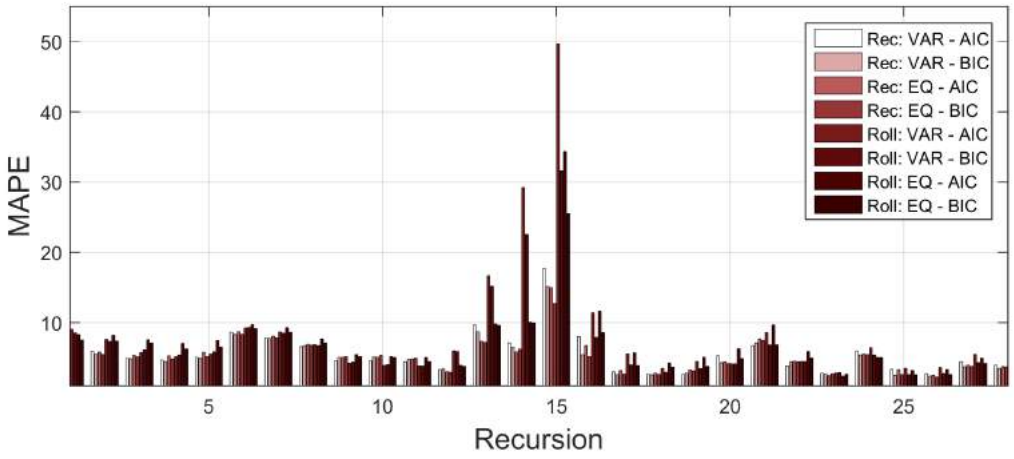
$$MAPE_{n,j} = \sum_{s=1}^{S} \frac{\sum_{h=1}^{H} \frac{|\hat{Y}^{M^n_{s,t+j}}_{s,t+h+j} - y_{s,t+h+j}|}{y_{s,t+h+j}}}{S \times H} \tag{2.11}$$

Figure 2.10: Forecast stability

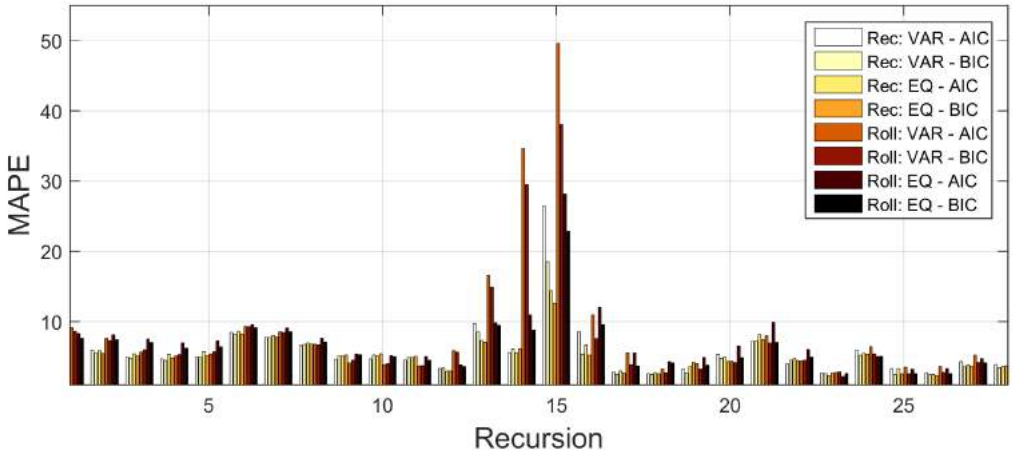
2.10.1 Smallest Model Set ($M^1_{s.t+i}$)



2.10.2 Intermediate Model Set ($M^{125}_{s.t+i}$)



2.10.3 Largest Model Set ($M^{250}_{s.t+i}$)



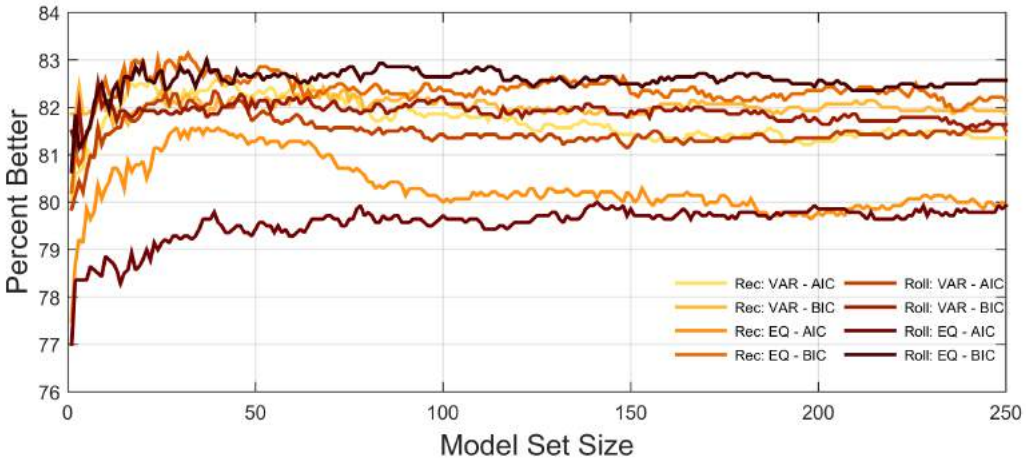
For conciseness, we confine our analysis to a comparatively small ($M^{1}_{s,t+j}$), an intermediate ($M^{125}_{s,t+j}$), and a large ($M^{250}_{s,t+j}$) model set. While the average forecast error for both the smallest and largest sets provide generally small average mean absolute percentage errors, there is a noticeable spike between 2008q3 and 2009Q2, which is a period generally regarded as the height of the financial crisis. The higher forecast error here may be attributable not only to the fact that house prices are more volatile over this period, but also that the corresponding series (i) entering into each of our bi-variate models are more volatile given the broad, all-encompassing nature of the crisis, and are not necessarily subject to the same structural break.

2.4.6 Beating the benchmark

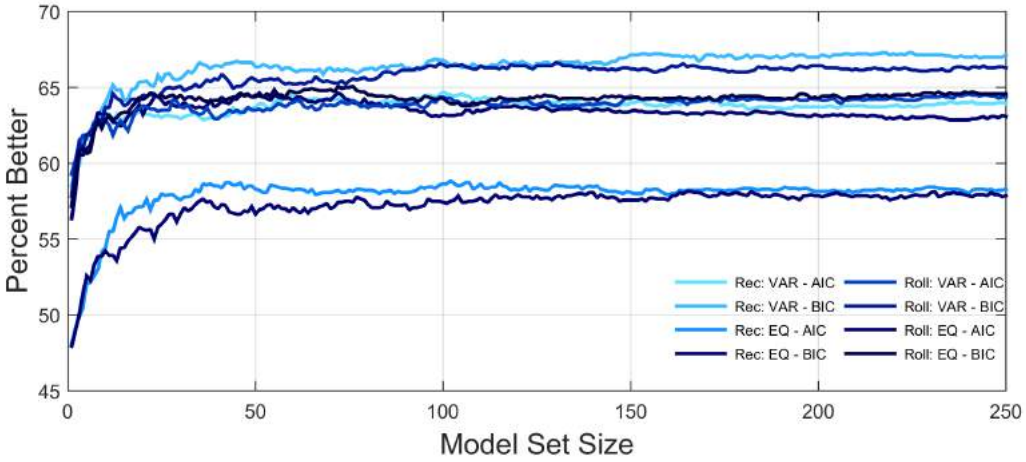
In order to validate our forecasting approach, we compare the results of each of our model sets against the eight forecasting approaches with the FAVAR(1,k), ARIMA(p,d,q) and mean benchmarks discussed above. The results, seen in Figures 2.11 and 2.12, show the percentage of times a forecast for each of the model sets is lower than the corresponding benchmark and how the average ratio for three model sets ($M^{1}_{s,t+j}$, $M^{125}_{s,t+j}$, $M^{250}_{s,t+j}$) changes across recursion. Commenting solely on the frequency of the out-performance, we can see that for datasets of this magnitude, a FAVAR(1,k) model performs extremely poorly on average, and is beaten approximately 80% of the time across each state and recursion, and the average forecast error is almost double that of our model selection and weighting routines. The mean model performs slightly better, and, depending on the modelling strategy it is competing against, is bested in approximately 55-65% of instances. The most competitive of the three benchmarks is the ARIMA selection method akin to Hyndman and Khandakar (2008), which is actually able to outperform the worst performing combination of our modelling strategies over 50% of the time for all model set sizes. However, our most competitive mode modelling strategy - which we claim is the IMS forecasting method with models selected and weighted by the BIC - is still able to beat outperform the Hyndman and Khandakar (2008) style algorithm in around 55% of all instances. As in Figure 2.9, in Figure 2.11 we can see a gradual increase in out-performance as the model set size increases, validating our former suggestion of a set size in the region of 30-50 in order to increase performance but ensure computational tractability. Figure 2.12

Figure 2.11: How frequently can models beat the benchmark?

2.11.1 Beating the FAVAR



2.11.2 Beating the mean model



2.11.3 Beating Hyndmann and Khandakar (2008) algorithm

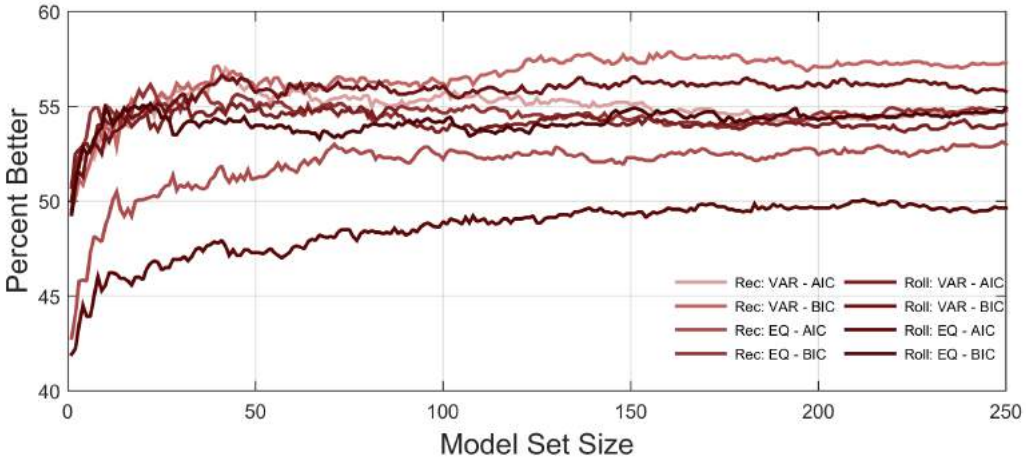
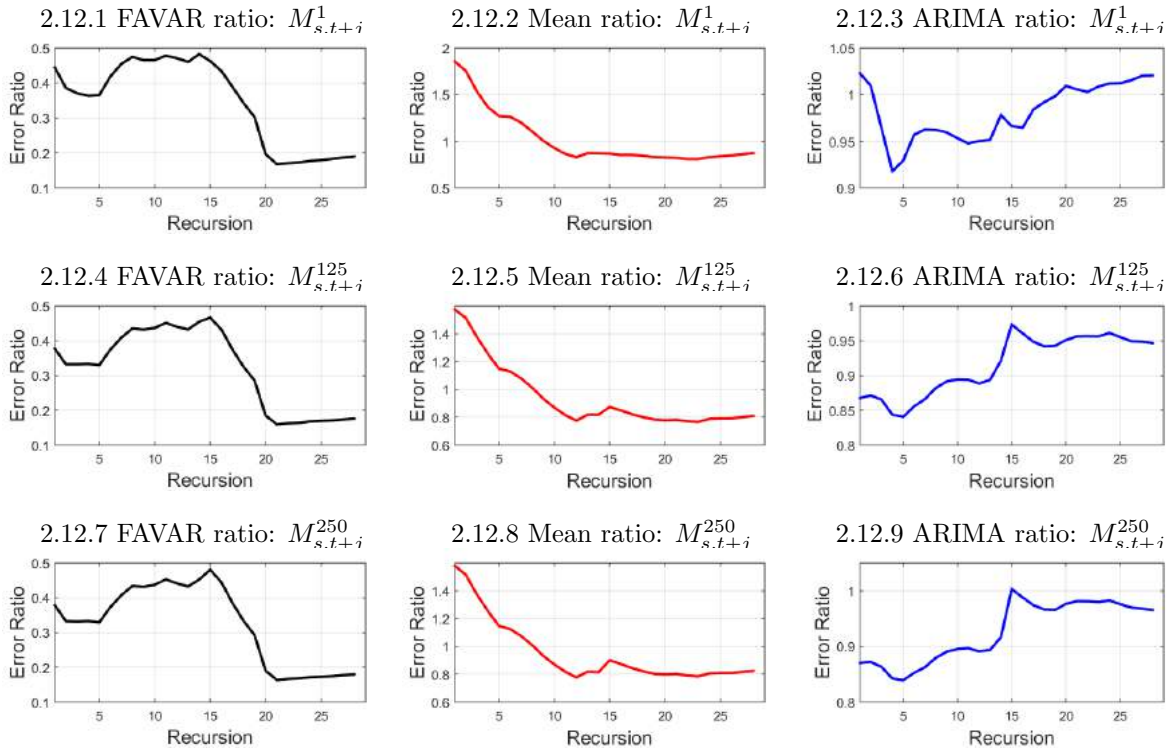


Figure 2.12: Benchmark ratios: IMS forecasts, BIC selection and recursive estimation



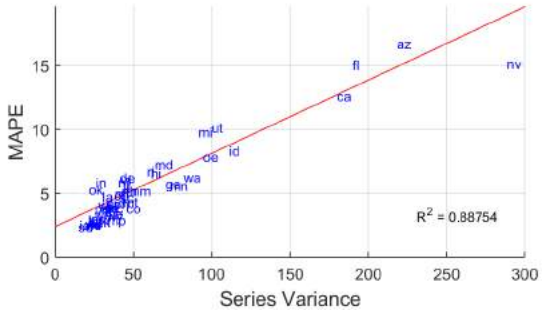
displays the fact that for two of the benchmarks, forecast error ratios are considerably lower both during, and after, the onset of the Great Recession. This is indicated by a metric common to the macroeconometric forecasting literature: where the forecast error ratio is less than one. However, the 'mean' benchmark performs admirably in times of stability: recursions one through ten.

2.4.7 Forecast error and variance

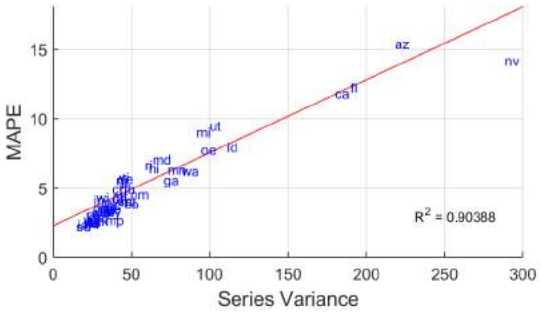
The final metric which we consider is how the mean absolute forecast error (across recursion, per horizon) varies with the variance of the underlying series which is being forecast (the annualized log real growth rate), with results seen in Figure 2.13. We report the results from the largest model set only ($M^{250}_{s,t+j}$), although all other results are of course available on request. Each element on each panel represents the average mean absolute percentage error (as per Equation 2.10) for each state averaged over all recursions plotted against the applicable state-level variance, with each panel representing a different forecasting approach. The result is consistent across all eight approaches, and is completely in line with econometric theory, in that the more volatile

Figure 2.13: Forecast variance

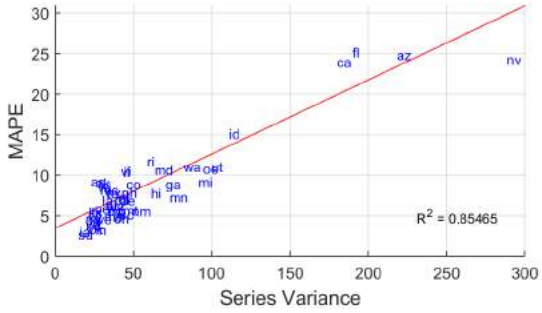
2.13.1 Recursive: IMS (AIC)



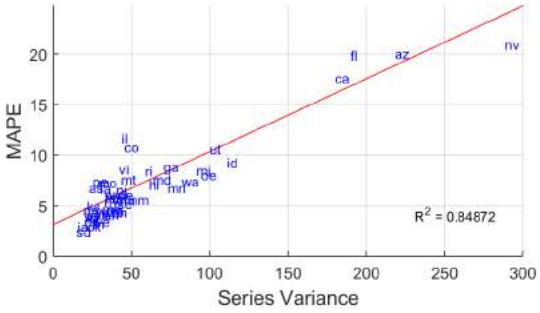
2.13.2 Recursive: IMS (BIC)



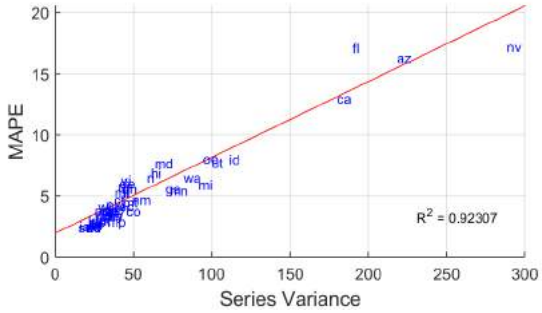
2.13.3 Rolling: IMS (AIC)



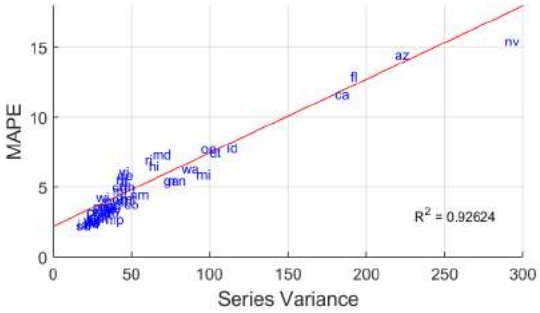
2.13.4 Rolling: IMS (BIC)



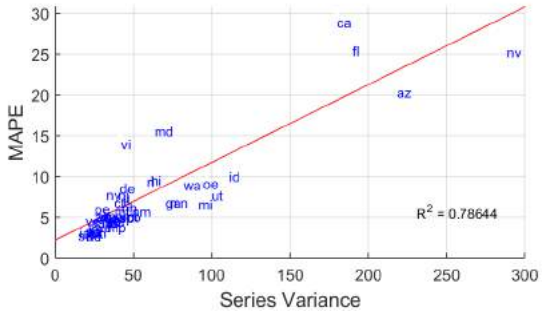
2.13.5 Recursive: DMS (AIC)



2.13.6 Recursive: DMS (BIC)



2.13.7 Rolling: DMS (AIC)



2.13.8 Rolling: DMS (BIC)

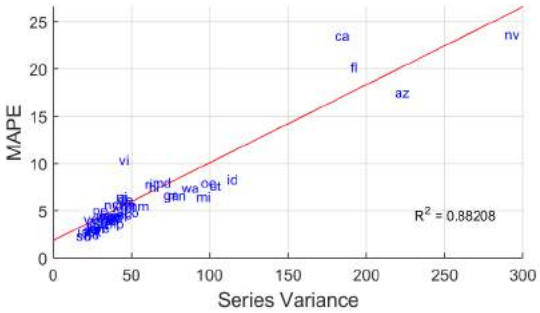
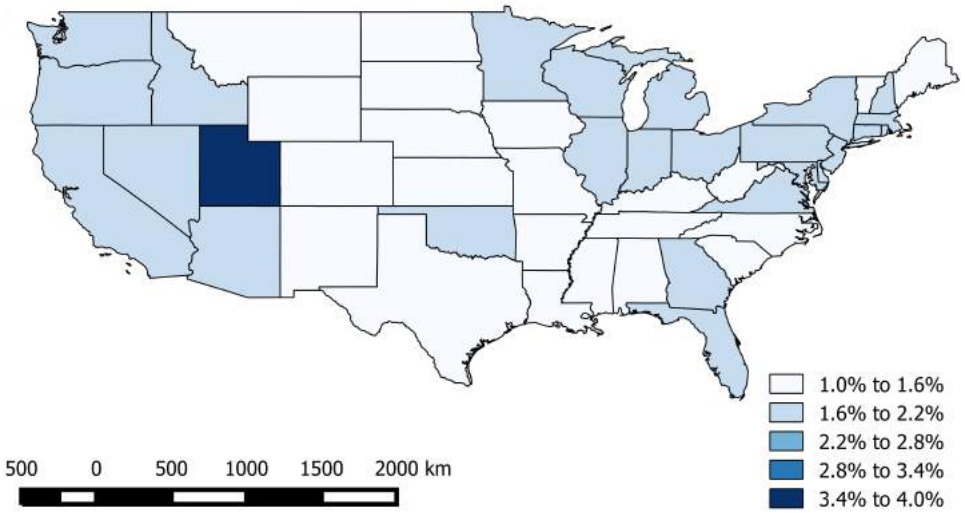


Figure 2.14: Average MAPE to standard deviation ratio



series are more difficult to forecast using autoregression style models. The correlation between forecast error and variance is 0.89, 0.90, 0.85, 0.85, 0.92, 0.93, 0.79 and 0.88 . While none are as quite as high as the 0.97 reported in Bork and Møller (2015), the results are almost equivalent, with higher correlations occurring with the DMS forecast methods. The choropleth map in Figure 2.14 shows the average MAPE (of the recursive, IMS forecast from BIC weights) to series standard deviation ratio at state level, and emphasizes a result that is completely in accordance with Rapach and Strauss (2009), that there are 'important differences in the forecastability of real housing price growth across US states, especially between interior and coastal states'.

2.5 Conclusion

This chapter has expanded on earlier work in the literature which attempts to construct and evaluate forecasts from large 'macro-panel' style datasets. In particular, we outline a method of filtering and then incorporating the huge cross-section of macroeconomic variables found in the FRED database to generate a large number of forecasts in an application to state-level house prices. Our evaluation routines show several stylized facts. Most notably, it is found that forecasts generated from IMS models are able to generate lower forecast errors more regularly than their DMS counterparts and that the approximate BMA weighting method is able to

provide lower forecast errors for such models than the SAIC method. The dominance of recursive, rather than rolling schemes when the rolling window is short, is most likely due to our choice of window size and the in-sample period (Great Moderation). The benefits of increasing the model set size for weighting peaks at between 30-50 models, not only as regards computational efficiency, but also due to the potential to include a variable subject to outlying values which may distort performance. We are able to show the stability of our forecast errors; a quality which diminishes at the peak of the financial crisis, as expected. The volatility of the underlying series is shown to be significantly correlated with the forecast error as predicted by econometric theory and recent literature, and finally, we can show the ability of our combinatorial approach to perform admirably when comparing average eight-step forecast errors against three prevalent benchmarks. For this reason, we feel that the aforementioned results encourage the incorporation of such algorithms into the modelling approach at policy institutions such as central banks, especially as processing power and data availability increases over time.

Given the limited amount of work in the area of datasets of this size (with most macro-panel datasets in this sub-discipline featuring between 100-300 cross-sections only), we propose a number of avenues for future exploration. Natural extensions as the time dimension of macro-databases (such as FRED/ALFRED) grows would be to consider different rolling window sizes, rather than a fixed 10 year in-sample estimation period as done in this chapter, or to compare the results against another large international dataset such as the Bank of England Statistical Interactive Database. Comparison of the two main model averaging techniques (SAIC and BMA) used in this chapter against a wider range of model selection or model averaging methods (such as the Mallows Model Averaging, mean-squared error ranks, or jack-knife averaging) would also undoubtedly be an extremely rich avenue for future research. Lastly, future research might explicitly compare techniques such as the above, which lie firmly within the econometrics literature, with machine learning ensemble meta-algorithms such as bagging or boosting.

PREFACE TO CHAPTER 3

In comparison to the other chapters in this thesis, here we instead consider micro ('unit level') transactions. The focus is therefore no longer specifically on out-of-sample forecasting, automation or even specifically on model selection (although all three feature heavily). Specifically, we consider the choices related to the recently emerging spatio-temporal weighting matrix specification. Assuming the weighting matrix in spatial error models is set exogenously, we use simulated data and two freely available real estate datasets well studied in the literature to show how key statistics and coefficients vary over the choice set. In particular, we focus on the effect of increasing the time and distance thresholds which determine how many neighbouring properties are taken into account, and the magnitude of weight applied to nearer neighbours (in space and time). Our results show significant variation across three-dimensional correlograms, spatial Hausman tests, Moran's I, the spatial autocorrelation parameter, a number of coefficients of interest and two commonly considered evaluation metrics. We also evaluate competing methods of matrix construction. All code is vectorized and ran in parallel where appropriate, reducing computational time from months to a matter of hours. However, it remains infeasible to run Section 3.4 with the entire dataset on a local desktop machine due to the inversion of matrices with approximately 650,000,000 elements. Forthcoming extensions include running this on a high performance computer, such as the Birmingham Environment for Academic Research (BEAR) cluster, with the full dataset and smaller convergence criteria.

CHAPTER 3

WAIT BEFORE YOU WEIGHT: A COMPUTATIONAL APPROACH TO SPATIO-TEMPORAL MISSPECIFICATION

3.1 Introduction

While the issue of spatial dependence has been well studied in the academic literature, an emerging body of work is showing the necessity of a joint consideration with the time dimension. Most of the existing studies which do so consider spatial panels, but the present chapter focuses on unit-level real estate modelling, where transactions are spatially located, and, historically, merely aggregated (pooled) over the time dimension of the data collection process. We provide simulated and empirical evidence on not only the importance of modelling both dimensions jointly (in existing estimation frameworks), but primarily on the hitherto almost arbitrary *a priori* construction of an essential component of such models: the spatio-temporal weighting matrix. Our analysis is split into two primary parts, using both simulated data and two of the most prominent, freely available spatial datasets originating from Lucas County, Ohio and the San Francisco Bay Area.

As early as Court (1939) in his analysis of automotive sales, followed by Lancaster (1966) and then the seminal work of Rosen (1974), hedonic techniques have increasingly been used to determine the implicit prices associated with the attributes of differentiated products. Of particular interest is Harrison and Rubinfeld (1978), which extends hedonic house price analysis to consider the demand for clean air. Malpezzi (2008) claims that hedonic modelling has been applied to every permanently inhabited region of the globe. The first notable publication in the field of spatial econometrics is almost certainly the volume of Paelinck and Klaassen (1979), which provides an outline of the field and preliminary methodologies, with several other well cited papers published in the same year, such as Hordijk (1979).

House price realizations (P) are due to a combination of factors at both the macro- and micro-levels. At the macro level, considerations such as interest rates, government policy and output productivity may contribute to demand, while at the micro level, structural (S), environmental (E) and locational (L) characteristics are all key determinants ($P = f(S, E, L)$). Meen (2001), however, notes that locational attributes are the most important determinant of all, and this helps to drive our modelling approach. Given the first law of geography, that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p.1), and the fact that one of the determining characteristics of real estate as an asset class is the

inherent spatial heterogeneity, simple hedonic techniques alone which do not take into account the explicit relationship between observations may be insufficient.[1] The adage of Miles (1994) that 'no two houses are the same' is particularly apt, with their primary distinctive characteristic being geographical position (in the form of latitude and longitude), or 'unique location-value signature'. While some hedonic models may incorporate variables such as accessibility indices and regional location dummies, information on the exact geo-spatial position is indispensable. Traditional analysis of housing markets regressed transaction prices on structural and non-structural attributes, estimating the implicit price of each individual characteristic. However, properties in close proximity to one another will often have similar structural characteristics, with clustering occurring due to size or age. A localized homogeneity in building quality is to be expected, given that neighbouring properties are likely to have been developed at the same time. In addition to this, several houses within a neighbourhood may be subject to the same common externalities. With the increased reporting of demographics by national statistical agencies at a disaggregated level, researchers are increasingly able to incorporate localized variables (such as population growth, educational attainment or unemployment) into their models.

Hedonic modeling with ordinary least squares is prolific in real estate analysis. A search on Google Scholar for "hedonic model real estate" currently yields approximately 23,000 results, and a number of papers show the benefits of augmenting the hedonic approach with a spatial component, estimating the model with maximum likelihood to achieve lower sum of squared errors and higher $\bar{R}^2$. However, that is not to say there are not alternative modeling approaches. Machine learning algorithms such as Random Forest and Support Vector regressions are increasingly applied to the problem in papers such as Antipov and Pokryshevskaya (2010) and Park and Bae (2015). In practical private sector applications, machine learning algorithms are increasingly common at popular online real estate database companies such as Zoopla, Zillow, Rightmove and HouseCanary which typically employ a multitude of different approaches and average their estimates across them based on recent performance. Despite this, hedonic and spatial models still saturate the academic literature.

Our primary concern is the fact that the values of a variable within a certain geographic

[1]Despite this, hedonic techniques have been used around the world in the creation of house price indices (including, for example, the index created by the Office of Deputy Prime Minister (ODPM) in the U.K.)

area show a distinct spatial pattern, causing a violation of the assumptions required of standard econometric tools such as hedonic regression. In their comprehensive guide to the subject, LeSage and Pace formally note: "Spatial dependence reflects a situation where values observed at one location or region... depend on the values of neighbouring observations at nearby locations" (LeSage and Pace, 2009, p.2). However, in this chapter, we are particularly concerned with how both the spatial and temporal elements are *jointly* incorporated into spatial weighting matrices. In particular, we stick closely to common specifications seen in the literature, focusing on the effect which key modeler decisions have on the estimation output.

This chapter, while extremely similar in design to Chapter 2, has different intentions. The purpose is not to estimate a large number of potential specifications, search for best model, and then conduct a comparative out-of-sample forecasting exercise. It is instead to estimate a large number of potential specifications and show how various conclusions drawn from this model space might vary. In particular, we are motivated by LeSage and Pace (2014), who attempt to show that potential mis-specification of the spatial matrix is largely inconsequential, and that to think otherwise is 'The Biggest Myth in Spatial Econometrics'. We expand on their research design by incorporating more datasets, different weighting schemes and a more computational approach across a larger set of permissible (spatio-temporal) specifications.

The ordering of the chapter is as follows. In Section 3.2, we review the most influential papers in the literature and those which guide and motivate our analysis. Following this, in Section 3.3, we describe a simulated approach to show some properties of the three-dimensional correllograms of various spatio-temporal weighting specifications. In Section 3.4, we first detail the model to be used (Section 3.4.1) before outlining the two different datasets which we employ (Section 3.4.2). We then show how key spatial statistics and coefficients of interest vary across the maximum spatial and temporal distance 'cut-offs' or 'thresholds' (which we will later term 'δ') available to the spatial econometrician (Section 3.4.3). In Section 3.4.4 we use the same model specifications and datasets, but instead focus on the movement across the set of permissible matrix element powers (which we will later term 'α') which determines how heavily more closely related observations are weighted. In Section 3.4.5 we directly compare competing methods of matrix construction, and finally, we conclude in Section 3.5.

3.2 Literature

While the literature in the field is extensive, we attempt to give an over-view of the hedonic and spatial models relevant to the chapter at hand. One prevalent approach in accounting for space and time in the literature is to use location and time dummies under the assumption that properties within the same region, or sold in the same time period, are subject to the same forces. One example is the work of Fletcher *et al.* (2004) for the East Midlands in the UK, which focuses solely on a specific sub-market, accounting for a short time horizon and spatial component with postcode sector and monthly dummies respectively. A second method of incorporating location into hedonic models is to include polynominals of the latitude and longitude (e.g. Cartesian $\{x, y\}$ co-ordinates) of specific observations, known in the geography literature as 'trend surface analysis'. For example, Jackson (1979) used census tract rents to create a continuous measure of location value across space through specifying lot size as an interactive term. More recently, Pavlov (2000) built a model with {x,y} coordinates, with their coefficients taking on various functional forms, thus allowing characteristic variables to vary over space. A development of this approach is taken in Fik *et al.* (2003), who run four different semi-log specifications; i.) an 'aspatial' model, ii.) a model with locational dummies, iii.) a model with interactive $\{x, y\}$ co-ordinates, and iv.) both $\{x, y\}$ co-ordinates and locational dummies. Using an in-sample estimation with 2,471 observations and 500 reserved for out-of sample evaluation based on the mean absolute error, the authors find their fourth specification prevails for the nine contiguous Multiple Listing Service (MLS) divisions of the city of Tucson in 1998. However, Pace *et al.* (1998) derides this method of dealing with spatial dependence, calling it 'the crudest and least satisfactory', citing things such as the multicollinearity caused by a large number of polynomials and the generation of an overly smooth surface which ignores natural or man-made barriers such as rivers or roads. With a similar emphasis on accessibility as Fik *et al.* (2003), Bruyne and Hove (2013) estimate municipality level house prices for all 589 Belgian municipalities in 2001. They focus on modelling the distance and travel time to important economic centres which offer services and employment, with a reduced form specification which includes fixed effects for provinces.

One directly relevant paper of interest is Bracke (2014), which also uses both locational

dummies and Cartesian co-ordinates for properties within a proprietary dataset from a leading real estate consultancy. Filtering down from 21,012 sales and 48,597 rental agreements, the author analyzes the determinants of the price-to-rent ratio in Central London during the 2006-2012 period. While his specific hypothesis is of less concern to this chapter, he finds location to have a significant impact on rent-price ratios at a property specific level, and that the R^2 is higher for postcode dummies (0.193) than a latitude/longitude polynomial (0.180) for an otherwise equivalent specification.

Several papers such as Pace *et al.* (1998), for example, come to the conclusion that the spatial weightings approach is best suited for real estate analysis due to the inherent structure of housing markets, claiming that distance based weighting matrices 'subsume two basic paradigms of real estate work' (p.8). Even Dubin (1998) admits, despite the author being a proponent of geographically weighted regression models, that the approach of modelling the process directly in the form of a spatial weighting matrix is 'more common' in the real estate literature. It is this primer of Dubin (1998) which forms the basis for our simulation routine in Section 3.3.[2] Dubin identifies the arbitrary 'a priori' way in which the weightings matrix is specified by researchers in this field, showing correlograms for different weighting schemes and different values of the autocorrelation parameter in a Spatial Error Model (hereafter SEM) approach. One particularly often cited empirical study is that of Can (1990), which places an early emphasis on the influence of spatial variability on the value of housing attributes and 'neighbourhood dynamics'. The paper examines two competing generalized weight matrix specifications; 'inverse distance' and 'inverse distance squared' with an additional two which are limited to a distance under 3 miles and 5 miles within a (relatively small) dataset of 577 single-family houses for the Franklin County portion of the Columbus MSA in the year 1980. This kind of curiosity regarding specification choices encourages the approaches taken within this chapter, in that not only do we consider variations in the cutoff (albeit in a far more computational way), but also the denominator power on inverted space and time distances.

Pace and Gilley (1997) take the now famous Harrison and Rubinfeld (1978) data (Bostom SMSA with 506 observations), re-specifying the model to create a more independent and identically distributed error structure. They claim that estimating the model as a SEM with a

[2]Replication codes for the most relevant part of Dubin (1998) can be found in Appendix A.

distance based spatial weighting matrix not only generates more theoretically correct coefficient signs, but also causes the estimated sum-of-squares to fall by 44% relative to OLS models. Wilhelmsson (2002) uses time series dummies in a dataset of 1,377 single family homes in Stockholm between January 2000 and May 2001. The author considers four different weighting matrices (inverse distance squared, inverse distance squared limited to within 600m, closest neighbour and contiguity), showing that within a SEM framework (which Wilhelmsson considers the most widely used in spatial real estate modelling, and is the approach taken in this chapter), the model with the inverse distance (maximum 600m) best explains the variation in price ($\bar{R}^2 = 0.689$, compared to a hedonic OLS model with $\bar{R}^2$=0.664). Wang and Ready (2005) estimate three models and consider an often overlooked modeler decision. Using 7,493 arms-length house sales between 1998 and 2002 in Pennsylvania, with 1,600 meters as the cutoff distance for spatial weighting matrices, they highlight the fact that row standardization better modeled the spatial structure for the SEM model, but within a Spatial Autoregression (hereafter SAR) context, the prices were better modeled by a non-standardized weighting matrix. Using 2,715 house prices for a single year (2002) with 61 independent variables, Ismail (2005) shows that an accessibility measure is more influential than a zonal accessibility measure because the former is better able to capture the micro effect of location on house price, showing that the adjusted R^2 increases from 0.758 with an OLS estimator up to 0.797 under the SEM specification.

In a more relevant spatio-temporal context, Can and Megbolugbe (1997) identify and utilize only (recent) 'comparable sales': properties within a pre-specified distance which sold within an arbitrarily determined 'a priori' time period. Specifically decomposing their weighting matrix into four subsidiary components based on lower triangular time (T) matrices, Pace *et al.* (1998) propose another method, termed the spatio-temporal auto regression (STAR). With a focus on sparse matrices and computational efficiency, the authors use a dataset of 70,822 observations on real estate transactions from 1969 to 1991 in Fairfax County Virginia. The authors show that their spatio-temporal autoregression with twelve variables is able to reduce the median absolute error by 37.35% in comparison to an indicator-based specification containing twenty-six variables. Taking a similar approach, Pace *et al.* (2000) more specifically consider the forecasting implications of the earlier model by some of the same authors. They restrict their spatial matrix to the 15 nearest neighbours (NN), and, critically, their lower triangular matrix T to have

a constant weight for each neighbour only (a 'naive' approach), claiming that they 'did not examine extensively weighting elements of T in a more complicated fashion' [p.236]. This, however, is one of the key contributions of our simulated and empirical approaches. Using 5,234 sales from Baton Rouge, the authors show the 'substantial benefits' of explicitly modelling the temporal dependence of the errors, claiming: 'the spatio-temporal autoregression with 14 variables produced 46.9% less SSE than a 12-variable regression using simple indicator variables for time'.

Building on this earlier work, Nappi-Choulet and Maury (2011) also incorporate a filtering process into the conventional hedonic function in an attempt to correct for spatial and temporal effects. Analyzing the inner Parisian housing market across over a million transactions (with data from the Paris Region notary office from 1990Q1 to 2005Q4), they merge the aforementioned STAR model of Pace *et al.* (1998) with a local estimation technique to create an LSTAR model. Their model, based on overlapping sub-samples, shows that spatial autoregressive effects are more pronounced in the historic centre of Paris than in the surrounding area. Sun *et al.* (2005) extend this framework and apply it to 54,282 transactions in the market for Singapore multi-unit high rise condominiums between 1990 and 1999. Their key innovation is the decomposition of the weighting matrix into W_1 and W_2, which separately identify building and neighbourhood effects respectively, and to use a Bayesian method to estimate the model to produce more robust coefficients. Smith and Wu (2009) develop a model which not only allows for the lagged spatio-temporal effects of previous sales, but also for general autocorrelation effects over time. Their contribution is the treatment of the unequal timing of their observations (sales, whereby their extension of the AR(1) model allows events (house sales) to occur on the same day), with both simulations and an empirical application of a small dataset in the Philadelphia area. An important precursor to the methodology discussed below is that of Huang *et al.* (2010), which claims to show 'substantial benefits' in comparison to pre-existing methodologies. These authors decompose the distance measure into one that is a function of both space (d^S) and time (d^T): $d^{ST} = d^S \circ d^T$, where $\circ$ could be a range of different operators. In the discussion of the work which follows, $\circ$ is taken to represent the Hadamard product.

Building on the formulations of Pace *et al.* (2000), additional work which inspires our approach to matrix construction in Sections 3.3 and 3.4 is that of Jean Dubé and Diègo Legros, and

their simulated and empirical work which makes use of city centre datasets for Paris, Quebec and Lucas County Ohio (the final of which is identical to the one we utilize in Section 3.4.3). In work such as Dubé and Legros (2013b), they provide a generalization of inverse distance spatial weighting matrices (proposed in the aforementioned papers) into matrices which respect the multi-dimensional spatial, and uni-dimensional temporal relation. The matrix they propose is essentially a constrained version of the spatial weights matrix typically seen in the earlier literature, whereby specific elements are set equal to zero according to when transactions take place, or weighted according to how recent the transactions were. In the paper Dubé and Legros (2014b), the authors estimate several specifications dependent on the distance threshold criterion (δ^S, discussed extensively below), which simultaneously motivates and informs this chapter. Their conclusions are backed up with Monte Carlo evidence in Dubé and Legros (2015), with a general theme showing that their spatio-temporal model outperforms a spatial autoregressive specification (and simultaneously the OLS specification) and that a simple spatial specification which ignores the temporal dimension significantly over-estimates the autoregressive coefficient values, influencing the magnitude of the β estimates.

While not the same approach as the ones taken above, it is necessary to consider recent work which treats spatial weighting matrices as endogenous. Of particular interest is Qu and Lee (2015), which presents three estimators of the SAR model with this new assumption: a two-stage instrumental variable method, a quasi-maximum likelihood approach, and a general method of moments estimator. Bhattacharjee and Jensen-Butler (2013) show that in the same SEM context as in this chapter (with a slight reformulation), the spatial weights matrix is only partially identified, but fully identified under the constraint of symmetry. They propose a new method of estimating the weighting matrix, which is then applied to a spatial panel for UK housing markets. However, much work still remains to be done on the theory of weights construction, especially with regard to the time dimension. While a large number of approaches are taken in the literature, there appears to be no consensus, or no significant discussion of the sensitivity of the modelling choices involved other than a cursory consideration of a couple of competing specifications. This further motivates our study of not only the optimal distance threshold parameter, but also the optimal threshold parameter relating to time (the upper bound traditionally fixed by modelers), recently introduced in the aforementioned literature, as well as

the mass of weight attributed to closer observations, determined by the exponents on the inverse time and distance elements.

3.3 Simulated analysis based on Dubin (1998)

In this section we follow Dubin (1998) and simulate a large number of houses located on a grid, advancing the previous design by accounting for the time dimension which we argue is of critical importance in applications to house price modelling. We consider three spatio-temporal weighting matrix specifications which featured in Section 3.2. Our continuous (inverse distance), as opposed to binary (contiguity based) approach to the weighting matrix is also motivated by some of the founding literature in the field, where it is claimed that "Ease of computation apart, there is no virtue in using binary weights" (Cliff and Ord, 1969, p.50). Therefore, we use the SEM and specify the equation for the error terms using standard notation:

$$y = X\beta + u \tag{3.1}$$

$$u = \lambda W u + \varepsilon \tag{3.2}$$

Using the same notation that we will further develop in Section 3.4.3, let y represent an unspecified $N \times 1$ vector containing the sales price of various houses, X represent an unspecified $N \times K$ matrix containing various attributes of houses to be discussed more in Section 3.4.3 and β to be an unspecified $K \times 1$ vector of coefficients relating to each attribute. The $N \times 1$ vector ε refers to uncorrelated residual terms which are normally distributed with variance σ^2 and our autocorrelation parameter is λ, which equals zero if there is no spatial correlation. From Equation 3.2, as in Dubin (1998), we can show that the variance-covariance matrix is defined as:

$$u = (I - W)^{-1}\varepsilon \tag{3.3}$$

$$V = \sigma^2 (I - \lambda W)^{-1}(I - \lambda W')^{-1} \tag{3.4}$$

From this matrix (V), we calculate and plot three-dimensional correlograms as:

$$Corr_{i,j} = \frac{V_{i,j}}{\sqrt{V_{i,i} \times V_{j,j}}} \tag{3.5}$$

We begin by simulating the data; we simulate 100 (a number chosen to prevent 'over-crowding' of results in Figures 3.1-3.3) randomly distributed points (representative of houses) on a square grid of dimension 1000, 1000 (chosen to represent the typical size of a city centre with meters as units of measurement). We randomly assign simulated points in time across which the houses were sold (60 periods, which is representative of the five years of monthly data typically used in studies such as those discussed in Section 3.2, and approximately equal to one of the datasets discussed in Section 3.4.2). Combined, these two attributes of each observation resemble housing units distributed across a city sold at random intervals in time.

While the debate about which spatial model is most suitable under certain assumptions about the data generating process (see, for example LeSage and Pace (2009)) and whether the SAR or SEM is more appropriate in the house price context is ongoing, we stick to the SEM employed by Dubin (1998) to enunciate the effect of the addition of a dimension to the spatial weight matrix and to focus upon modeler choices. What is of critical importance, however, is the specification of the $N \times N$ weighting matrix W, where $W_{i,j}$ represents the relationship between observations i and j (where, in our simulated example, $N = 100$). Dubin (1998) examines the issue using n nearest neighbours, the case where the maximum distance between two house price sales (i and j) is less than a specific distance D, and an inverse distance criteria ($W_{i,j} = \frac{1}{D^{\alpha}_{i,j}}$), all with zeros on the main diagonal.[3] In our approach throughout this chapter, we consider three distinct weighting matrices and two key critical modeler choices. In each approach, the first step is to calculate our $N \times N$ distance matrix D from our two $N \times 1$ Latitude (Lat) and Longitude (Lon) vectors, where each element is calculated as the distance between the two points ($d_{i,j} = \sqrt{(Lat_i - Lat_j)^2 - (Lon_i - Lon_j)^2}$). We take a similar approach with our time matrix (T), generating $t_{i,j}$ from our $N \times 1$ vector of simulated dates ($Date$) as $t_{i,j} = Date_i - Date_j$. Our time ($T$) and distance ($D$) matrices are then combined in three different ways to create three final matrices ($W_A^{\alpha,\delta}$, $W_B^{\alpha,\delta}$ and $W_C^{\alpha,\delta}$) to be discussed below. These matrices are fed into Equations 3.1 - 3.5. Our D and T matrices are constructed as:

[3] As noted by Dubin (1998), each of these specifications has featured frequently in the real estate literature.

$$D = \begin{bmatrix} 0 & d_{1,2} & \dots & d_{1,j} \\ d_{1,2} & 0 & \dots & d_{2,j} \\ \vdots & \dots & \ddots & \vdots \\ d_{1,j} & d_{2,j} & \dots & 0 \end{bmatrix} \qquad T = \begin{bmatrix} 0 & t_{1,2} & \dots & t_{1,j} \\ -t_{1,2} & 0 & \dots & t_{2,j} \\ \vdots & \dots & \ddots & \vdots \\ -t_{1,j} & -t_{2,j} & \dots & 0 \end{bmatrix} \tag{3.6}$$

By design, D is symmetric, and T is symmetric with signs reversed, as shown in Equation 3.6. The first group of matrices ($W_A^{\alpha,\delta}$) is created as per Dubé and Legros (2013b) whereby element $w_{i,j,A}^{\alpha,\delta}$ is defined as:

$$w_{i,j,A}^{\alpha,\delta} = \begin{cases} \frac{1}{t_{i,j}^{\alpha}} \times \frac{1}{d_{i,j}^{\alpha}} & \text{if } 0 < t_{i,j} < \bar{t} \text{ and } d_{i,j} < \bar{d} \text{ and } i \neq j \\ 1 \times \frac{1}{d_{i,j}^{\alpha}}, & \text{if } 0 = t_{i,j} \text{ and } d_{i,j} < \bar{d} \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

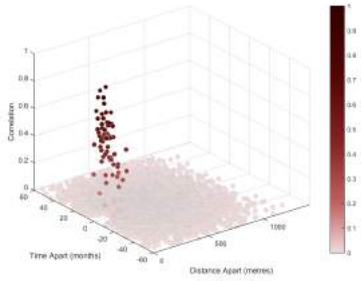
To consider the effect of modeler choice, we vary $\alpha = \{1, 3, 5\}$, explicitly considering smaller intervals across α-space in Section 3.4.4. The intuition behind this choice corresponds to the relative weight placed on more closely located observations (in time and space). The second matrix, denoted as $W_C^{\alpha,\delta}$, applies the same inverse *distance* weights as used above and in more classical spatial weighting specifications, merely restricted to observations which have already occurred through the natural passage of time. What is of critical importance to re-emphasize at this point is that if one observation happens before another ($t_{i,j} < 0$), then the element $w_{i,j,A}$ and $w_{i,j,B}$ is set to zero, highlighting the uni-directional time dimension under examination:

$$w_{i,j,B}^{\alpha,\delta} = \begin{cases} \frac{1}{d_{i,j}^{\alpha}} & \text{if } 0 < t_{i,j} < \bar{t} \text{ and } d_{i,j} < \bar{d} \text{ and } i \neq j \\ \frac{1}{d_{i,j}^{\alpha}} & \text{if } 0 = t_{i,j} \text{ and } d_{i,j} < \bar{d} \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

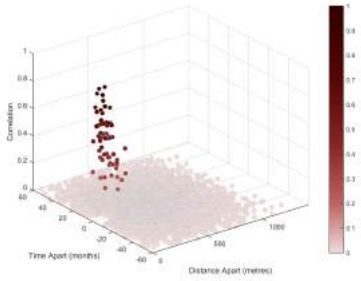
In both specifications, observations at period $t_{i,j} = 0$ might be thought of as 'under offer' or 'subject to completion'. The third of the three methods can be expressed as $W_C^{\alpha,\delta}$ (with the subscript referring to the more classical *spatial*, as opposed to *spatio-temporal* weighting matrix) where each element is constructed as:

Figure 3.1: $W_A^{\alpha,\delta}$ Specifications

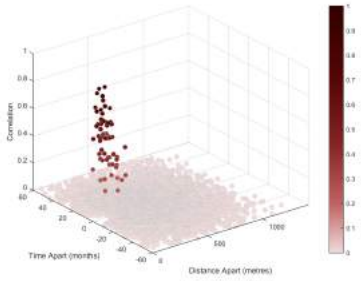
3.1.1 $\alpha = 1$ and $\delta = 0.1$



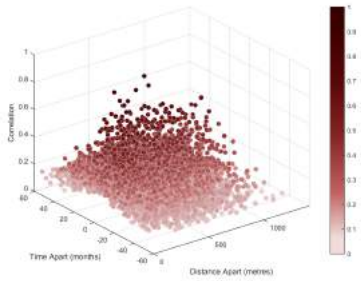
3.1.2 $\alpha = 3$ and $\delta = 0.1$



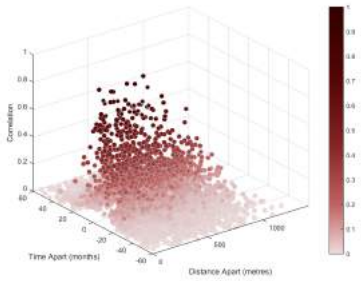
3.1.3 $\alpha = 5$ and $\delta = 0.1$



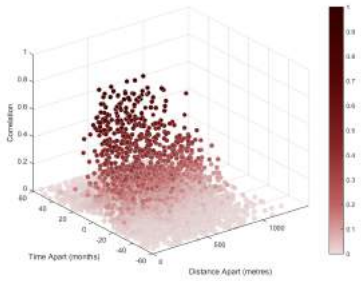
3.1.4 $\alpha = 1$ and $\delta = 0.5$



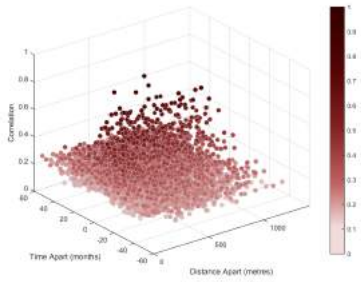
3.1.5 $\alpha = 3$ and $\delta = 0.5$



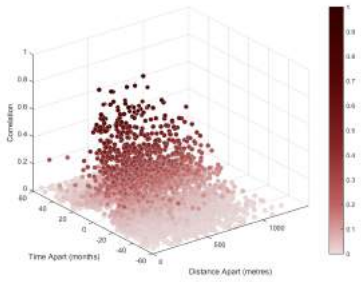
3.1.6 $\alpha = 5$ and $\delta = 0.5$



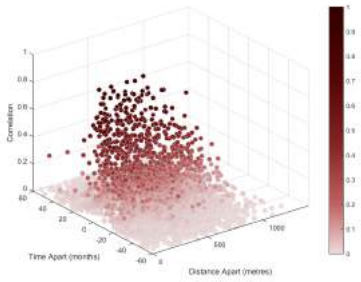
3.1.7 $\alpha = 1$ and $\delta = 0.9$



3.1.8 $\alpha = 3$ and $\delta = 0.9$



3.1.9 $\alpha = 5$ and $\delta = 0.9$



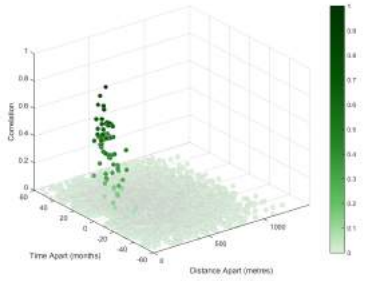
$$w_{i,j,C}^{\alpha,\delta} = \begin{cases} \frac{1}{d_{i,j}^{\alpha}} & \text{if } d_{i,j} < \bar{d} \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{3.9}$$

Each of these matrices is then row standardized. A useful observation is that the weighting schemes are increasingly general between $W_A^{\alpha,\delta}$, $W_B^{\alpha,\delta}$ and $W_C^{\alpha,\delta}$ and that they are nested within one another. For fixed α and δ, it is trivial to show that $W_A^{\alpha,\delta}$ is a nested within $W_B^{\alpha,\delta}$ and that they are equivilent when $t_{i,j}^{\alpha} = 1$. We can then further show that $W_B^{\alpha,\delta}$ is nested within $W_C^{\alpha,\delta}$ and that they are equivalent when $\bar{t} > t_{i,j}, \forall i, j$.

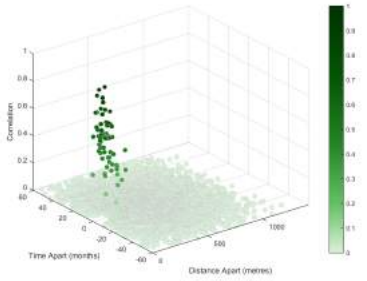
As alluded to above, and as seen in Figures 3.1-3.3, we not only vary α, but also a modeler choice which we will define as δ. This parameter is defined as the fraction of the maximum

Figure 3.2: $W_B^{\alpha,\delta}$ Specifications

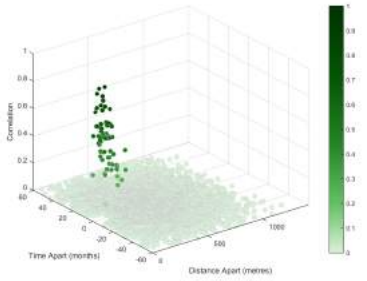
3.2.1 $\alpha = 1$ and $\delta = 0.1$



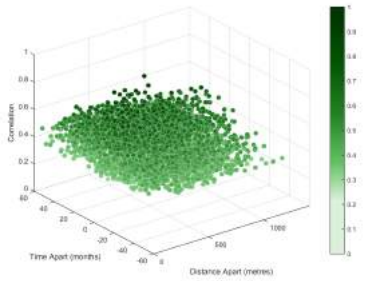
3.2.2 $\alpha = 3$ and $\delta = 0.1$



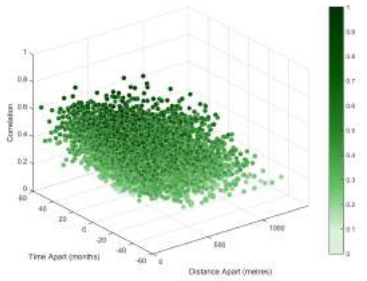
3.2.3 $\alpha = 5$ and $\delta = 0.1$



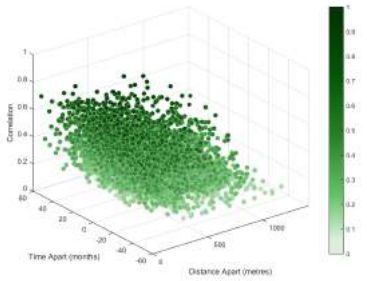
3.2.4 $\alpha = 1$ and $\delta = 0.5$



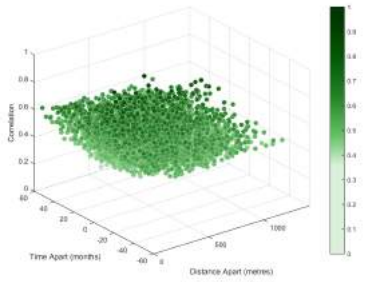
3.2.5 $\alpha = 3$ and $\delta = 0.5$



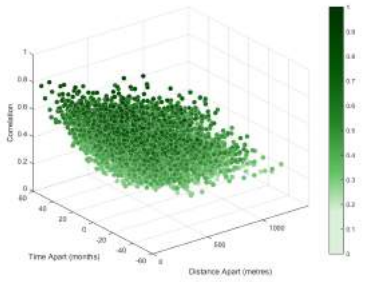
3.2.6 $\alpha = 5$ and $\delta = 0.5$



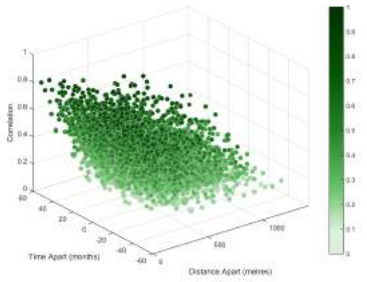
3.2.7 $\alpha = 1$ and $\delta = 0.9$



3.2.8 $\alpha = 3$ and $\delta = 0.9$



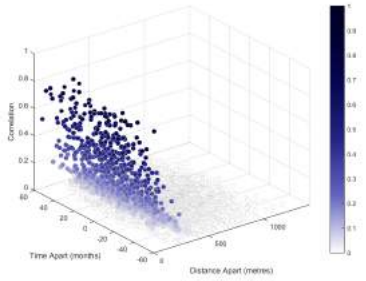
3.2.9 $\alpha = 5$ and $\delta = 0.9$



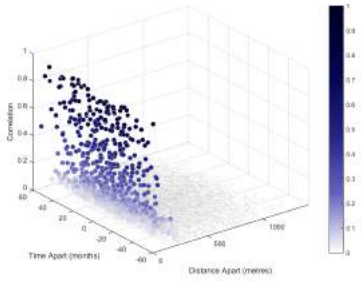
distance along either time or space: for $\bar{t}$ and $\bar{d}$ where $\bar{t} = \delta \times \max_{i\in\{1,\ldots,N\},j\in\{1,\ldots,N\}}\{t_{i,j}\}$ and $\bar{d} = \delta \times \max_{i\in\{1,\ldots,N\},j\in\{1,\ldots,N\}}\{d_{i,j}\}$, we calculate outputs respectively for each of $\delta = \{0.1, 0.5, 0.9\}$. The variations of $\alpha = \{1, 3, 5\}$ and $\delta = \{0.1, 0.5, 0.9\}$ in this section correspond directly to those examined (albeit with a more computational approach) in Section 3.4. The intention in this chapter is to consider the effect of exogenously predetermined values for δ and α. In the empirical application in Section 3.4.3, we decompose δ into movements across both the time (δ^T) and space (δ^S) dimensions separately. The intuition behind these scalars is that they represent the maximum values permissible (a 'threshold') for non-zero values in matrices $W_A^{\alpha,\delta}$, $W_B^{\alpha,\delta}$ and $W_C^{\alpha,\delta}$, determined by the longest distance between sales in time and space respectively as a relation to the maximum space and time distance in the entire sample. A larger δ^S means

Figure 3.3: $W_C^{\alpha,\delta}$ Specifications

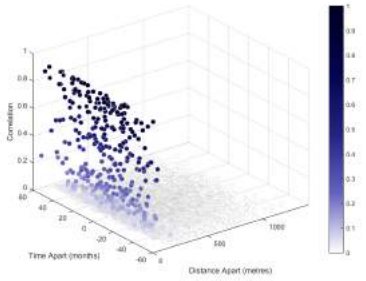
3.3.1 $\alpha = 1$ and $\delta = 0.1$



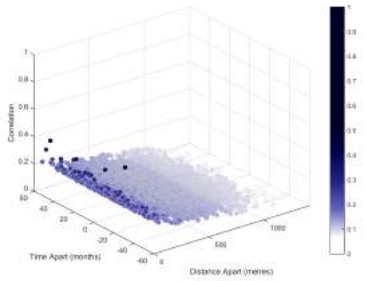
3.3.2 $\alpha = 3$ and $\delta = 0.1$



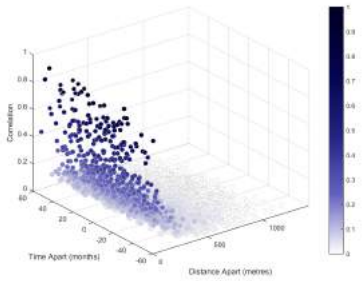
3.3.3 $\alpha = 5$ and $\delta = 0.1$



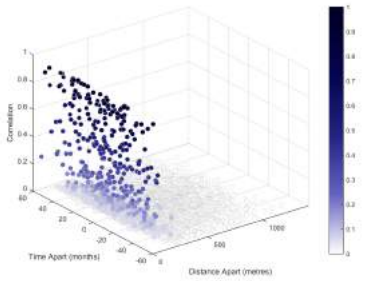
3.3.4 $\alpha = 1$ and $\delta = 0.5$



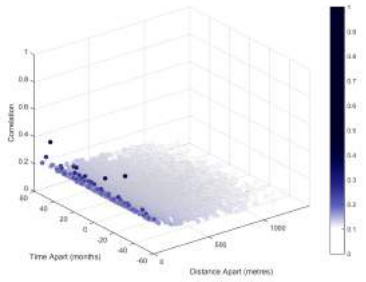
3.3.5 $\alpha = 3$ and $\delta = 0.5$



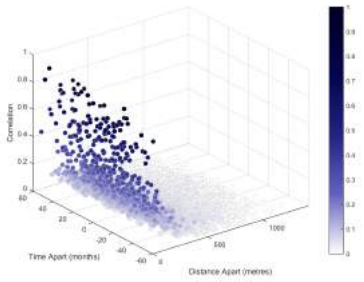
3.3.6 $\alpha = 5$ and $\delta = 0.5$



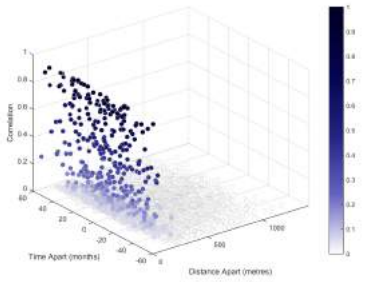
3.3.7 $\alpha = 1$ and $\delta = 0.9$



3.3.8 $\alpha = 3$ and $\delta = 0.9$



3.3.9 $\alpha = 5$ and $\delta = 0.9$



more information on observations located further away in space is included in the row for one specific observation, and a larger δ^T incorporates the inclusion of information on observations which occurred a longer time ago. The $W_C^{\alpha,\delta}$ ignores the entire temporal dimension. Variation over α, as in Section 3.4.4, has the effect of weighting nearer, more closely related observations (in time and space - $W_A^{\alpha,\delta}$ and $W_B^{\alpha,\delta}$ - or in space alone - $W_C^{\alpha,\delta}$) higher than those further away.

As in Dubin (1998), our approach is to plot the correlograms which are now also a function of the time separating the sales. We choose an arbitrary value of $\lambda = 0.75$, corresponding to between a medium and high level of spatial dependence according to Pace and LeSage (2008), and is a value firmly situated in the range of estimates in Section 3.4.3. Figure 3.1 displays the output for each combination of α and δ for the $W_A^{\alpha,\delta}$ specification, Figure 3.2 displays the

nine equivalent matrices for $W_B^{\alpha,\delta}$, and Figure 3.3 displays $W_C^{\alpha,\delta}$. Of critical importance is to consider how correlations change with separation distance in time and space and the relative weights between them, and that the relative choice of the parameters (δ and α) greatly affects the correlation pattern. As can be seen by the three sets of subplots, the 'peaked' nature of the points decreases as the boundary within which the weights are non-zero increases, as expected. The two spatio-temporal $W_A^{\alpha,\delta}$ specifications show, also as expected, a peaked distribution in time and space, as opposed to the more 'vertical' pattern seen in $W_C^{\alpha,\delta}$ specifications. This is because $W_C^{\alpha,\delta}$ omits the temporal consideration, creating no variation in the correlogram on the time (y-axis) dimension. More peaked patterns are seen in higher α values as higher relative weights are placed on those observations located more closely together. Considering the overall variation across the three figures, an important point to remember is that none of these weighting schemes are by any means implausible, yet the correlations between the observations can differ dramatically, especially when considering, for example, the top right (high α, low δ) and bottom left (low α, high δ) panels in each figure.

3.4 Empirical applications

We pursue a similar hypothesis here as in Section 3.3, albeit with a more computational, applied approach. As mentioned, in Section 3.4.3 we decompose δ into δ^S and δ^T in order to consider how key conventional spatial statistics and methods of model evaluation vary over the specification set, and in Section 3.4.4 we focus on the same criteria while creating variation in α. We split these analyses into these two primary routes of inquiry, examining the three weighting matrix specifications outlined and employed in Section 3.3, and we utilize two prominent, publicly available datasets which we discuss and describe in Section 3.4.2.

3.4.1 The SEM model

Our approach is to build a continuum of weighting matrices specified as per Equations 3.5-3.7. We adapt the notation of Equations 3.1 and 3.2 for our empirical application:

$$y_d = X_d \beta_{d,m}^{T,S,\alpha} + u_{d,m}^{T,S,\alpha} \tag{3.10}$$

$$u_{d,m}^{T,S,\alpha} = \lambda_{d,m}^{T,S,\alpha} W_{d,m}^{T,S,\alpha} u_{d,m}^{T,S,\alpha} + \varepsilon_{d,m}^{T,S,\alpha} \tag{3.11}$$

where superscripts ^{T,S} denote estimations made from the various combinations of δ^T and δ^S respectively and α denotes variation over the ratio of weights between observations. The subscript $_d$ indicates that we are estimating over two different datasets, and m denotes each of the three methods of matrix construction discussed in Section 3.3 ($m = \{A, B, C\}$). The log of each retail sale price of each observation is expressed in elements of the $N \times 1$ vector y_d, where N is the number of observations in each dataset. Spatial dependence of unit level house prices on nearby units (and contemporaneous/recent sales) is measured by the scalar coefficient $\lambda_{d,m}^{T,S,\alpha}$, the variation of which across the grid is explicitly analyzed in the following sections. Columns of X_d include variables such as a constant, data on the age, the log of the living area, the number of bedrooms, the log of the house lotsize and additional dataset specific variables. Our goal is to analyze how results change across variations in the 20×20 grid in δ and the 1×400 grid in α space. It should be noted that in the chapter, we are not using a spatial autocorrelation model (more commonly called SAR – a spatial autoregressive model), but a spatial error model (SEM). However, a common factor restriction between these two models is implied under a specific condition, as discussed in Burridge (1981): Writing a SAR model in matrix notation as: $(I - \beta W)Y = (\gamma_0 I + \gamma_1 T)X + \varepsilon$, Burridge shows that when $T = W$, and $\gamma_1 + \beta\gamma_0 = 0$, a common factor can be extracted, simplified as: $Y = \gamma_0 X + u$ where $u = \beta W u + \varepsilon$ - the SEM used in this chapter.

3.4.2 The datasets

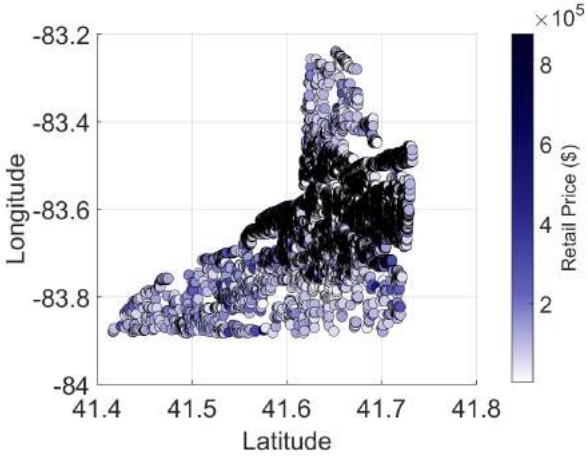
3.4.2.1 The Ohio dataset

The first dataset which we detail is the somewhat infamous housing sales dataset of Lucas County Ohio, distributed as part of the library of datasets associated with James LeSage's spatial econometrics toolbox which is available at http://www.spatial-econometrics.com.

There are approximately 140,000 properties in the area, with between 4,000 and 8,000 sold per year on average. Specifically, the dataset is for single-family residential sales between 1993-1998, first used in the study by LeSage and Pace (2004). Provided by the county auditor, there are 24 columns of information including, but not limited to selling price, latitude, longitude, sale date, and various structural characteristics such as year built and total living area. The dataset provides 25,357 observations, but due to computational restrictions, and a desire to more accurately match sample sizes commonly seen in the literature, we create a sample of $10,000$ observations drawn randomly without replacement from the dataset. This practice of estimating models on subsets of a large dataset is by no means uncommon, and in Dubé and Legros (2014a), the author's estimations use only 10% of their Quebec database, and generate

Figure 3.4: Spatial distribution of the datasets

3.4.1 Ohio dataset



3.4.2 San Francisco dataset

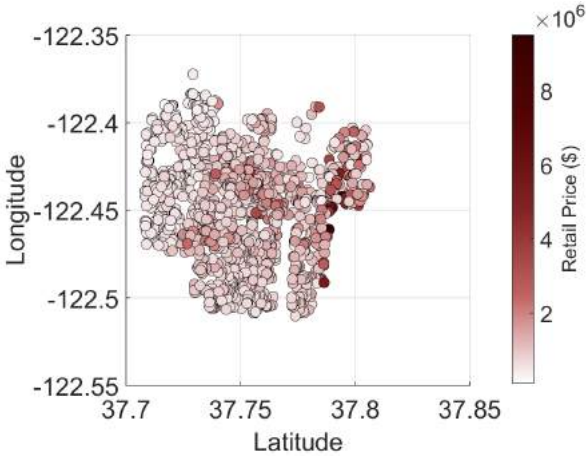
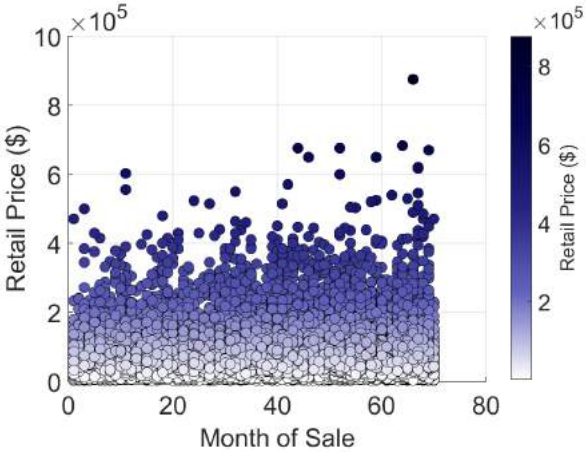


Figure 3.5: Temporal distribution of the datasets

3.5.1 Ohio dataset



3.5.2 San Francisco dataset

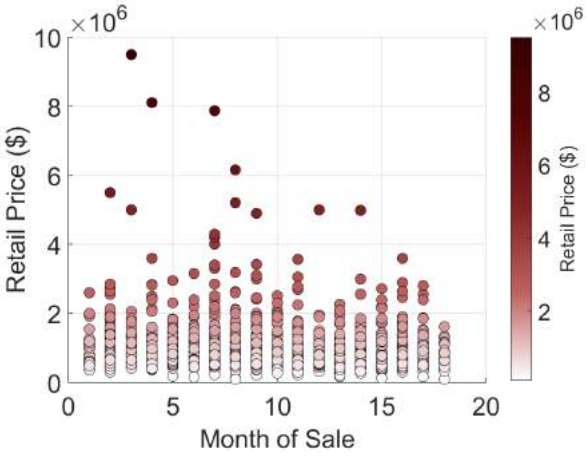


Figure 3.6: The four main variables: San Francisco

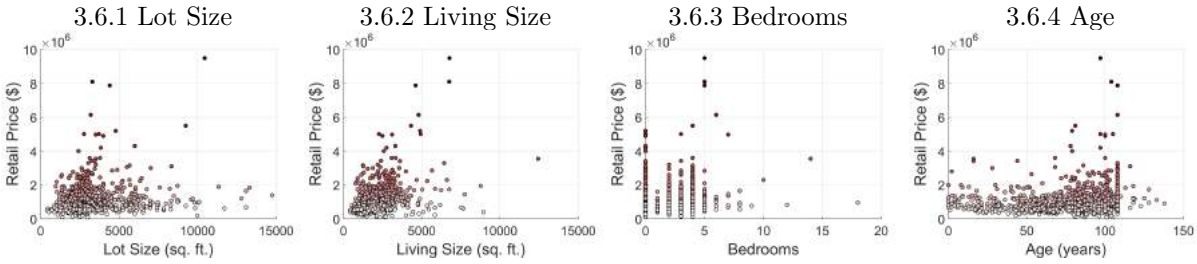
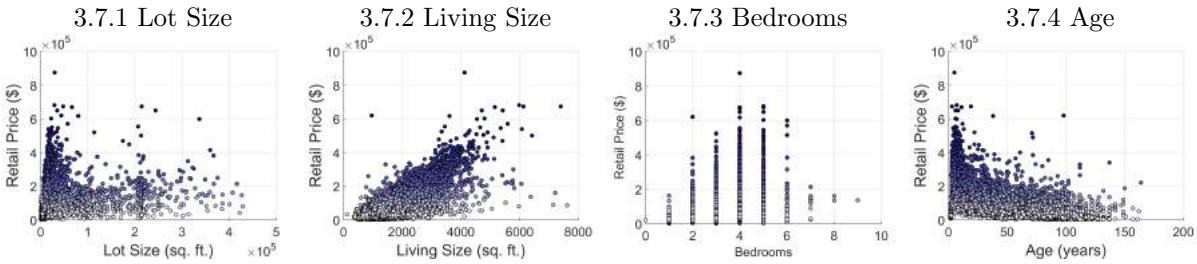


Figure 3.7: The four main variables: Ohio



three samples of 10,000 observations each for their Paris dataset. Our sample is somewhat the result of computational necessity in the way that we approximate the log determinant (using the MC method of Barry and Pace (1999)) and the fact that we are estimating 400 models per dataset and method.[4] The maximum distance between any two properties in the entire sample is measured at 59.31 kilometers, and sales occur across 70 months between January 1993 and October 1998.

3.4.2.2 The San Francisco dataset

The second dataset which we utilize is a dataset on the San Francisco Bay area compiled by Joseph Adler and described in Adler (2010). The original dataset contains details on 3,280 sales between February 13, 2008 and July 6, 2009. While the exact date of sale is provided, we use months as the time dimension. Data is distributed across thirteen months, and the largest distance between any two observations is 9.59km. The dataset contains data on the sold price, lotsize, interior square feet, the number of bedrooms, the year built, a string of neighbourhoods, and latitude and longitudes. From this dataset, we also construct a 'condolike' variable, conditional on whether the address contains an apartment number (as in Adler (2010)). In addition

[4] A tertiary consideration is that we are keen to make our replication files operable on local machines as opposed to high performance/distributed systems.

to this, we also generate thirty two neighbourhood and seventeen monthly dummies. In Figure 3.4 we plot the spatial distribution of each of the samples across latitude and longitude, and it is from these latitude-and-longitude that we calculate distance matrices using the Haversine formula. Figure 3.5 displays the distribution of retail sale prices across time. Figures 3.6 and 3.7 show two dimensional scatter plots of each of the four main variables against retail price.

3.4.3 Variation over delta

As discussed above, we create variation over the modeler choice set by re-estimating the spatial error model with sequentially specified weighting matrices where we expand the maximum threshold permitted for non-zero entries across the distance (δ^S) and time (δ^T) dimension. Specifically, we specify 400 different matrices constructed as per each of the methods in Equations 3.7-3.9 in twenty intervals between each of $\delta^S = \{0.1, \ldots, 0.9\}$ and $\delta^T = \{0.1, \ldots, 0.9\}$. This is done for each of $W_A^{\alpha,\delta}$, $W_B^{\alpha,\delta}$ and $W_C^{\alpha,\delta}$. In Subsubsections 3.4.3.1-3.4.3.6, as in Subsection 3.4.4, we consider the effect of these variations across Spatial Hausman tests, Moran's I, in-sample mean absolute percentage errors (MAPE), $\bar{R}^2$, the spatial autocorrelation parameter ($\lambda_{d,m}^{T,S,\alpha}$) and four key coefficients related to four variables which feature in each of our two datasets. In Subsection 3.4.3 we provide an exposition of the spatial statistics, evaluation criteria and coefficients (in addition to the results from variation across δ), and in Subsection 3.4.4 we focus solely on the movements of the same statistics, criteria and coefficeints for variation in α.

3.4.3.1 Spatial Hausman tests

An implicit assumption of the model which we are using is that our X_d matrix is well specified and that there are no omitted variables which are correlated with the included explanatory variables. We test the appropriateness of our modelling strategy using a formal testing procedure. The test which we employ is the 'spatial Hausman test' developed in Pace and LeSage (2008), based on a comparison of least squares estimates for the coefficient vector $\beta_{d,OLS}$ with those from the SEM ($\beta_{d,m}^{T,S,\alpha}$). Anselin (1988) shows that OLS and SEM parameter estimates should both be unbiased under SEM assumptions (irrespective of the choice of $W_{d,m}^{T,S,\alpha}$).

The application of the spatial Hausman test is a useful inclusion because under the null hypothesis there are two consistent estimators differing in efficiency, with an alternative hypothesis of misspecification the two estimators (OLS and maximum likelihood estimation of the SEM). As noted in the original (Pace and LeSage, 2008, p.4) paper: "Divergence between the coefficient estimates from SEM and OLS suggests that neither is yielding regression parameter estimates matching the underlying parameters in the DGP. This calls into question the use of either OLS or SEM for that set of variables"

The intuition which lies behind this testing procedure is that any significant differences between OLS and SEM estimates of β indicate model mis-specification. While omitted from several studies, this kind of comparison must form a natural part of all empirical work which uses spatial error models. Where $\tilde{\Lambda}_{d,m}^{T,S,\alpha} = \hat{\beta}_{d,OLS} - \tilde{\beta}_{d,m}^{T,S,\alpha}$ represents the difference between OLS and SEM estimates of the estimated coefficients, Pace and LeSage (2008) show that their Hausman test statistic ($\tilde{H}_{d,m}^{T,S,\alpha}$) can be defined (using our continuing notation) as:

$$\tilde{H}_{d,m}^{T,S,\alpha} = \tilde{\Lambda}_{d,M}^{T,S,\alpha^\top} (\tilde{V}_{OLS} - \tilde{V}_{d,m}^{T,S,\alpha})^{-1} \tilde{\Lambda}_{d,m}^{T,S,\alpha} \tag{3.12}$$

where $\tilde{V}_{OLS}$ is a consistent estimate of the variance-covariance matrix associated with the OLS model, and $\tilde{V}_{d,m}^{T,S,\alpha}$ is a consistent estimate of the variance-covariance matrix from the SEM. The null hypothesis which we are testing therefore is that $\beta_{d,OLS} = \beta_{d,m}^{T,S,\alpha}$, and surface plots of p-values[5] of this test are seen in Figure 3.8.

The extremely small p-values ($p < 0.01$) for a large percentage of the weighting matrix specifications indicate strong evidence against the null hypothesis that the model is correctly specified as an SEM . This is despite the fact that the set of independent variables which we use (X_c) is almost identical to a number of specifications used by a range of academic studies such as LeSage and Pace (2004)[6], Dubé and Legros (2013a)[7] and Zhu *et al.* (2011)[8]. While there

[5] We are grateful for the insightful code provided by Justin M. Ross at https://sites.google.com/site/jross08/matlabfunctions which we re-write, tailoring it for our specific purpose.

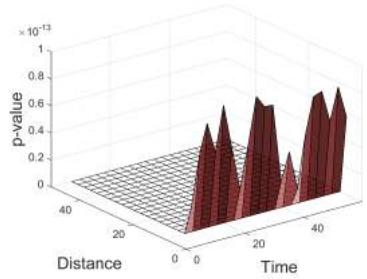
[6] In this study, explanatory variables were reported to be: a constant, age of the home, age squared, age cubed, size of the lot (logged), number of rooms, total living area (logged) and the number of bedrooms.

[7] In this study, explanatory variables were reported to be: a constant, age of the home, the log of the age, the log of the living area, log of the lot size, the number of bathrooms, garage size, whether the house had two or more stories and year dummies.

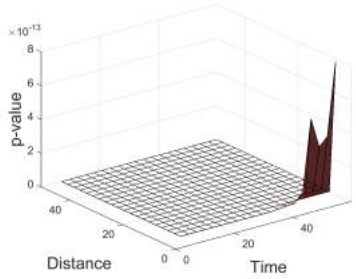
[8] In this study, explanatory variables were reported to be age, age squared, the log of size, the log of the living area, the number of rooms, the number of bedrooms, the log of the lot size squared, a dummy variable for using asbestos or composite tile shingles or stone or wood siding, dummy variable for brick

Figure 3.8: Hausmans p-value: delta variation

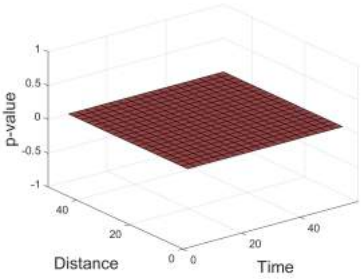
3.8.1 $W_A^{\alpha,\delta}$: Ohio



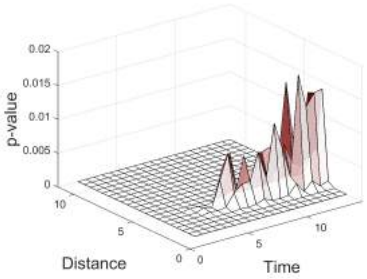
3.8.2 $W_B^{\alpha,\delta}$: Ohio



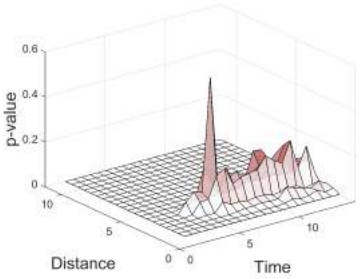
3.8.3 $W_C^{\alpha,\delta}$: Ohio



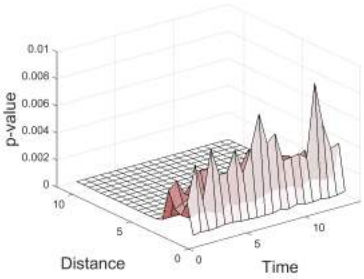
3.8.4 $W_A^{\alpha,\delta}$: San Francisco



3.8.5 $W_B^{\alpha,\delta}$: San Francisco



3.8.6 $W_C^{\alpha,\delta}$: San Francisco



is sporadic variation across m and d (the matrix specification and the dataset used) regarding the magnitude of the p-value, the result is almost entirely unambiguous, drawing into question the use of the SEM model with these variable specifications, regardless of the weighting matrix chosen. An interesting consideration is the location of the peaks in the panels. In all but two of the panels we observe peaks of sufficient magnitude for us to unilaterally reject the null regardless of the position in δ^T and δ^S space. Peaks occur in five of the panels of Figure 3.8, all occurring at the comparatively smaller values of δ^S. Pace and LeSage (2008) cite the necessity to therefore 'enrich' the spatial specification included in X_c in specifications of this type, with the intention of reducing the correlation between the stochastic disturbance term and the included variables. The top right sub-figure displays an entirely flat panel, for which every value is zero to at least six decimal places.

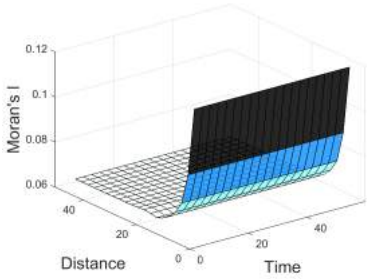
3.4.3.2 Moran's I

Of particular interest is the comment of Dubin (1998) related to Can (1992) that when using tests such as Moran's I, the detection of spatial autocorrelation is completely reliant on the

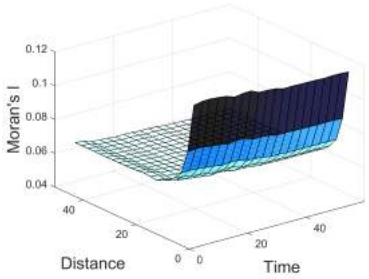
siding, the number of stories, a dummy variable for garage, quarterly sale dummies, and x/y co-ordinates.

Figure 3.9: Moran's I: delta variation

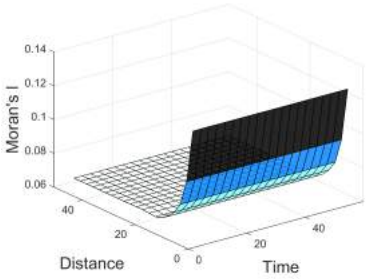
3.9.1 $W_A^{\alpha,\delta}$: Ohio



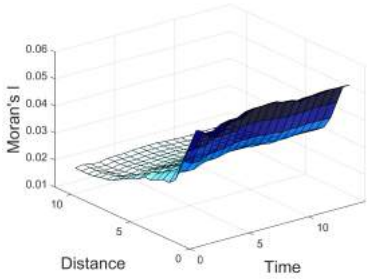
3.9.2 $W_B^{\alpha,\delta}$: Ohio



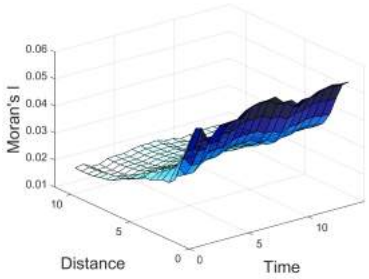
3.9.3 $W_C^{\alpha,\delta}$: Ohio



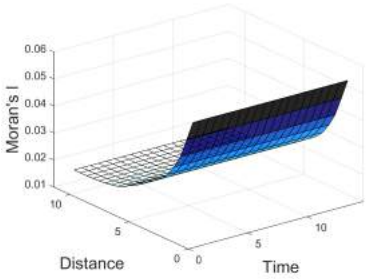
3.9.4 $W_A^{\alpha,\delta}$: San Francisco



3.9.5 $W_B^{\alpha,\delta}$: San Francisco



3.9.6 $W_C^{\alpha,\delta}$: San Francisco



specification of the weighting matrix. Keeping consistent sub- and super-script notation, the test statistic can be written as:

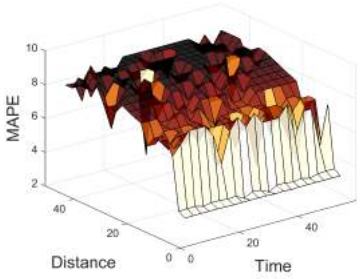
$$I_{d,m}^{T,S,\alpha} = \frac{N_d}{\sum_{i=1}^{N}\sum_{j=1}^{N}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,j}(y_{i,d} - \bar{y_d})(y_{j,d} - \bar{y_d})}{\sum_{i=1}^{N}(y_{i,d} - \bar{y}_d)^2} \tag{3.13}$$

While it is also possible to test the significance through transformation to a Z-score, for concision we only visualise the unstandardized values across all possible weighting matrices in Figure 3.9. However, in Can (1992), the inconclusive result is that three of the six combinations of weighting matrix and linear/semi-log specifications show spatial autocorrelation, and three do not. The unstandardized variant provides evidence on the magnitude of spatial correlation. Bogdon and Can (1997) claim that I statistics of between 0.17 and 0.20 indicate 'moderate to high' levels of spatial clustering. However, the main comparison is naturally with Figure 8 of Dubé and Legros (2013a). Taking a similar, albeit less computational approach, their figure plots the values of Moran's Index across eleven different 'optimal distances' used in the construction of their $W = S \circ T$ matrix (where T is lower triangular based on an ordering of the data). Our findings concur and replicate this study, in that as δ^S increases, Moran's I (spatial

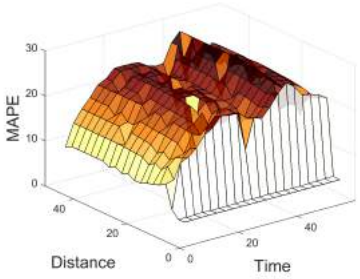
autocorrelation) decreases. In Dubé and Legros (2013a), the value drops from approximately 0.3 with a 500m threshold to 0.1 with a 3,000m threshold for the same Lucas county dataset, and ranges between 0.0819 (3,000m threshold, spatial weighting matrix) and 0.3044 (1,500m threshold, spatio-temporal weighting matrix) in Dubé and Legros (2014a) for their Quebec dataset. Our values differ due to the fact that our minimum value of $\delta^S = 0.1$ corresponds to a threshold distance of 5.06km, and their smallest threshold pertains to a 500m cutoff. We observe an identical pattern however, with the value decreasing to 0.06-0.08 as δ^S becomes larger and more observations (which are a greater distance apart in time and space) are set to non-zero values in our weighting matrices. While the pattern of results is consistent between datasets and weighting specifications (with no variation caused by the time dimension, as predicted by economic theory), the main contribution which we make (in addition to estimating a larger set of weighting matrices for a fixed δ^T alone) is to document the importance of variation in δ^T and δ^S.

Figure 3.10: MAPE: delta variation

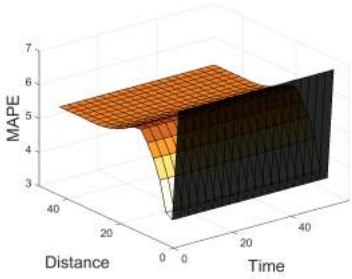
3.10.1 $W_A^{\alpha,\delta}$: Ohio



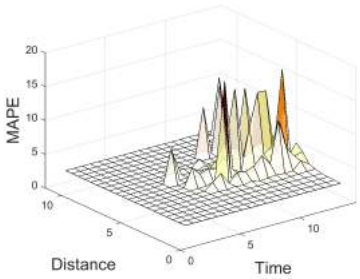
3.10.2 $W_B^{\alpha,\delta}$: Ohio



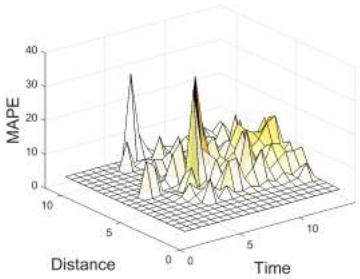
3.10.3 $W_C^{\alpha,\delta}$: Ohio



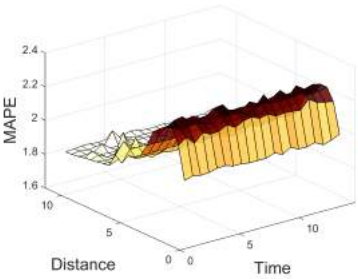
3.10.4 $W_A^{\alpha,\delta}$: San Francisco



3.10.5 $W_B^{\alpha,\delta}$: San Francisco



3.10.6 $W_C^{\alpha,\delta}$: San Francisco



3.4.3.3 Movement of the MAPE

In Figure 3.10 we consider the mean absolute percentage errors ($MAPE_{d,m}^{T,S,\alpha}$), calculated as:

$$MAPE_{d,m}^{T,S,\alpha} = \left| \frac{\sum_{i=1}^{N_d}(\hat{y}_{i,d} - y_{i,d})}{y_{i,d}} \right| \times \frac{1}{N_d} \tag{3.14}$$

The results from this evaluation metric need to be considered in conjunction with Section 3.4.3.4. Firstly, the figures show that for both of the spatio-temporal weighting matrices, a more conservatively sized matrix is able to consistently provide lower mean absolute percentage errors. For the conservatively estimated weighting matrix, we can see an average MAPE in the region of 1.5-3%, although deviation towards less sparse matrices can induce the possibility of MAPE values as high as 33.28% for a $W_B^{\alpha,\delta}$, or 16.11% for $W_A^{\alpha,\delta}$ in the San Francisco dataset. The slight variation across the time dimension which we observe in the bottom right panel (and also that which was seen in the above figures) is especially interesting, given that the specification of the matrices is invariant to the date of sale. This potentially anomalous observation is attributable solely to the fact that at every element in $\delta^S \times \delta^T$ space, we are re-estimating the SEM (essentially with twenty identical specifications for each δ^S for $W_C^{\alpha,\delta}$) with a convergence criteria of $1e-4$. This criteria is not sufficiently small to generate identical estimates from identical specifications.

3.4.3.4 Movement of the $\bar{R}^2$

We next consider the movement of $\bar{R}^2$ across δ^T and δ^S space. Recalling that it is a *lower* MAPE and a *higher* $\bar{R}^2$ which represents a better model fit, the results between Figure 3.10 and 3.11 concur. In this subsection, this is indicated by the fact that the specifications which yield the largest $\bar{R}^2$ are again systematically located in the sparser matrices for all specifications and both datasets, where matrices with fewer non-zero entries provide lower $\bar{R}^2$ values. The pattern in both graphs is similar, although, while the $\bar{R}^2$ is increasing as the distance threshold decreases for the Ohio dataset from the smallest distance threshold parameter (with surprisingly no variation across the time dimension for any weighting matrix estimator), there is a constant region above the δ^S minimum for which $\bar{R}^2$ is highest. For the San Francisco dataset, the two spatio-temporal specifications ($W_A^{\alpha,\delta}$ and $W_B^{\alpha,\delta}$) show a generally increasing $\bar{R}^2$ as the time

dimension increases and more observations, located further in the past are incorporated. Again, we see slight variation in the time dimension for $W_C^{\alpha,\delta}$, and we again attribute this to the effect of the convergence criteria on the estimated parameters.

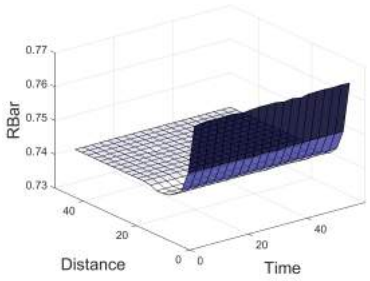
The results from the two datasets generally straddle those found in the literature discussed in Section 3.2. The $\bar{R}^2$ values are both higher and lower than those seen in Wilhelmsson (2002) (a range of 0.678-0.689), and both higher and lower than the equations for log price in the complete specifications in the hedonic model of Bracke (2014) (0.621 and 0.858, although significantly higher for regressions run on sub-samples of 'matched' properties). Our results for Ohio are almost comparable with Fik *et al.* (2003) (0.844) and their 'location value signature' approach, and, naturally, approximately equal (for more sparse weighting matrices) to the full sample estimations for the Lucas County dataset in Dubé and Legros (2014a).

3.4.3.5 Movement of λ

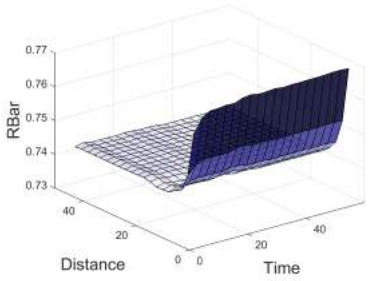
In Figure 3.11 we take the same approach as above, considering the movement of $\lambda_{d,m}^{T,S,\alpha}$ from Equation 3.11 across the specification space. The positive values of spatial correlation (where

Figure 3.11: $\bar{R}^2$: delta variation

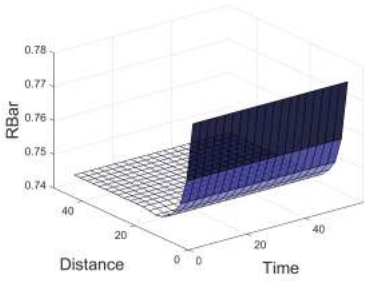
3.11.1 $W_A^{\alpha,\delta}$: Ohio



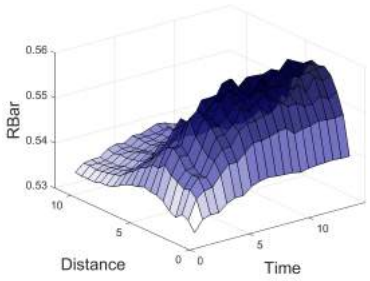
3.11.2 $W_B^{\alpha,\delta}$: Ohio



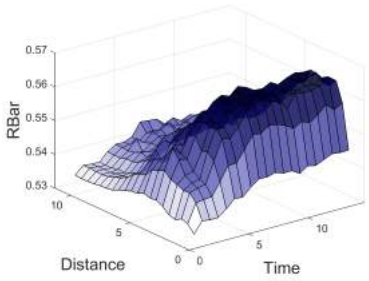
3.11.3 $W_C^{\alpha,\delta}$: Ohio



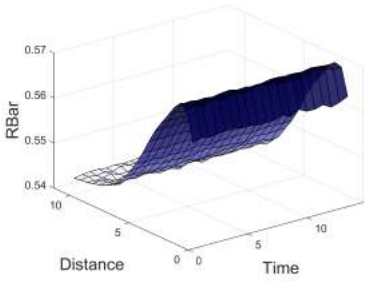
3.11.4 $W_A^{\alpha,\delta}$: San Francisco



3.11.5 $W_B^{\alpha,\delta}$: San Francisco



3.11.6 $W_C^{\alpha,\delta}$: San Francisco



$\lambda_{d,m}^{T,S,\alpha}$ is in the range $[-1, 1]$) imply the (expected) clustering of similar error values, violating independence assumptions. Different to the SAR model, in our case, dependence enters only through the error terms. This restricts observations to being related through unmeasured locational factors correlated across distances, rather than directly through a relationship between observations in our independent variable. The results show a far stronger heterogeneity in terms of range and spread of mass than the previous statistics considered, with the highest values (0.990) typically occurring when $W_{d,m}^{T,S,\alpha}$ is less sparse (large δ^t and δ^s). This value corresponds directly to the results reported in Dubé and Legros (2014a) (in the third column of their Table 1) for the 'spatial only' specification (which is approximately equivalent to the case where δ^T is large), citing the possibility of spurious results from spatial non-stationarity; or at the least, an over-estimation of the spatial co-efficient.

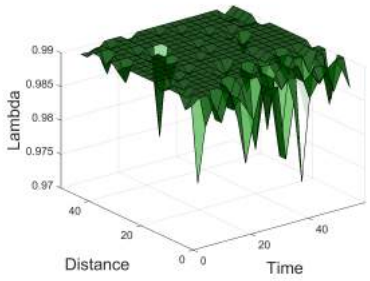
Figure 3.12 again features the occurrence of sporadically distributed peaks in the surface plots for $W_A^{\alpha,\delta}$ and $W_B^{\alpha,\delta}$. While this is potentially due to the aforementioned convergence criteria, we argue this is also potentially indicative of a sensitivity of $\lambda_{d,m}^{T,S,\alpha}$ to $W_{d,m}^{T,S,\alpha}$, whereby these peaks could be caused by as few as a couple of non-zero entries (potential outliers) due to changes in δ^T and δ^S. Our results are significantly higher than those reported in Dubé and Legros (2013a) (which are 0.5809, 0.6840 and 0.6850), and this is again representative of the fact that our minimum cutoff values are substantially higher than their maximum. Conversely, for the San Francisco dataset, the estimated parameter $\lambda_{d,m}^{T,S,\alpha}$ is noticeably significantly lower when fewer observations are incorporated (smaller δ^T and δ^S). We see huge variations across the model specification space, ranging from 0.9755, 0.99 and 0.99 (in the region of less sparse weighting matrices), to 0.1177, 0.1130, 0.7290 - where matrices are more sparse (the 'six o'clock' region).

3.4.3.6 Further Analysis

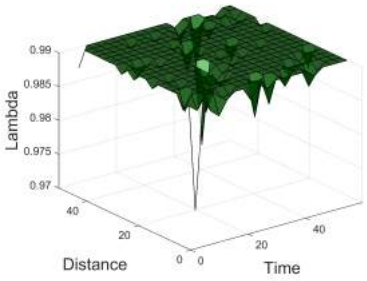
To provide an additional dimension to the metrics visualized above, Table 3.1 shows the range and mean of the coefficient estimates across each of our 400 weighting matrix specifications, which is an avenue we also examine visually in Section 3.4.4. One notable feature to mention is the constancy of each metric *between* matrix specification and dataset. We again focus on the variability which the modeler is subjected to through his weighting matrix specification

Figure 3.12: $\lambda_{d,m}^{T,S,\alpha}$: delta variation

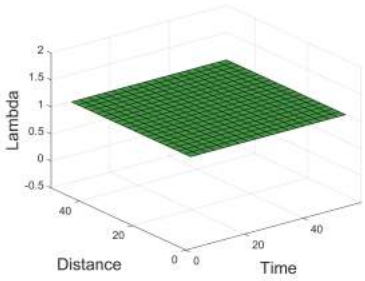
3.12.1 $W_A^{\alpha,\delta}$: Ohio



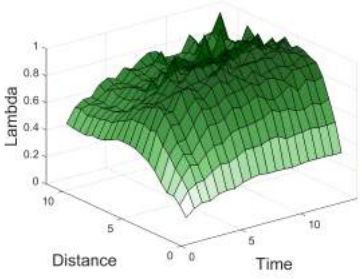
3.12.2 $W_B^{\alpha,\delta}$: Ohio



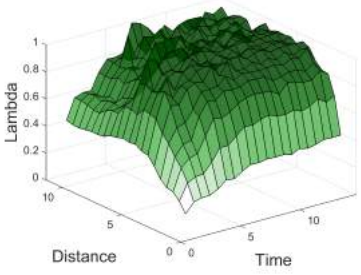
3.12.3 $W_C^{\alpha,\delta}$: Ohio



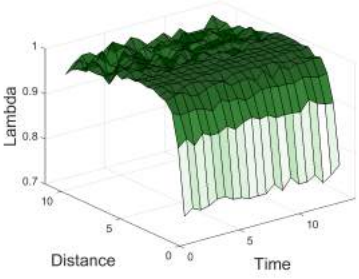
3.12.4 $W_A^{\alpha,\delta}$: San Francisco



3.12.5 $W_B^{\alpha,\delta}$: San Francisco



3.12.6 $W_C^{\alpha,\delta}$: San Francisco



choices. The largest variation typically comes from the β^{age} coefficient. The variation is so large that it reverses sign between its maximum and minimum value, to an equal, but opposite magnitude for the $W_B^{\alpha,\delta}$ matrix, and changes sign and doubles its magnitude for the $W_C^{\alpha,\delta}$ for the San Francisco dataset. The variation is almost equally as large for this coefficient in the Ohio dataset, approximately halving in value between the maximum and minimum estimate in δ space. The variance across the coefficients is smaller for other variables, albeit still in the magnitude of approximately 30%, which may have significant consequences on the conclusions drawn from a modelling exercise. Our results are broadly consistent with each of the three datasets and all models employed in Dubé and Legros (2014a), with an interesting observation being that the coefficient on the logarithm of living area (β^{living}) is almost unilaterally 2-4 times higher than the coefficient on the logarithm of lot size (β^{lotsize}). Naturally, as in earlier sections, there are no comparable applications for the San Francisco dataset as yet. We supplement Table 3.1 with plots of the histograms shown in Figures 3.13-3.18, and these subplots allow us to more fully consider the consistency of the distribution of coefficients.

Table 3.1: Variation across key coefficients in δ space

	Ohio			San Francisco		
Coefficient	$W_A^{\alpha,\delta}$	$W_B^{\alpha,\delta}$	$W_C^{\alpha,\delta}$	$W_A^{\alpha,\delta}$	$W_B^{\alpha,\delta}$	$W_C^{\alpha,\delta}$
β_{max}^{age}	0.9530	1.1022	1.0324	5.2300	5.2858	2.5975
β_{min}^{age}	0.5118	0.5142	0.5956	-2.8568	-4.7821	-5.7028
β_{mean}^{age}	0.9013	1.0163	0.9881	1.8674	0.8714	-0.9422
β_{max}^{living}	0.8549	0.8602	0.8571	0.2923	0.2927	0.2763
β_{min}^{living}	0.8084	0.8042	0.8078	0.2562	0.2504	0.2287
β_{mean}^{living}	0.8485	0.8526	0.8511	0.2778	0.2773	0.2568
β_{max}^{beds}	0.0093	0.0109	0.0114	0.0140	0.0140	0.0138
β_{min}^{beds}	0.0006	0.0007	0.0024	0.0102	0.0099	0.0103
β_{mean}^{beds}	0.0022	0.0027	0.0039	0.0124	0.0124	0.0126
$\beta_{max}^{lotsize}$	0.1853	0.1847	0.1849	0.1325	0.1398	0.1412
$\beta_{min}^{lotsize}$	0.1594	0.1547	0.1609	0.1161	0.1176	0.1242
$\beta_{mean}^{lotsize}$	0.1758	0.1744	0.1757	0.1215	0.1252	0.1313

3.4.4 Variation across alpha space

In Section 3.4.3 we concerned ourselves with a narrow specification of the spatio-temporal matrix in order to consider how key spatial statistics varied over threshold (δ) space. In this section we consider the same three common, competing variations of the matrix specification across the choice of exponent α when it was previously constant (set to $\alpha = 1$), but now assuming a fixed value of $\delta = 0.5$. We sequentially assign an exponent to each element in intervals of $\alpha = \{1.01, 1.02, \ldots, 5\}$, re-estimating the model and storing results. For each weighting matrix ($W_A^{\alpha,\delta}$, $W_B^{\alpha,\delta}$, $W_C^{\alpha,\delta}$) we row standardize, and apply the four hundred exponents as described above, again focusing exclusively on the spatial error model. For each estimation, we store and then plot three commonly examined spatial statistics in Figure 3.19, the movement of four commonly estimated coefficients (Figure 3.20), and then two evaluation criteria (the $\bar{R}^2$ and the MAPE) in Figure 3.21. The metrics considered correspond exactly to those which featured in Section 3.4.3. We again utilize the exact same two datasets and an identical set of dependent and independent variables with the sole desire of examining variation caused by modeler choice. In each of the following graphs, the red solid lines represent $W_A^{\alpha,\delta}$ specifications, the black dashed lines represent $W_B^{\alpha,\delta}$ and the blue dotted lines represent $W_C^{\alpha,\delta}$ specifications.

Figure 3.13: Ohio coefficient variation: $W_A^{\alpha,\delta}$

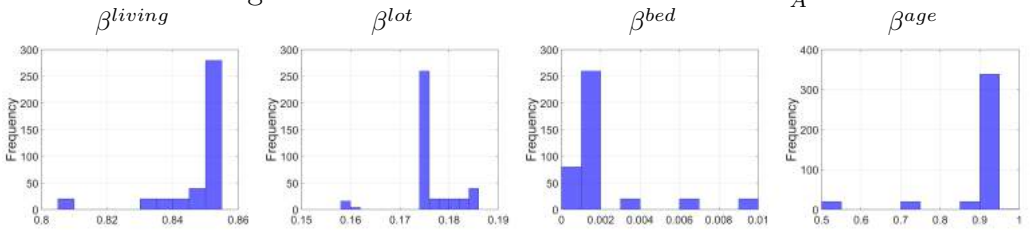


Figure 3.14: Ohio coefficient variation: $W_B^{\alpha,\delta}$

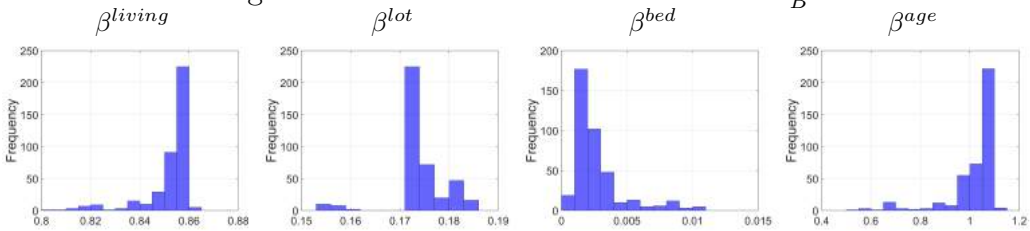


Figure 3.15: Ohio coefficient variation: $W_C^{\alpha,\delta}$

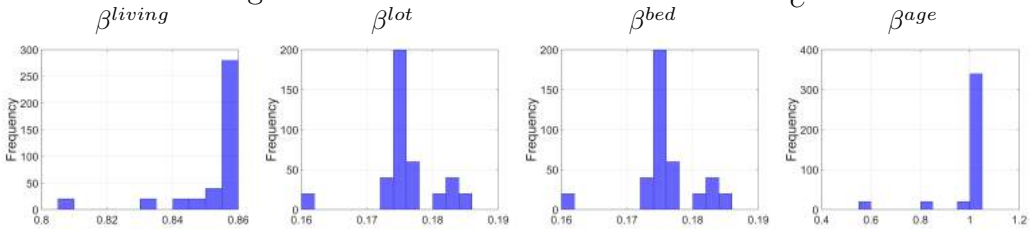


Figure 3.16: San Francisco coefficient variation: $W_A^{\alpha,\delta}$

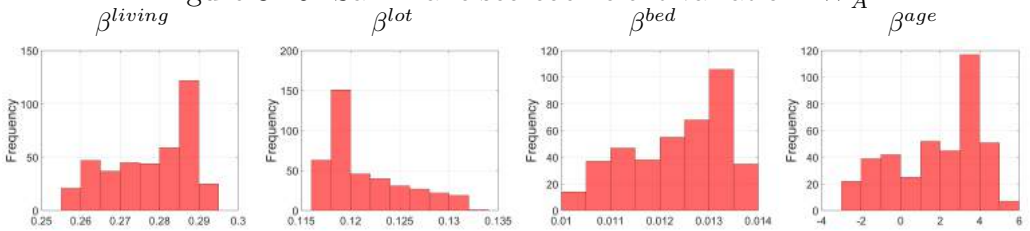


Figure 3.17: San Francisco coefficient variation: $W_B^{\alpha,\delta}$

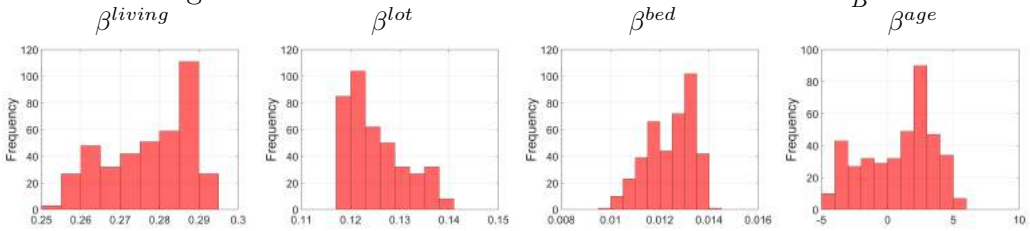


Figure 3.18: San Francisco coefficient variation: $W_C^{\alpha,\delta}$

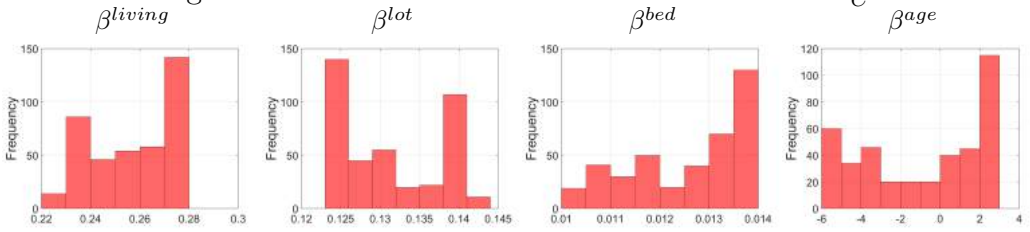
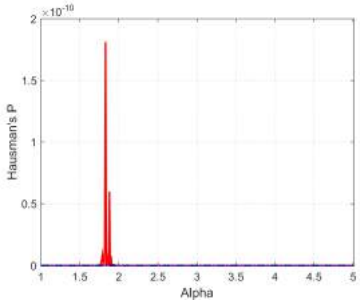
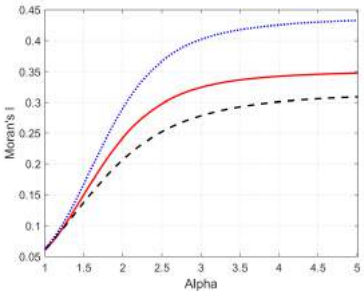


Figure 3.19: Movement of spatial statistics in α space

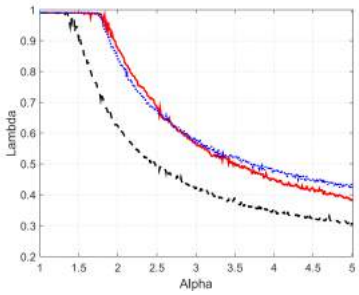
3.19.1 Hausman P-value: Ohio



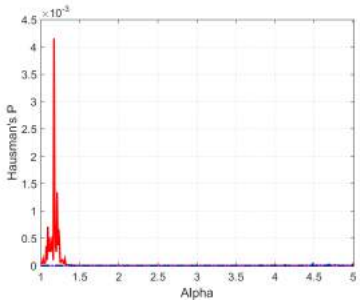
3.19.2 Moran's I: Ohio



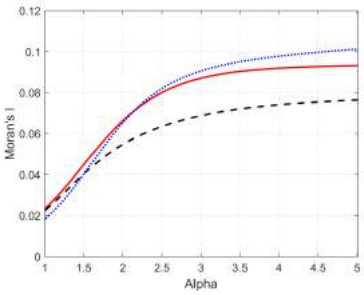
3.19.3 Lambda: Ohio



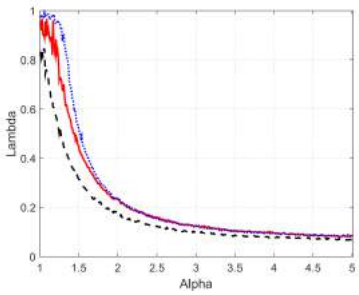
3.19.4 Hausman P: San Fran.



3.19.5 Moran's I: San Fran.



3.19.6 Lambda: San Fran.



3.4.4.1 Movement across spatial statistics

Firstly, the results of Figure 3.19 again show (as in Section 3.4.3) that the spatial Hausman test is highly sensitive to the specification of W. The middle panel of Figure 3.19 documents the movement of the unstandardized Moran's I, and the third panel shows the increasing degree of spatial autocorrelation as observations further away in distance and time are given a lower proportional weight (the interpretation of increasing α). It is important to identify the fact that at $\alpha = 1$, the results (should) correspond to the specific case when $\delta^T = \delta^S = 0.5$ (subject to convergence criteria) as seen in Section 3.4.3. Given this, what is of particular interest is how λ shifts from the previously observed 0.99 to as low as 0.0840 for $W_C^{\alpha,\delta}$ in the San Francisco dataset.

3.4.4.2 Movement across coefficients

In Figure 3.20, the conclusion to be drawn is that the differing matrix estimators themselves provide relatively similar coefficient ranges, frequently displaying identical patterns of movement across α space. It is instead the movement along the permissible range of $\alpha = [1, 5)$ which drives

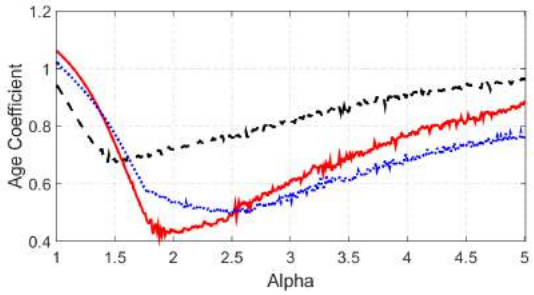
the variability in coefficient estimates. The majority of this variance occurs in the range of between $\alpha = 1$ and $\alpha = 3$, with estimates showing some convergence as the exponent gets larger, indicated by the increasingly horizontal shape of each line in each sub-figure. Again, we see the largest coefficient variance for β^{age} using the San Francisco dataset, which ranges from as low as -1.8954 at $\alpha = 1.28$, to as high as 5.2740 at $\alpha = 5.00$. An interesting comparison arises from an unreported result for the dummy variable 'Condo', which provides coefficient estimates extremely similar to the two equivalent estimates provided in Bracke (2014) for apartments in Central London (-0.110 and -0.076).

3.4.4.3 Movement across evaluation metrics

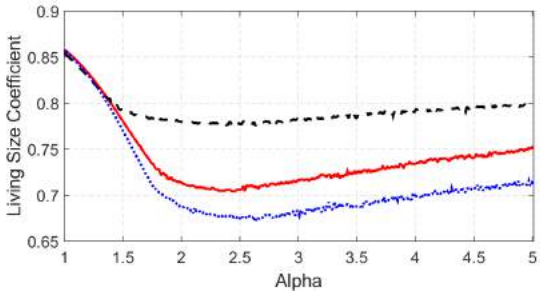
Finally, Figure 3.21 shows the behavior of the two in-sample evaluation criteria across α-space for the three weighting matrix specifications. A direct comparison of the number of times which each of these specifications (used in conjunction with an SEM framework) performs best, and how frequently they are able to outperform OLS and (equivalent SEM based) nearest neighbour specifications is the specific focus of Section 3.4.5. While the focus of this section is on variation across α as opposed to constancy, a brief comparison with the results found in Section 3.4.3 provides valuable insight. Increasing α above the commonly used value of $\alpha = 1$ appears to have the ability to diminish the weight attributed to potential outliers, which causes the MAPE 'spikes' seen in this section, and to dramatically reduce the average MAPE seen in Section 3.4.3 where α is set to one for all specifications. The elimination of spikes for higher values of α is consistent with the underlying intuition regarding what α represents. One of the most interesting results of the section can be seen in the top right panel of Figure 3.21. In this figure, it is shown that all three weighting specifications start at approximately the same $\bar{R}^2$ value (0.74). However, they quickly rise to values which far surpass any value obtained through variation of δ in Section 3.4.3 (the highest being $\bar{R}^2 = 0.8195$ for $W_C^{\alpha,\delta}$ at $\alpha = 2.10$); an increase of over 10%, while still close to the most commonly used value ($\alpha = 1$). This is especially striking given that $\alpha = 2.1$ represents an entirely plausible modeler choice. Considering the variation between matrix construction method, it provides an entire 7.3% increase in $\bar{R}^2$ against an exactly equivalent $W_B^{\alpha,\delta}$ specification.

Figure 3.20: Movement of coefficients in α space

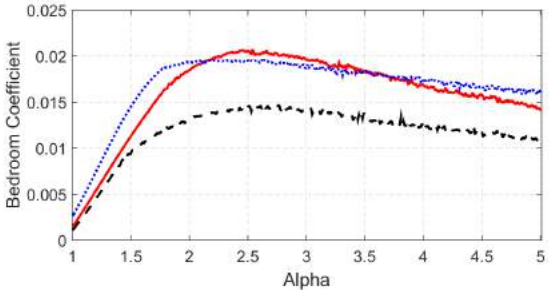
3.20.1 β^{age}: Ohio



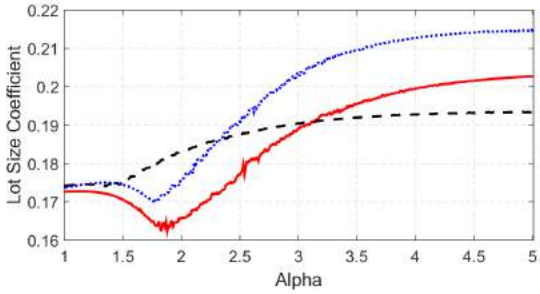
3.20.2 β^{living}: Ohio



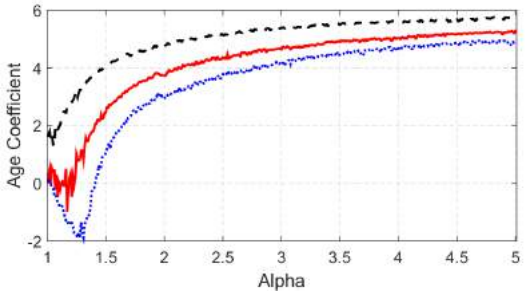
3.20.3 $\beta^{bedrooms}$: Ohio



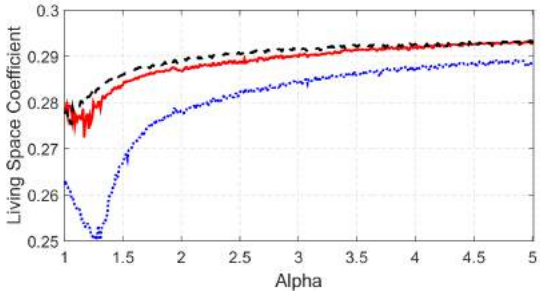
3.20.4 $\beta^{lotsize}$: Ohio



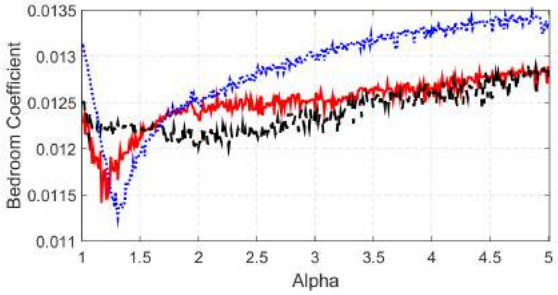
3.20.5 β^{age}: San Fran.



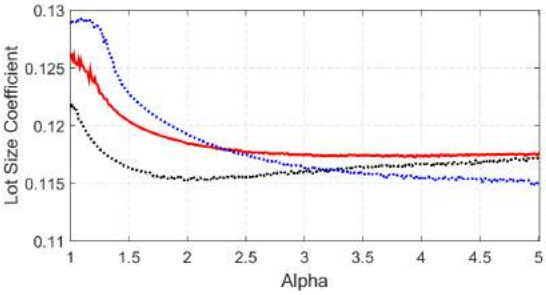
3.20.6 β^{living}: San Fran.



3.20.7 $\beta^{bedrooms}$: San Fran.



3.20.8 $\beta^{lotsize}$: San Fran.



3.4.5 Direct comparison between specifications

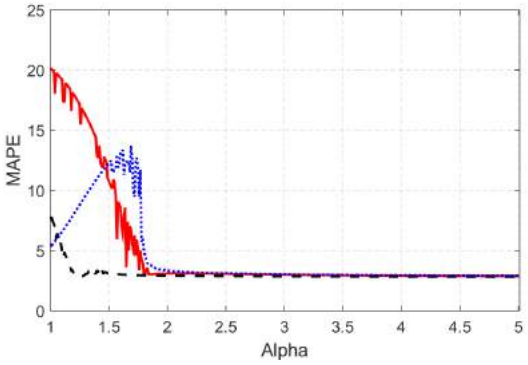
In this section we amalgamate the two sources of variation and the two datasets used in Sections 3.4.3-3.4.4 to directly evaluate each of the three matrix specifications. Noting that in each of the two preceding sections we create 400 intervals over both α and δ space, we now consider how

many times a specific method of weighting matrix specification performs best at each interval for each of the two datasets, where performance is again measured either by the $\bar{R}^2$ or MAPE. We also estimate 'static' hedonic OLS and nearest neighbour SEM comparisons (set to five nearest neighbours, and a second order Delaunay triangulation, again normalized to have row-sums of unity) as comparable benchmarks. The results are shown in Tables 3.2 and 3.3 for the $\bar{R}^2$ and MAPE respectively.

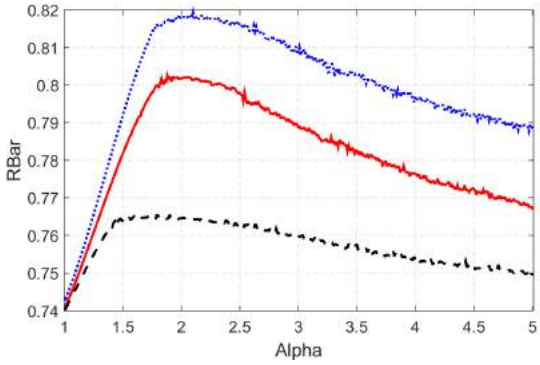
The results firstly allow a review of the statistics from the relevant earlier sections of the chapter. For example, we can see the highest $\bar{R}^2$ achieved occurs using the $W_C^{\alpha,\delta}$ specification for an α variation for Ohio (0.8195), while the lowest occurs using a δ variant of the $W_B^{\alpha,\delta}$ matrix (0.7388). We can again witness the fact that the range of $\bar{R}^2$s from the Bay Area dataset is significantly lower than corresponding Ohio values, and that the largest 'peaks' in the aforementioned in-sample MAPE values generally occur when using the $W_B^{\alpha,\delta}$ specification. Evaluating them directly, the specification which is able to generate the highest absolute $\bar{R}^2$

Figure 3.21: Movement of evaluation metrics in α space

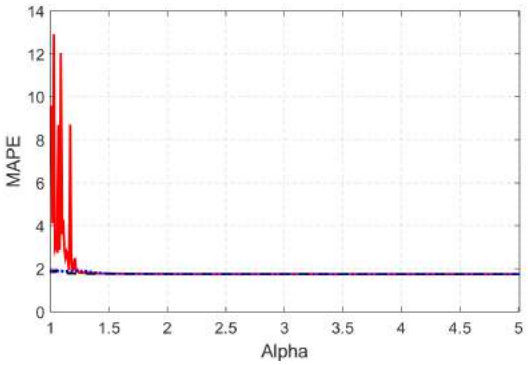
3.21.1 MAPE: Ohio



3.21.2 $\bar{R}^2$: Ohio



3.21.3 MAPE: San Fran.



3.21.4 $\bar{R}^2$: San Fran.

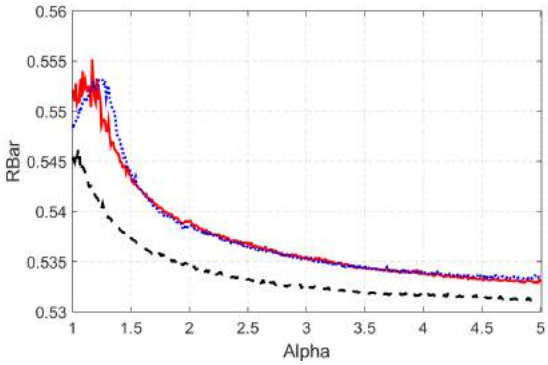


Table 3.2: Comparative evaluation of the $\bar{R}^2$

α - **San Francisco**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	0.5463	0.5311	0.5337	0	100	100	8.25
$W_B^{\alpha,\delta}$	0.5551	0.5328	0.5373	201	0	100	22.5
$W_C^{\alpha,\delta}$	0.5531	0.5330	0.5374	199	0	100	22
α - **Ohio**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	0.7655	0.7404	0.7575	0	400	100	0
$W_B^{\alpha,\delta}$	0.8025	0.7410	0.7823	0	0	100	0
$W_C^{\alpha,\delta}$	0.8195	0.7430	0.7994	400	0	100	0
δ - **San Francisco**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	0.5578	0.5310	0.5432	0	373	100	58
$W_B^{\alpha,\delta}$	0.5640	0.5309	0.5471	132	22	100	81.25
$W_C^{\alpha,\delta}$	0.5689	0.5407	0.5522	268	5	100	25
δ - **Ohio**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	0.7633	0.7386	0.7413	0	384	100	0
$W_B^{\alpha,\delta}$	0.7676	0.7388	0.7423	0	16	100	0
$W_C^{\alpha,\delta}$	0.7735	0.7415	0.7448	400	0	100	0

Table 3.3: Comparative evaluation of the MAPE

α - **San Francisco**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	1.928	1.751	1.758	171	144	0	83.75
$W_B^{\alpha,\delta}$	12.92	1.750	1.909	167	0	0	78.75
$W_C^{\alpha,\delta}$	1.895	1.749	1.766	62	256	0	79
α - **Ohio**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	7.722	2.828	3.010	0	394	0	93
$W_B^{\alpha,\delta}$	20.06	2.875	4.940	48	0	0	79.75
$W_C^{\alpha,\delta}$	13.71	2.902	4.393	352	6	0	70.5
δ - **San Francisco**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	16.11	1.753	2.437	19	270	0	2.5
$W_B^{\alpha,\delta}$	33.28	1.753	3.999	200	21	0	2.5
$W_C^{\alpha,\delta}$	2.241	1.780	1.945	181	109	0	0
δ - **Ohio**	Max	Min	Av	# Highest	#Lowest	% Beats OLS	% Beats NN
$W_A^{\alpha,\delta}$	8.422	2.948	7.079	9	8	100	93.25
$W_B^{\alpha,\delta}$	26.26	2.828	17.61	351	38	100	90
$W_C^{\alpha,\delta}$	6.667	3.372	5.112	40	88.5	0	0

($W_C^{\alpha,\delta}$) is also the one which performs best out of the three most frequently in this criterion, that is 1,267 out of 1,600 instances. Correspondingly, the $W_A^{\alpha,\delta}$ specification generates the lowest $\bar{R}^2$ a total of 1,557 times across both datasets and sources of variation. Compared to benchmarks, each

of the three different sets of 1,600 spatial specifications generates an $\bar{R}^2$ higher than an equivalent hedonic model estimated with simple OLS, although the nearest neighbour specification produces a unilaterally superior performance for the Ohio dataset. Comparing performance across the MAPE, we see the converse becoming true: the nearest neighbour specification is well beaten in three of the four variation/dataset combinations. In addition, the OLS model produces a lower MAPE 1,200 out of 1,600 times for $W_A^{\alpha,\delta}$ and $W_B^{\alpha,\delta}$, and the full 1,600 times for $W_C^{\alpha,\delta}$. Directly comparing between the specifications, $W_A^{\alpha,\delta}$ performs best of the three 816 times compared to 725 times for the $W_C^{\alpha,\delta}$ matrices, with $W_B^{\alpha,\delta}$ the worst performing, generating the highest MAPE 766 times, and the lowest a mere 59.

3.5 Conclusion

We have analyzed the effect of modeler choice on the method of, and parameters related to, constructing spatial weighting matrices. In Section 3.3, we showed how correlations change given a fixed λ in the SEM model for nine different methods of specifying the weighting matrix (W) when using simulated data. In Sections 3.4.3 and 3.4.4 we used two commonly available, widely used datasets to emphasize the variation of key spatial statistics, evaluation methods and coefficients across a large number of possible modelling choices. Finally, in Section 3.4.5, we showed a broad preference for purely spatial ($W_C^{\alpha,\delta}$), as opposed to spatio-temporal ($W_A^{\alpha,\delta}$ or $W_B^{\alpha,\delta}$) weighting matrices. The chapter highlights the need for further work on the choice set related to the weighting matrix, with weighting schemes typically chosen at the discretion of the modeler.

In particular, this research might focus on three dimensions. First of all, future work might further consider the effect of misspecifying other exogenous weighting schemes in addition to those based on inverse time and distance, or specifically on the accepted procedure of row standardization. Secondly, future work by this author will develope a procedure which calculates the Bayesian log-marginal posterior for the spatial autoregressive model, and then uses this log-marginal posterior to compare the different weighting specifications. Along a parallel dimension, the pursuit of endogenously estimating weighting matrices remains a critical item on the research agenda, in addition to how key statistics vary by choices across estimation methods.

PREFACE TO CHAPTER 4

While the role of housing markets in the run up to the global recession of 2008-2012 is well documented, in this chapter we consider how housing markets have since responded to unconventional monetary policy shocks in the form of innovations both in total assets and the monetary base. We create unique quarterly and monthly housing market datasets for eight OECD countries, estimating a range of specifications with various identification strategies and estimators, and we consider several methods of response analysis. Despite cross-country heterogeneity from single country models, our mean group panel estimators show a positive, persistent response of house prices which peaks at between one and two years following a policy shock. Residential investment responds more strongly than house prices, but takes longer to peak, consistent with the conveyancing procedures involved in construction. While constrained to be negative in five of our six identification schemes, mortgage rates are shown to initially decrease by between five and six basis points before returning to zero. Both this chapter and Chapter 5 make use of a more conventional macroeconometric model of housing markets, and Chapter 4 represents the most policy orientated chapter of the thesis in general.

CHAPTER 4

HOUSING MARKETS AND UNCONVENTIONAL MONETARY POLICY

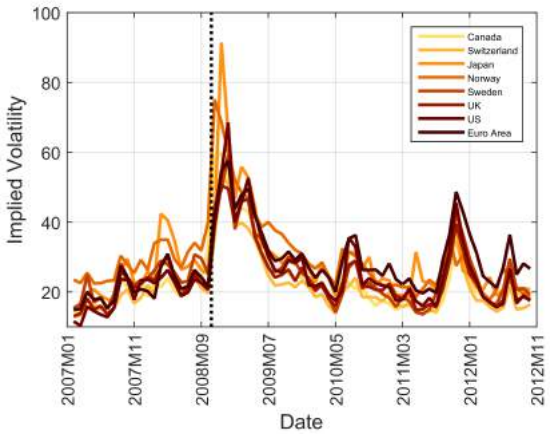
4.1 Introduction and policy background

The global recession of 2008-2012, developing out of the faltering United States mortgage market and two collapsed Bear Sterns hedge funds, began its trajectory as early as March 2007. With the threat of global financial and economic instability looming large in the world's most developed economies, governments and central banks began to pursue unorthodox policies in the face of recessionary and deflationary fears in order to prevent a collapse akin to the Great Depression of 1929. While most countries have pursued a mix of monetary and fiscal policies, some actions taken by central banks in response to the crisis have often been termed 'unconventional'. Such policies quickly became the primary monetary policy tool after interest rates began their descent to the zero lower bound (ZLB). Following the collapse of Lehman Brothers Holdings Inc. (formerly the fourth largest investment bank in the United States) on the 15$^{\text{th}}$ of September, 2008, economic confidence indicators fell and implied stock market volatility rose dramatically (Figure 4.1).

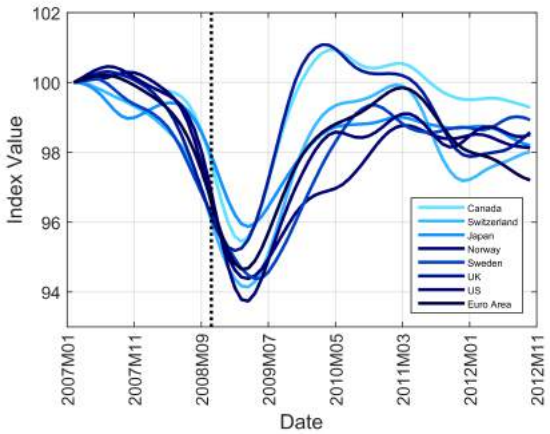
The role of housing is vital in the transmission mechanism of monetary policy. The first potential channel - the contentious housing wealth effect - has originated a significant body of literature in its own right, with varied empirical findings. There is also the closely related potential for a substitution effect, specifically concerned with the relative price of housing to other goods and services. In addition, it is possible that buoyant housing markets create aggregate

Figure 4.1: Economic indicators during the 'Great Recession'

4.1.1 Implied volatility of national stock markets



4.1.2 Confidence leading indicators



Source: OECD and volatility indices, dotted - collapse of Lehman Brothers Inc

demand in the form of transaction costs associated with high housing turnover, such as surveying fees, fixtures and furnishings, or a transaction related tax. Academic models postulate that house prices are an important determinant of residential investment (and, in particular, housing starts - Leamer (2007)) which can stimulate the supply side of an economy and create real activity. Another mechanism by which monetary policy may affect housing markets is through the credit channel by means of collateral values, as described in the model outlined in Iacoviello (2005). An important example of this channel is provided by Fuster and Willen (2010) who show that large reductions on agency mortgage-backed security rates around the 25$^{\text{th}}$ of November 2008 were quickly followed by large reductions in residential mortgage rates offered by lenders to households, something explicitly considered in this chapter. Unconventional monetary policy reduces long term real interest rates, reducing the user cost of housing, and therefore increasing the demand for (and price of) houses.

Following the financial crisis, it has never been more important to estimate the impact of policies on the value of residential assets. Any asset whose value is linked to residential real estate is an essential component of the aggregate portfolio of financial intermediaries (with increasing effects through leverage), and therefore significant for the wider economy. The need for this study can also be identified throughout the early literature on quantitative easing stemming from the crisis, where it is claimed: 'Analysis suggests that the purchases have had a significant impact on financial markets and particularly gilt yields, but there is clearly more to learn about the transmission of those effects to the wider economy' (Joyce *et al.*, 2011, p.4).

However, because of the nature of the crisis and the policies involved (due to the presence of the ZLB), standard econometric tools which would have otherwise included estimation across a pre-crisis period are not applicable[1]. Therefore, we take a rigorous, computational approach to not only consider the response of housing markets, but also how impulse responses vary dependent on the type of sample (panel or single country), the frequency of the data (quartly or monthly), the method of assessment (median, mean, mean group, or median target response), the type of restrictions employed (zero or sign) and the model specifications used. In particular, we follow three distinct identification strategies: a standard Cholesky decomposition, a range of

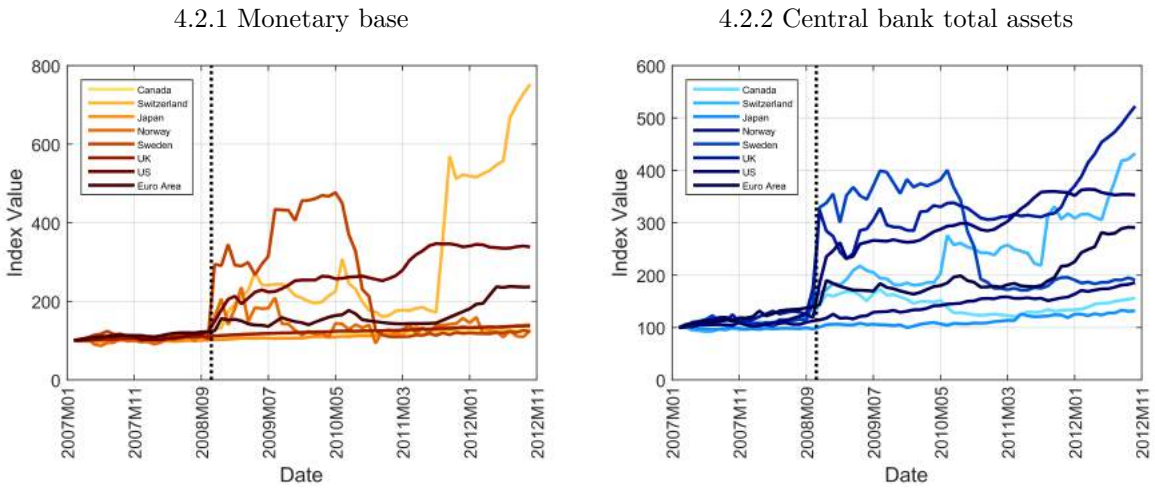
[1]Standard (policy orientated) structural vector autoregression (SVAR) models do not allow for a lower bound on the interest rate, making the empirical monetary SVAR model unsuitable for studying the effect of quantitative easing.

responses in an approach akin to that of Uhlig (2005) (an 'agnostic' sign restricted model), and the mean group and individual country responses from a panel vector auto-regression (PVAR) which is identified through a range of zero and sign restriction in the form of Gambacorta *et al.* (2014). It is the last of these studies which informs our sample selection and set of model specifications, whereby we expand their sample along the time dimension across eight developed economies which have employed quantitative easing measures in response to the financial crisis (US, UK, Japan, Canada, the euro area, Norway, Sweden and Switzerland), and we also use central bank total assets as the policy variable.

Our contributions are threefold. Firstly, we create a unique dataset for both monthly and quarterly housing variables across the crisis period, informed by pre-crisis studies such as Goodhart and Hofmann (2008) and Assenmacher-Wesche and Gerlach (2008) which consider the response of house prices to *conventional* monetary policy shocks in PVAR models. Secondly, we use this dataset (and variants of the baseline specification which is composed of economic activity, house prices, housing supply, a mortgage rate and an unconventional policy variable) to analyze the response of housing markets following an unconventional monetary policy shock. Finally, we attempt to show the variability across a range of commonly chosen econometric modelling choices, as mentioned above and discussed below in Section 4.4.

Figure 4.2 shows the time paths of central bank total assets and the monetary base. While there has been a strong cross-country correlation in the rise of central bank total assets as

Figure 4.2: Central bank responses to the crisis



Source: Central bank websites, dotted - collapse of Lehman Brothers Inc

a policy tool, with the balance sheets of the Federal Reserve and the European Central Bank tripling, the size of the balance sheet of the Bank of England rose more than fivefold. The central banks of smaller economies, such as the Swedish Riksbank, the Swiss National Bank, and, to a lesser extent, the Bank of Canada also expanded their balance sheets sharply after September 2008 (and such actions guide our sample selection in the remainder of the chapter). However, such policies have been predominantly heterogeneous in the ways they were undertaken. The transmission of such policies to housing markets is also likely to be strongly characterized by the institutional structure regarding how mortgage finance is provided. In general, the countries in our sample are those which have experienced faster and deeper innovations in mortgage markets, such as the United Kingdom, the United States, Canada and the Nordic countries, and tend to have a higher stock of mortgage debt as a ratio to GDP, compared to the international averages.

As argued for in both the theoretical and empirical literature (Section 4.2), countries with a larger stock of mortgages may be expected to experience stronger effects of monetary policy shocks. This will not only depend on the proportion of owner-occupiers, but the amount of debt outstanding per residence. Areas with a higher proportion of variable (as opposed to fixed) rate mortgages may be more sensitive, and variation in loan-to-value (LTV) ratios may cause a heterogeneous degree of pass-through, magnified by variations in mortgage/credit market efficiency. The remainder of this chapter is structured as follows. We next discuss the literature along four inter-related paths in Sections 4.2.1 to 4.2.4. Following this, we detail the intricate construction of our monthly and quarterly housing market datasets across eight OECD countries in Section 4.3. In Section 4.4 we discuss the various methods used to estimate and identify our models and construct and select our impulse responses, and in Section 4.5 we document our results. Finally, Section 4.6 concludes with policy implications and avenues for further research.

4.2 Existing literature

4.2.1 Housing in general equilibrium models

While not explicitly the focus here, a number of recent studies worth mentioning have emerged within a general equilibrium framework which emphasize the necessity of researching relation-

ships between monetary policy, house prices and other macroeconomic variables. Of particular interest is the work by Rubio (2014), which uses a two-country new Keynesian general equilibrium model with housing and collateral constraints to show that consumption reacts more strongly to a monetary policy shock in countries with high LTV ratios, a higher proportion of borrowers and a higher proportion of variable-rate mortgages. Aoki *et al.* (2004) consider a general equilibrium model with frictions in credit markets which are used by households, where houses provide services and themselves serve as collateral to reduce the cost of debt. Their result is that this effect amplifies and propagates the effect of monetary policy shocks on housing prices, housing investment and consumption.

4.2.2 Unconventional monetary policy shocks

In related work along the unconventional monetary policy dimension, Boeckx *et al.* (2014) also estimate the effects of exogenous innovations to central bank (ECB) total assets within a (single country) structural VAR framework with zero and sign restrictions, also identifying the same monetary policy shock as used in this chapter. They find that an expansion leads to a significant, albeit temporary rise in output and prices, with results similar to those found in the literature for conventional policy rate innovations. With interesting implications for our study, they found smaller effects in countries within the ECB where the banking system is less well-capitalized. In one of the first studies on the subject, Peersman (2011) estimates an SVAR model (with sign restrictions lasting up to four months) for the euro area between 1999M1-2009M12, also finding 'similar macroeconomic consequences' when comparing unconventional policy shocks with traditional interest rate innovations. As mentioned in Section 4.1, of particular inspiration is the paper by Gambacorta *et al.* (2014), where the authors estimate a PVAR with monthly data across the eight aforementioned countries using exogenous increases to the central bank balance sheets (total assets and monetary base) as policy shocks.[2] In accordance with the two previous aforementioned studies in this subsection, their conclusion is that unconventional policies have similar effects to policy rate shocks, albeit the impact on the price level is weaker and less persistent.

[2]We follow a similar identification strategy to this paper, and this is discussed in Section 4.4.

4.2.3 Single country VAR models of housing

The most relevant paper from a conceptual perspective is that of Gabriel and Lutz (2014), who use a structural FAVAR model (with factors extracted from daily US series) to show, amongst other things, that 'the estimated effects (on housing markets) are generally large in magnitude and similar in size to those found for equity markets'. A number of papers estimate single country SVAR models for housing markets. For the UK, the most relevant is Elbourne (2008), which takes a one-step approach to analyzing two questions using an eight-variable model. They find a predictable role for housing wealth in the transmission mechanism, albeit less than suggested by other authors. Aoki *et al.* (2002) estimate a small ('core five variable') single country SVAR for the UK to consider how house prices react to a 50 basis point short-term interest rate shock, concluding that the peak decline of house prices occurs after five quarters (0.8%).

Giuliodori (2005) considers monetary policy shocks across Europe by estimating nine individual country-specific VAR models over the period preceding the European Monetary Union. The identification scheme is standard in that the policy variable is ordered last, however, they also augment the baseline with mortgage market variables and exchange rates. The study finds evidence of a stronger credit channel in countries with more developed housing and mortgage markets. Vargas-Silva (2008b) estimates regional VAR models for the United States using the sign restriction methodology proposed by Uhlig (2005), comparing the results with a standard Cholesky decomposition, and Vargas-Silva (2008a) and Gupta and Kabundi (2010) use the two-step FAVAR methodology of Bernanke *et al.* (2005) to answer similar questions. It is this difference between the two identification procedures which is especially interesting, as within their sign restricted models, the response of housing market activity to a monetary policy shock is not only smaller, but also lasts for a shorter period.

Iacoviello (2002) estimates an SVAR for six countries. After considering the co-integrating relationships in the quarterly model, they show that countries with low transaction costs, high LTV ratios, a large owner-occupied ratio and a large proportion of variable-rate mortgages experience a higher sensitivity to monetary policy shocks. Using a similar methodology, Iacoviello and Minetti (2008) use short and long-run restrictions in four different VAR models across four countries, each time with a different variable such as the 'mix' (the fraction of housing loans by non-bank institutions), showing a clear relationship between the credit channel and the efficiency

of housing finance. Jarocinski and Smets (2008) show that while developments in U.S. GDP alone cannot explain developments in housing markets, adding a policy variable to the information set can. Finally, considering synthetic euro area data for the period 1980-1998, Smets and Peersman (2001) analyze the impulse responses from an area-wide shock to monetary policy on house prices, finding (as may have been expected due to the frictions present in housing markets) that they respond more sluggishly than share prices.

4.2.4 Panel VAR models of housing

Despite the extensive work that has gone into the subject, individual country VAR models tend to produce estimates and forecasts which can often be easily improved, both in terms of their precision and efficiency. Sá *et al.* (2011) consider not only the effect of monetary policy on housing booms, but also the role of capital in-flows and financial innovation combined with lax regulation. They use a PVAR with sign restrictions, splitting their sample according to the 2008 International Monetary Fund index of high and low mortgage market development. They find a result similar to that of Iacoviello (2002) - that economies with more developed housing and mortgage markets are more responsive to interest rate shocks. In the research most similar in methodology to the current study, Goodhart and Hofmann (2008) uses a quarterly fixed effects PVAR in the pre-crisis period for 17 countries to show a strong link between house prices and monetary variables and the economy, especially in a more recent sub-sample (1985-2006: after mortgage market deregulation following the highly variable inflation rates of the early 1980s). Augmenting their PVAR with dummy variables, they also find evidence that money and credit growth are useful indicators of house price booms.

Assenmacher-Wesche and Gerlach (2008) use a quarterly pre-crisis (1986-2006) PVAR to consider the effect of monetary policy on asset (housing and equity) prices across countries using the mean group estimator as proposed by Pesaran and Smith (1995). While the paper predictably finds a noticeable effect of monetary policy on asset prices, one of the main contributions of this chapter to the literature is to analyse spillover effects, showing the significance of US monetary policy shocks internationally (but not housing market shocks). Carstensen *et al.* (2009) use a pre-crisis (1996-2006) model split across European countries, dependent on

how pronounced the response of house prices is to a monetary policy shock. They find that where there is a stronger transmission of the shock through housing, there is also a stronger transmission to other macroeconomic variables.

4.3 Data

As indicated in Section 4.1, one of the key contributions of this chapter is the creation of two unique balanced panels of quarterly and monthly macroeconomic time series from 2007M01 to 2014M09. Their construction arises from the need to undertake research on housing markets in an international, cross-country context, and are subsequently used in the remainder of this chapter. Despite the fact that house prices are not always directly comparable across countries, due to both the definition of residential property and data collection methodologies, we attempt to harmonize the data used in estimation.[3] One key insight to appreciate is that in all of our quarterly and monthly specifications, we attempt to use equivalent representations. Where a series is available only in a naturally occurring monthly frequency, we convert using the third monthly observation of the quarter to construct a quarterly representation (as opposed to a summation across all three quarters, given that we only have 'stock' variables of this type, such as total assets and the monetary base). Where a series is available in only quarterly frequency, we obtain its monthly representation using the Chow and Lin (1971) procedure and one or two 'related series' as discussed below in Section 4.3.2.[4] The procedure is a common method for temporal disaggregation which uses the relationship between the low and higher frequency indicator variables. Where there is one indicator, this is a univariate, and where there are multiple indicators, a multivariate regression is used, where the regression coefficients are estimated at the low frequency level where both the 'target' and the 'indicator' are available. All series are adjusted using the X-13 ARIMA-SEATS Seasonal Adjustment algorithm. Our sample contains data on Canada (CA), Switzerland (CH), Japan (JP), Norway (NO), Sweden (SE), United Kingdom (UK), United States (US) and the euro area (XM), and where required,

[3] Unless otherwise specified, series are collected from Datastream, and a list of the appropriate identifiers is available on request from the author.

[4] A former version of the work also considered the Litterman (1983) procedure, and generated no difference in the results.

an index is re-based (2007M1=100).

4.3.1 Quarterly dataset

For all specifications containing 'GDP', we use gross domestic product at constant PPP from the OECD database. In Models Two and Three, we use the final monthly observation of retail sales and industrial production, discussed below. For our quarterly house price specifications, we use indices provided by Oxford Economics. Quarterly residential investment comes from OCED series on investment in private dwellings, measured in constant values - other than in the case of Norway (due to availability difficulties), which is a series for gross fixed capital formation for construction. For mortgage rates, we use data from Oxford Economics, apart from the euro area where we use an arithmetical average of all available European countries from the euro area (France, Italy, Spain, Germany and Ireland); Norway, where we use an average of mortgage companies (provided by Statistics Norway) and an average of different variable mortgage rates for Switzerland (provided by the Swiss National Bank). For central bank total assets and the monetary base, we use the final monthly (or seldom weekly) observation of the quarter, discussed below.

4.3.2 Monthly dataset

The monthly series for GDP are interpolated using both retail sales and industrial production as related series, common in the macroeconometrics literature. Industrial production series are volume indexes of total production excluding construction (from the OECD MEI series), apart from Switzerland, where no appropriate series could be found for the required frequency, so a quarterly series was created with cubic spline interpolation. Retail sales are measured as total retail sales in volume, also from the OECD MEI series. The series which are used to interpolate quarterly house price indexes are in the majority country specific indexes (which are usually constructed in different ways, motivating the interpolation of the standardized Oxford Economics series, as opposed to using a monthly series alone) such as the ONS National House Price Index in the UK or the Valueguard, HOX Housing index in Sweden. Residential investment is interpolated using a somewhat heterogeneous range of series such as housing starts or building

Figure 4.3: Quarterly and monthly house prices

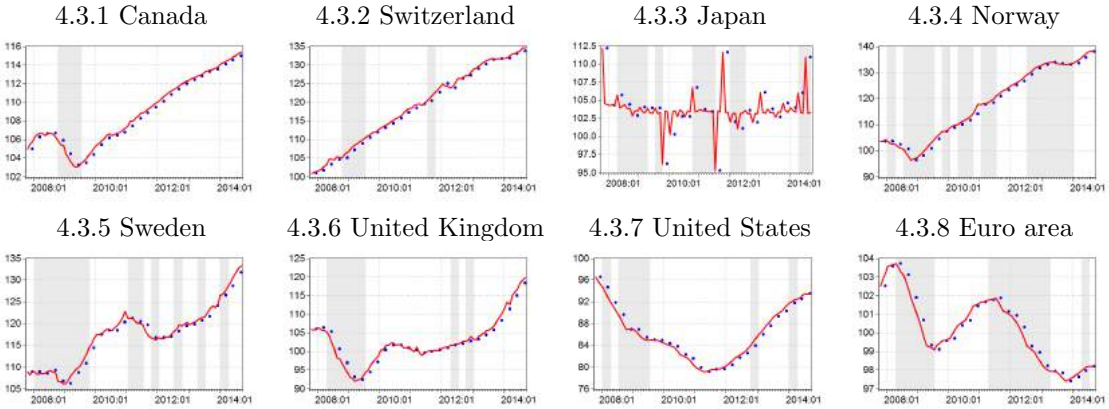


Figure 4.4: Quarterly and monthly residential investment

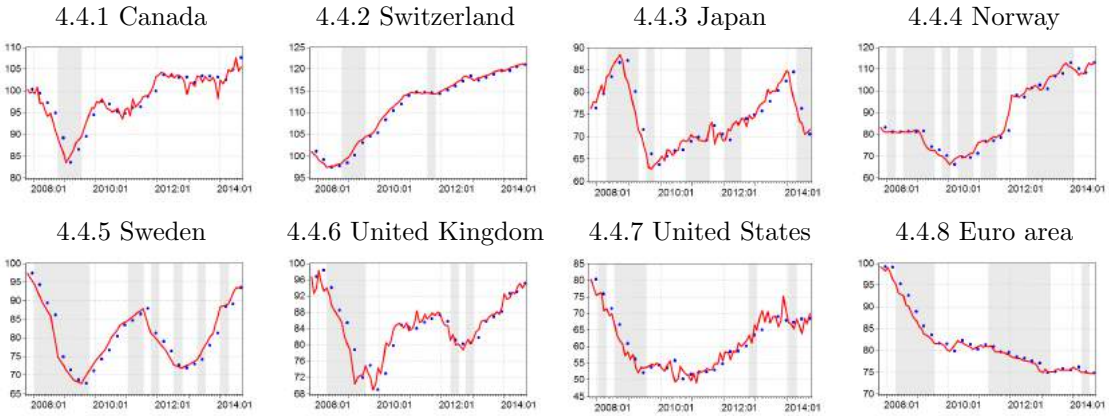
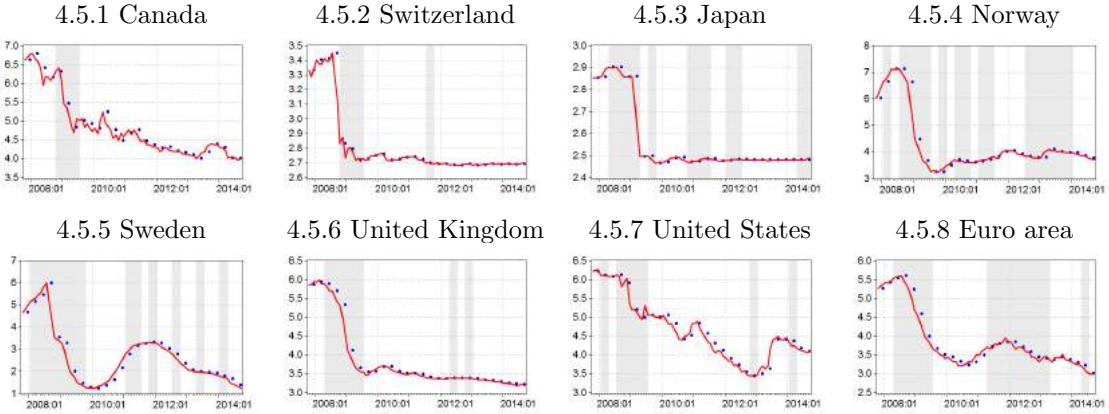


Figure 4.5: Quarterly and monthly mortgage rates



Shaded regions represent quarters of negative GDP growth specific to the country. Lines represent monthly data, symbols represent quarterly data.

permits granted (or for the UK, a volume index of building production, or for Switzerland, a monthly index on building composite materials imported into the country) from a range of national statistical offices, wholly dependent on availability. For monthly mortgage related series, we typically use private sector mortgage market indicators (such as the US conventional home fixed mortgage rate on 30 year products), apart from in Sweden, Japan and Norway, where no such series is available, so we instead use 10 year government bond yields to capture movements in capital markets. Of critical importance are our two policy variables, both of which are predominantly located on central bank websites and reconstructed from backdated balance sheets, using starting values from the dataset graciously made available by the authors of Gambacorta *et al.* (2014) as a check to ensure consistency. We end our sample in 2014M9 due to the discontinuation of the appropriate series by the Bank of England. In Figures 4.3 to 4.5 we show the close relation between our monthly and quarterly housing market series, and show that the interpolation procedure used generates entirely reasonable monthly series.

4.4 Methodology

Proposed as an alternative to traditional large-scale macroeconomic models, structural VARs have rapidly become the workhorse of macroeconomic research; in particular, studies which focus on the effects of an identified monetary policy shock, such as Faust (1998) and Uhlig (2005), are a primary focus of attention. We take a computational approach to the problem, comparing across eight countries, two data frequencies, five entirely reasonable competing model specifications, three common identification strategies and methods of summarizing sets of responses. Potentially most importantly, on the identification front, we follow the short-run restriction methodology which was broadly first outlined by Sims (1980), extending it for the 'agnostic' sign restriction approach of Uhlig (2005) and then an FGLS based panel variant as per Gambacorta *et al.* (2014) which utlizes a mixture of sign and zero restrictions.

4.4.1 Model specifications

We estimate five varying specifications each containing five variables revolving around a core housing market orientated[5] reduced form VAR model. Variations include changing the measure of economic activity between GDP, industrial production (IP) and retail sales (RS - as in Elbourne (2008)), changing our measure of monetary policy from central bank total assets (TA) to the monetary base (MB - as in Gambacorta *et al.* (2014)) and incorporating a trend to our baseline model. The variations are summarized in Table 4.1.

Table 4.1: Overview of the five model specifications

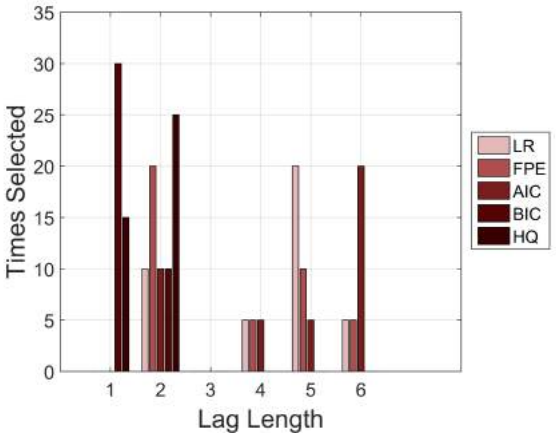
	GDP	RS	IP	HP	RInv	Mort	TA	MB	Trend	Constant
Model 1	✓	×	×	✓	✓	✓	✓	×	×	✓
Model 2	×	✓	×	✓	✓	✓	✓	×	×	✓
Model 3	×	×	✓	✓	✓	✓	✓	×	×	✓
Model 4	✓	×	×	✓	✓	✓	×	✓	×	✓
Model 5	✓	×	×	✓	✓	✓	✓	×	✓	✓

Even though the original study of Uhlig (2005) provides explicit rationale for not including a constant, we include one across all specifications and identification strategies for consistency. Table 4.1 also describes the ordering of the variables - an essential clarification of the zero restricted models - where the measure of economic activity is ordered first, followed by house prices, housing supply, mortgage rates, and finally, the policy variable is ordered last (to be discussed below). All of our VAR models described above are estimated through 2007M10/2007Q4 to 2014M9/2014Q3. Figure 4.6 shows the distribution of suggested lag lengths using three common criteria across each of our 40 reduced form models (eight countries, five specifications) for both frequencies. When setting the maximum lag length of six months for monthly data, and three months for quarterly data due to our relatively short samples, the BIC (chosen for its well known parsimonious qualities) suggests a lag of one on aggregate for both frequencies, and therefore, our eighty (eight countries, five specifications, two frequencies) reduced form VARs are modeled as VAR(1) models.

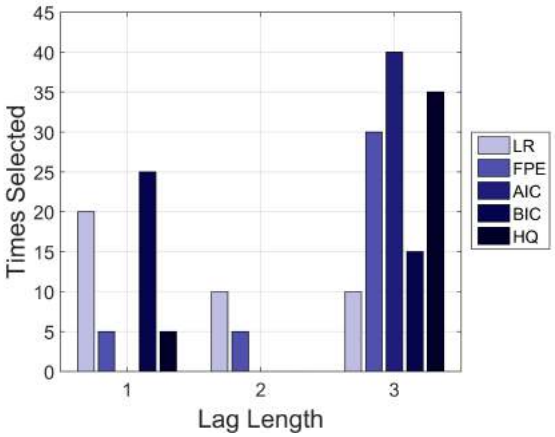
[5]For each specification we include house prices(HP), residential investment(RInv) and mortgage rates (Mort) as per the 'lens of fundamentals' of Gattini and Hiebert (2010) and other papers discussed in Section 4.2.3 and 4.2.4.

Figure 4.6: Lag length selection

4.6.1 Monthly data



4.6.2 Quarterly data



LR: Likelihood ratio, FPE: Final Prediction Error, AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion, HQ: Hannan-Quinn Information Criterion

4.4.2 Estimation and identification strategies

The use of different estimation strategies and identification schemes can alter results significantly, and we describe our methodology across the three main approaches in Section 4.4.2.1 to 4.4.2.3 below, with a brief discussion of the selection of impulse responses.

4.4.2.1 Identification: Cholesky

We begin with one of the most popular strategies used in the SVAR literature, where restrictions are placed on the contemporaneous effects of reduced form shocks (typically represented by matrix B_0 in the literature) - an approach dating back to well cited studies such as Sims (1980) (who calls it 'triangularizing the system'). In particular, we impose the necessary $\frac{m^2-m}{2}$ restrictions to exactly identify the system through a Cholesky decomposition of the variance-covariance matrix, making the assumption that B_0 is lower triangular, as in earlier housing market orientated SVAR papers such as Giuliodori (2005). Our ordering is as described in Table 4.1, and, critically, as per the majority of the earlier housing market orientated literature, we order the policy variable last. While it would be possible to impose a non-lower triangular specification of B_0 as in Elbourne (2008) (based on Kim and Roubini (2000)) as an alternative to a Cholesky decomposition, we retain the lower triangular structure as this is a popular comparison to the sign restriction methodology discussed below (as in papers such as Uhlig (2005)

and Vargas-Silva (2008b)). At this point, it is important to note that although standard unit root tests indicate some variables are non-stationary, we estimate the system in levels.[6]

4.4.2.2 Identification: Sign restrictions

Early pioneers proposed that structural inference using VARs might be based solely upon hypotheses about the signs of the impacts of some specific shocks in the system, and this approach has been adopted in numerous applied studies in various contexts thereafter (please see Fry and Pagan (2011) for a review of the various economic sub-disciplines that sign-restricted techniques have been utilized within). In particular, Uhlig (2005) introduced the key concept of the *impulse vector*, $a \in R^m$ (where m is the number of fundamental innovations), if and only if there is some matrix A such that $AA^\top = \Sigma$, so that a is a column of A (where Σ is the variance-covariance matrix of the VAR). His method considers the distribution of space across impulse vectors conditional on whether the response to these impulse vectors meets a given set of sign restrictions (which last for 'a few periods' - typically the medium run), and if it does, keep it. We define our monetary impulse vector as an impulse vector such that the impulse responses to a of the mortgage rates are not positive and that those for central bank total assets are not negative for six months (or two quarters, depending on frequency being used). Further discussion on the sign restrictions used for identification can be found in Section 4.4.2.3.

We generate impulse vectors and keep 1,000 draws which satisfy the restrictions[7]. In particular, we take the 'cleaner' *pure sign restriction* approach (based on a Bayesian method), assuming all impulse vectors satisfying the sign restrictions are equally likely. Given the insight of Fry and Pagan (2011) that it is problematic to structurally interpret the median of sign-restricted impulse responses, we implement their median target (MT) estimator method (which chooses responses from the model closest to the median response,) and report it in addition to the median and average impulse response.[8] Simple modifications to the program files would easily

[6] Sims *et al.* (1990) show that even if cointegration exists, the dynamics of the system are consistently estimated with a VAR in levels, whereas first differencing may discard relevant information. Regardless, we are prohibited from an explicit consideration of the long-run behavior due to our short sample, and we are keen to follow the informative work of Gambacorta *et al.* (2014) as closely as possibile.

[7] Unreported experiments to redraw 1,000 accepted responses show no qualitative difference (with differences occuring at the fourth or fifth decimal place only).

[8] Fry and Pagan (2011) succinctly summarize the situation, noting that sign restrictions are able to solve the parametric identification problem, but leave the model identification problem unresolved in that

allow the reporting of the 16th and 84th percentiles, as is common in the literature, although we leave them unreported due to the comparative nature of our approach. One notable benefit of using identification through this method as opposed to a pure Cholesky decomposition is that the results will be invariant to the ordering of the variables.

4.4.2.3 Identification: Zero and sign restrictions

As our final method of estimation and identification, we modify the method used in Gambacorta *et al.* (2014) discussed above, reporting the individual and mean group responses. Panel techniques allow a substantially more efficient estimate compared to country-specific SVAR models through the exploitation of the cross-sectional dimension. Estimation through the FGLS method of Zellner (1962) helps to capture information about common dynamic relationships which would have been lost by OLS estimation. In order to increase efficiency, each equation of the VAR is estimated for each country to take into account the correlation between the residuals of the same endogenous variables. We then bootstrap the system of equations having grouped each variable together across countries. As in Gambacorta *et al.* (2014), we use the mean group estimator of Pesaran and Smith (1995) as opposed to the fixed effects estimator, as it allows for cross country heterogeneity (required for comparison with the aforementioned approaches), without the imposition that the dynamics of each economy is the same.

To identify a structural re-parameterization of the system, we use a mix of zero and sign restrictions. While combinations of zero and sign restrictions have previously been used in papers such as Peersman (2011), Gambacorta *et al.* (2014) claim that the imposition of additional zero restrictions reduces the number of permissible models, 'sharpening' the identification. We start with a Cholesky decomposition of the country-specific covariance matrices, with a recursive ordering of variables as discussed above. We generate new estimated shocks (through the 'Q' matrices) to specific variables. It is the fact that this combination of shocks has an identical covariance matrix, yet a different impact upon our endogenous variables which gives us the 'ability to create a large number of candidate shocks that is the basis of sign restriction methods' (Fry and Pagan, 2011, p.6), and it is from here that we generate impulse responses which we use for our identification scheme (Equation 3 in Gambacorta *et al.* (2014)), based on economic

there are many structural VAR models to satisfy the sign restrictions.

theory. The matrix Q (an orthogonal matrix where $QQ^{\top} = I$ is partitioned in order to isolate monetary policy shocks (where Q is a function of θ and θ is a quasi-Bayesian draw from the range $[0, \pi]$), and responses of the fourth and fifth variables are treated as candidates for satisfying the sign restrictions.

Table 4.2: Identification of an unconventional monetary policy shock

Economic Activity	House Prices	Supply	Mortgage Rate	Total Assets
0	0	0	≤ 0	> 0

As in Section 4.4.2.1, we assume that policy shocks have only a lagged impact on residential supply, economic activity and house prices. The valuation, negotiation and conveyancing procedures involved in housing transactions justify such a restriction. However, as in Section 4.4.2.2, we also impose the restriction that mortgage rates must not increase in response to a positive policy shock to reflect the way that lenders may instantly adjust their lending rates following general economic theory and as in more structural papers such as Garriga *et al.* (2013). From this follows the 'complementary' restriction that total assets increase in response to the mortgage rate rising - a fact resembled in the reaction of central banks to mounting financial uncertainty. Utilizing the fact that the policy variable is the most endogenous of all, with policymakers reacting to information on all other variables in the system when determining how to optimally react, we order central bank total assets last; this gives us the ordering in Table 4.1 and sign restrictions as in Table 4.2 (where the restrictions again are imposed for six months or two quarters, as in Section 4.4.2.2). The imposition of the sign, as opposed to a zero restriction on mortgage market rate, allows monetary policy interventions to immediately influence mortgage markets. We argue that conditioning on the mortgage rate is able to 'disentangle exogenous innovations to central bank balance sheets from endogenous responses to financial market risk perceptions', as per (Gambacorta *et al.*, 2014, p.623), although we acknowledge it to be a weaker indicator than in their study. Again, if a draw does not satisfy the imposed restrictions, it is discarded, and we repeat until we have 1,000 satisfactory responses.

4.5 Results

4.5.1 Monthly data

A recurring theme across all plots is that the choice of identification and method of impulse response analysis cause some significant variation in the results. For house prices, this is seen no more so than in Japan, where the median target estimator predicts initial (zero period) increases or decreases by as much as 1.2%, a hundred basis points larger than the nearest competing response (although this MT response decreases after the initial period). Information on the peak of the mean group responses for all variables and specifications can be seen in Tables 4.3 and 4.5 (where the month or quarter of the peak of the response is shown in brackets), and the distribution of mean group responses themselves (for Model One) can be seen in Figures 4.10 and 4.14. We focus on these responses explicitly because not only do they give the most reasonable response from an intuitive perspective, but also because they are generated by the most rationalizable identification and specification schemes in general.

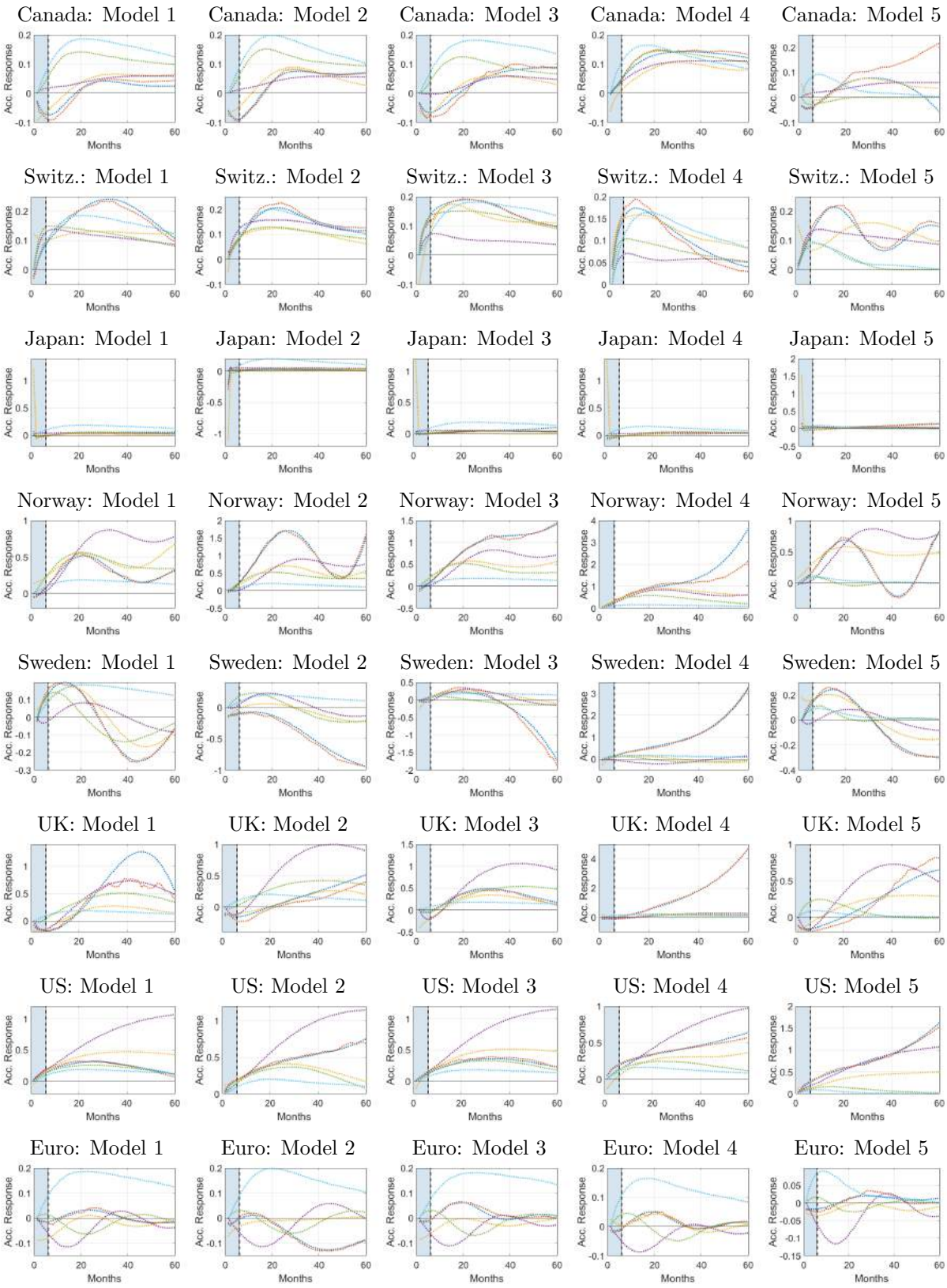
For monthly data, the results on the mean group response suggest that the response of house prices peaks at approximately 0.2% after between one and two years, and between 0.25% and 0.35% after approximately between two and three years in the case of residential investment[9]. The longer delay until the peak of the response in comparison to Gambacorta *et al.* (2014) is natural and expected, given the bureaucratic intricacies involved in the planning, financing and conveyancing of the construction of residential developments in the majority of our OECD countries. The first period is the most negative in magnitude for the response of mortgage rates (which are constrained to be negative for six periods under this identification scheme) which shows that following the shock, they never decrease further for any of the five models.

Table 4.3: The magnitiude and peak of mean group responses: Monthly data

	Model One	Model Two	Model Three	Model Four	Model Five
House Prices	0.186% (21)	0.199% (20)	0.181% (25)	0.165% (17)	0.092% (8)
Residential Investment	0.302% (30)	0.374% (28)	0.332% (32)	0.275% (26)	0.245% (15)
Mortgage Rates	-0.053% (1)	-0.057% (1)	-0.0549% (1)	-0.058% (1)	-0.0480% (1)

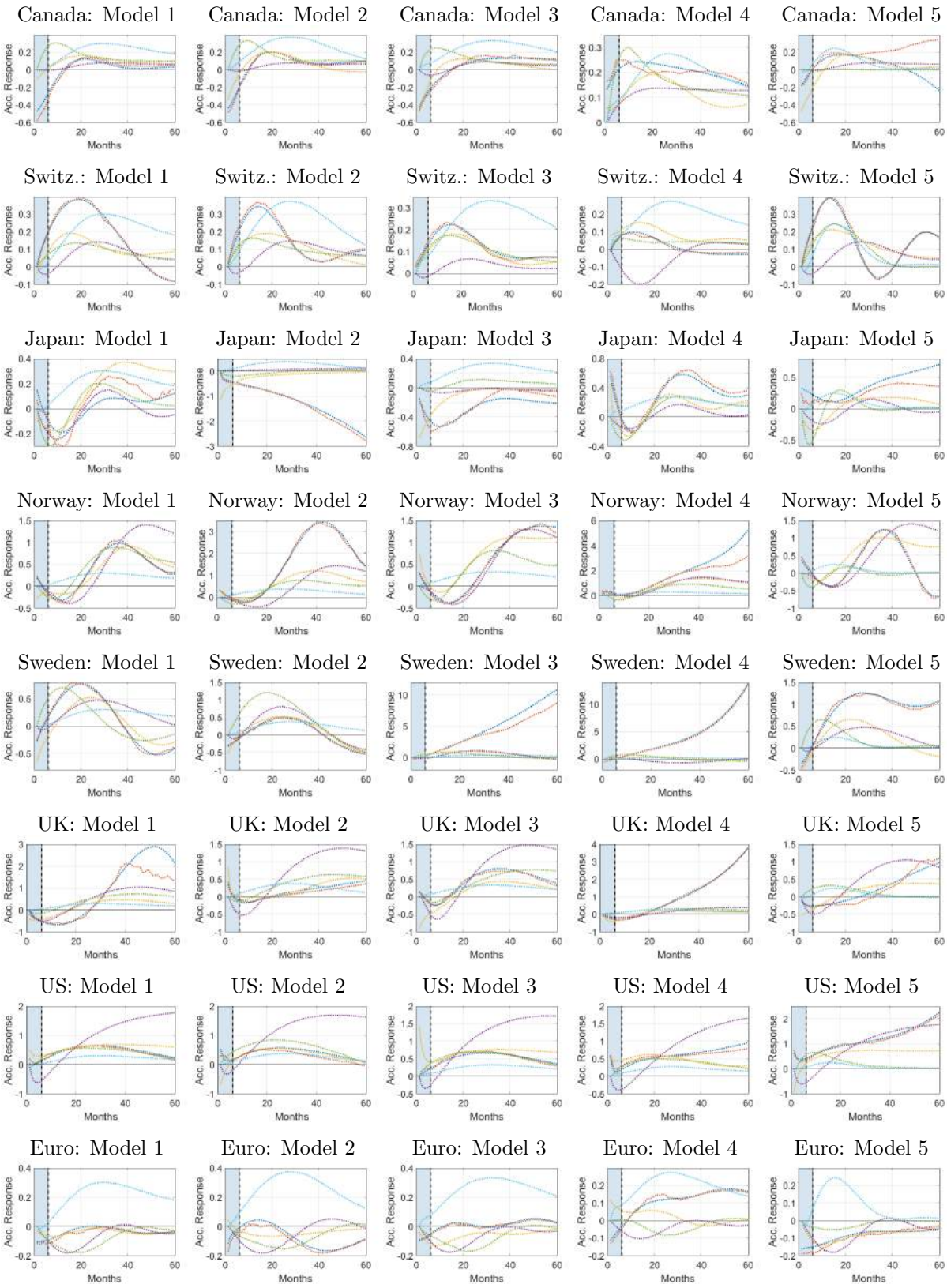
[9]Experiments with changing the lag length yield qualitatively indifferent results, whereby, for example considering Model One, house prices peak at 0.1735% after 19 periods for two lags, and 0.1667% at the 15th period for three lags.

Figure 4.7: Monthly house prices



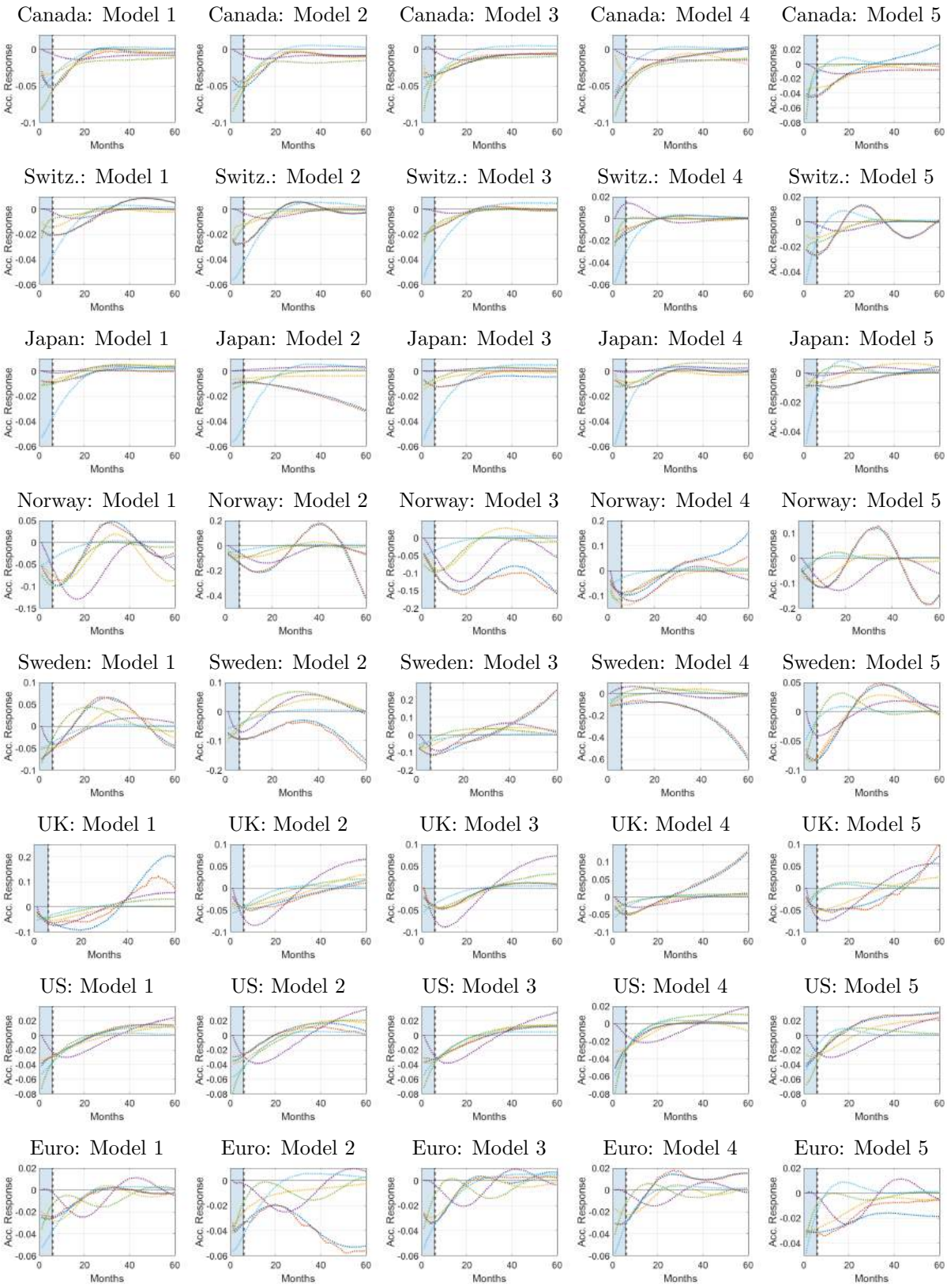
Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

Figure 4.8: Monthly residential investment



Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

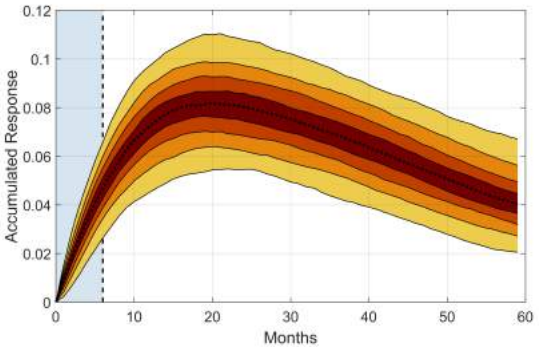
Figure 4.9: Monthly mortgage rates



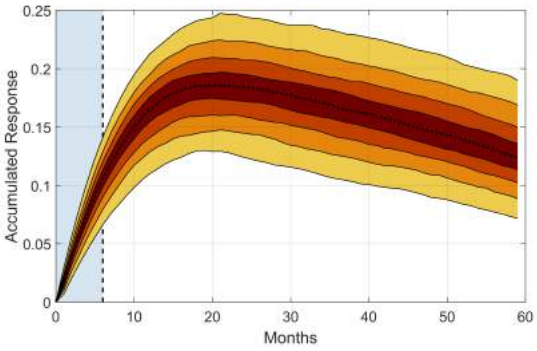
Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

Figure 4.10: Monthly mean group responses for Model One

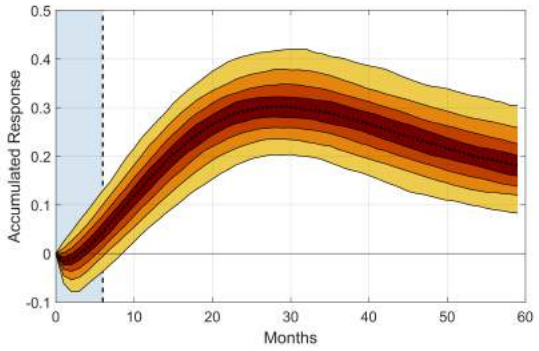
4.10.1 GDP



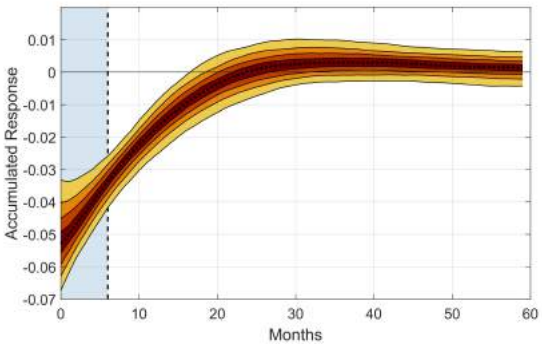
4.10.2 House prices



4.10.3 Residential investment



4.10.4 Mortgage rates



Firemap covers the eight deciles (80% probability bands), blue shade represents six periods of sign restrictions.

In general, as may be expected, the mean and median impulse responses from the pure sign restrictions models often provide almost identical results from a qualitative perspective. Comparing across model specifications, the results are broadly similar (across each of Figures 4.7 to 4.13), with some deviations occurring for Model Four - the specification where the policy variable is replaced with the monetary base. Considering just Model One (which might be considered our 'baseline' model), the monthly response of house prices is highest for the United Kingdom (where the response is in general strongly positive, although the response varies across method), and peaks after 46 periods for the sign restricted average (1.262%), 41 periods for the Cholesky decomposition (0.726%) and 39 periods for the sign restricted median (0.772%). The response of residential investment is more delayed also when considering the single country responses. For each of these five metrics, for this variable, at three months after the shock, 56% of our metrics report negative values compared with only 18% negative values after three months and three years. This is again attributed to the often protracted processes involved in

residential construction. The response of mortgages, while constrained to be negative through five of our six measures, is still negative at the (arbitrarily chosen) third period following a shock for 34 out of 40 (85%) instances (eight countries, five specifications) for our Cholesky decompositions, justifying our sign restriction identification scheme. For our three single country sign restricted response methods, it takes on average 7.26 periods for the first non-negative period in the response of the mortgage rate to occur, providing further validation. Table 4.4 shows the number of times (across all response steps, countries and models estimated on monthly data) one of the specific responses is largest in magnitude, indicating that the responses from the Uhlig (2005) identification approach are often smaller than both the Cholesky decomposition and the Gambacorta *et al.* (2014) based procedure when using monthly data. That the PVAR mean group response is repeatedly higher in both Table 4.4 and 4.6 is indicative of the fact that the response of the mean group is raised higher by the individual responses of the most responsive countries.

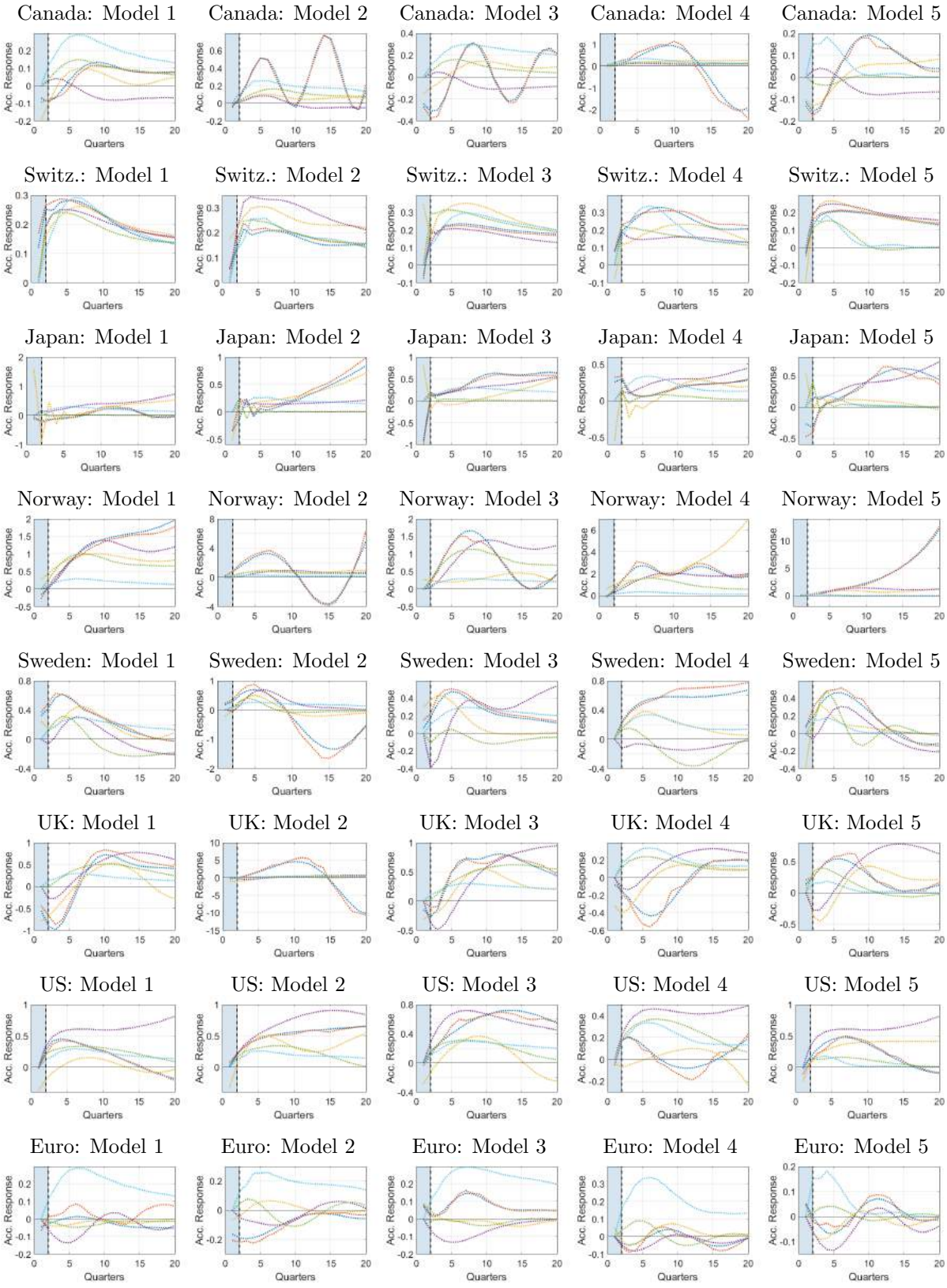
Table 4.4: Monthly data: Maximum absolute response of each element

	SR Average	SR Median	SR MT	Cholesky	PVAR Median	PVAR Mean
Model One	15.97%	14.10%	8.96%	21.53%	12.57%	26.88%
Model Two	9.86%	25.28%	2.64%	23.54%	10.97%	27.71%
Model Three	11.67%	16.46%	5.35%	24.31%	8.75%	33.47%
Model Four	20.35%	24.10%	7.50%	12.29%	10.90%	24.86%
Model Five	24.24%	30.00%	10.21%	23.40%	5.63%	6.53%

4.5.2 Quarterly data

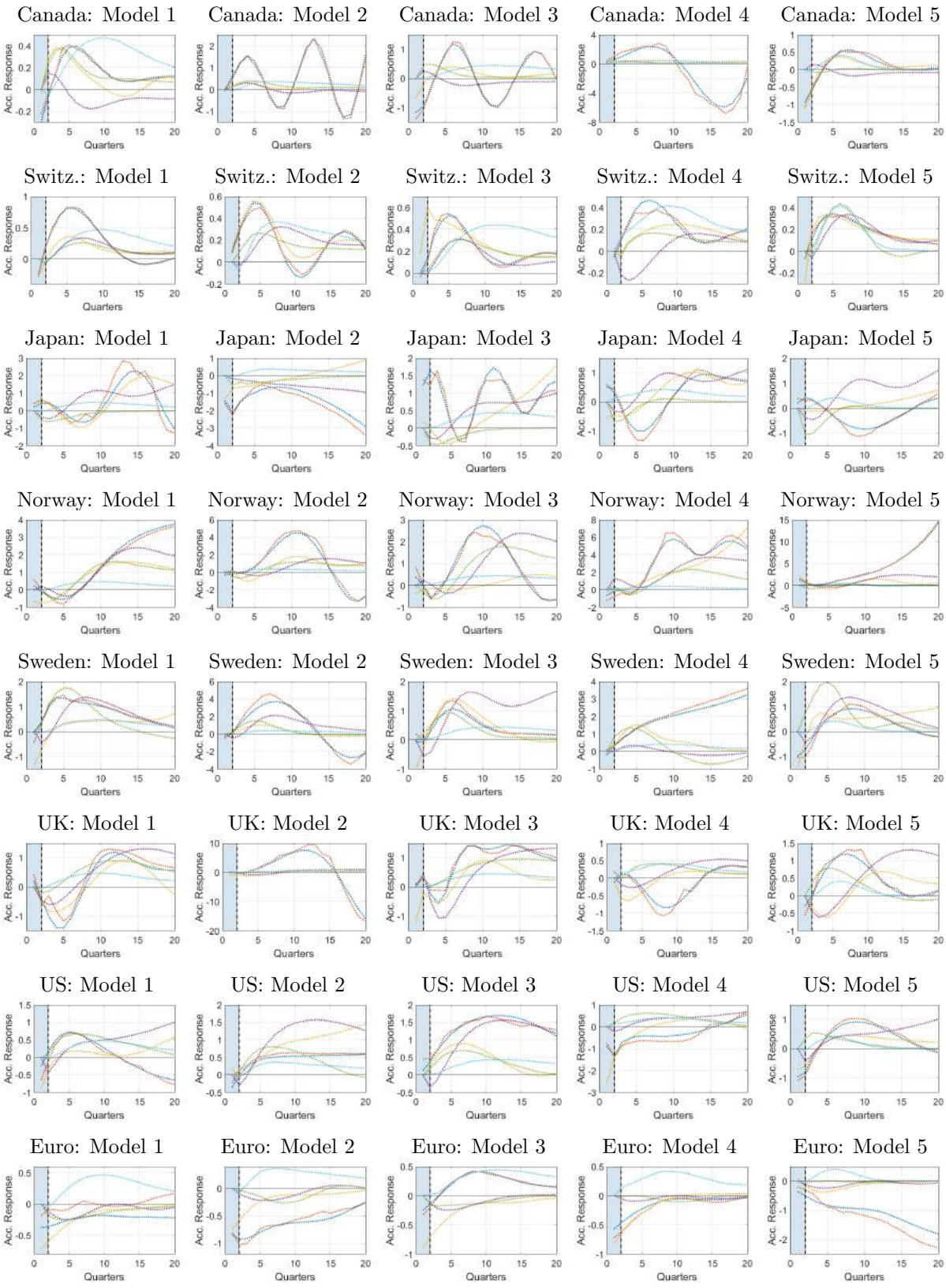
Turning to quarterly data, the mean group responses of house prices in the PVAR model (shown in Figure 4.14) are qualitatively similar in shape to the monthly results, albeit larger in absolute magnitude. Again, the peak response of the fifth specificaftion is lower (0.182%, compared with 0.092% above) than the first four models, which indicates a larger response of 0.257% and 0.337% at the peak, where the peaks are again between one and two years (six and seven quarters). The mean group response of residential investment is also larger in magnitude at the peak; whereas for monthly data the range was between 0.245% and 0.375%, it is now between 0.369% and 0.470% again with peaks occurring at a later point than any of those for house prices, a feature which also occurs with monthly data. To draw further comparison between the frequencies, for

Figure 4.11: Quarterly house prices



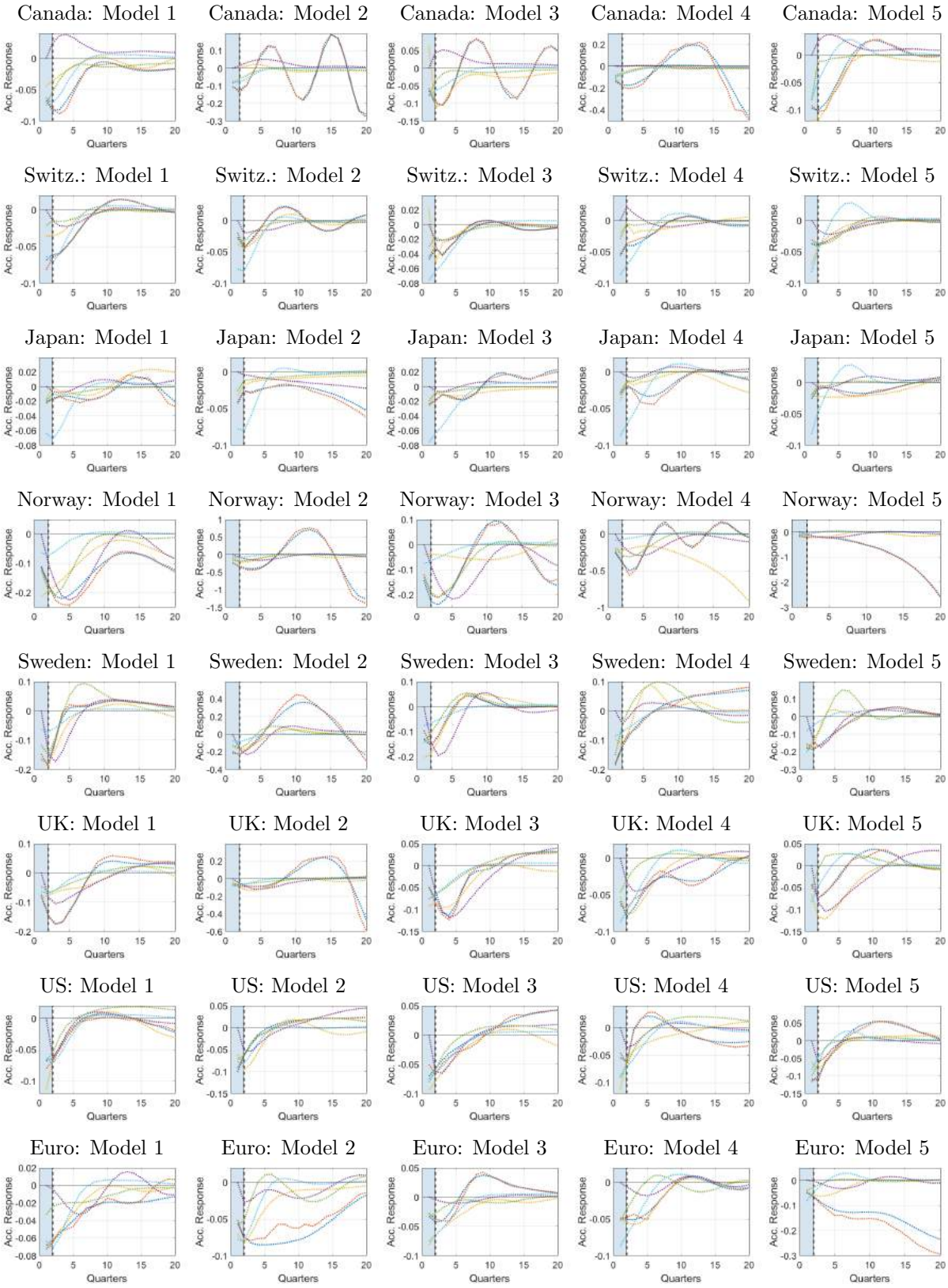
Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

Figure 4.12: Quarterly residential investment



Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

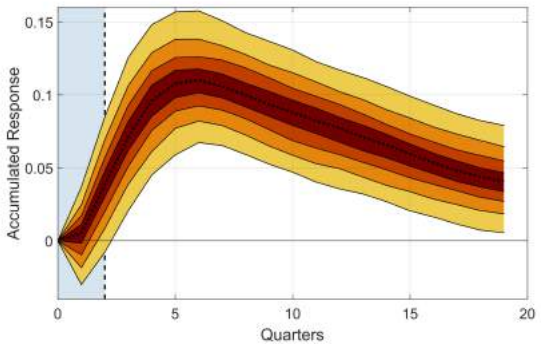
Figure 4.13: Quarterly mortgage rates



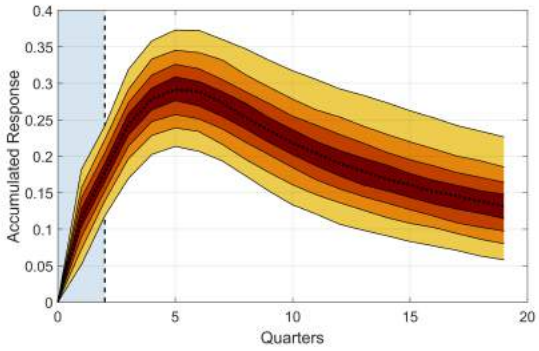
Dark blue line: sign restriction average, red: median target estimator, yellow: PVAR country specific median, purple: sign restricted median, green: Cholesky decomposition, light blue: PVAR mean response. Shaded: period for which sign restrictions bind. X-axis: horizon, Y-axis: accumulated response.

Figure 4.14: Quarterly mean group responses for Model One

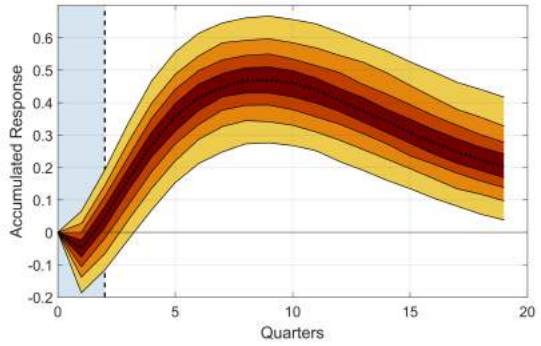
4.14.1 GDP



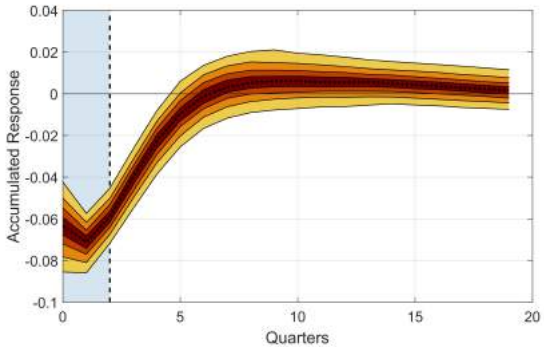
4.14.2 House prices



4.14.3 Residential investment



4.14.4 Mortgage rates



Firemap covers the eight deciles (80% probability bands), blue shade represents six periods of sign restrictions.

our four single country response metrics, the size of the largest absolute element (the 'peak') of the response vector is bigger for quarterly data 82.5% of the time.

Considering the individual country specific responses only, the largest response of house prices comes from the median and average response in Norway, despite some heterogeneity between specifications, coupled with the fact that not all responses return to zero. For all elements of all five individual country response methods, house prices display a positive response 69.125% of the time. Finally, considering Table 4.6, we see that the largest absolute responses come from the median method. On a comparative note, the response of house prices to a 'money

Table 4.5: The magnitiude and peak of mean group responses: Quarterly data

	Model One	Model Two	Model Three	Model Four	Model Five
House Prices	0.291% (6)	0.257% (6)	0.298% (7)	0.337% (6)	0.182% (4)
Residential Investment	0.470% (10)	0.369% (7)	0.438% (12)	0.413% (9)	0.417% (6)
Mortgage Rates	0.071% (2)	0.080% (2)	0.076% (1)	0.087% (1)	0.0820% (1)

shock' is in general smaller than in Goodhart and Hofmann (2008), which reports results for a sample split dependent on the rate of house price growth (Figure 7 in their paper), and is similar in magnitude, albeit slightly smaller (but necessarily opposite in sign) than Vargas-Silva (2008b), which shows a 0.5% decrease at the peak in response to a contractionary one standard deviation shock in conventional policy rates.

Table 4.6: Quarterly data: Maximum absolute response of each element

	Average	Median	Median Target	Cholesky	PVAR Median	PVAR Mean
Model One	14.79%	20.63%	13.75%	19.38%	10.63%	20.83%
Model Two	20.00%	42.92%	9.17%	16.46%	0.83%	10.63%
Model Three	23.13%	20.21%	13.13%	22.29%	3.96%	17.29%
Model Four	11.88%	36.25%	15.42%	16.25%	3.96%	16.25%
Model Five	10.42%	34.38%	13.96%	28.33%	6.46%	6.46%

4.6 Conclusion

Despite the heterogeneity in the results, the preferred specification and identification scheme (PVAR – Model One) provide strong evidence in favour of the intuitive economic theory that an unconventional monetary policy shock, in the form of an increase of central bank total assets, has an effect not only on house prices, but also on residential supply and mortgage markets. However, it must be noted that the varied undertaking of policy measures, combined with the heterogeneous institutional mortgage market structure, has contributed to a heterogeneous response of variables across countries. We have also provided evidence on the strong variation in results depending on the method of response analysis, frequency of data used, estimation and identification strategy and specification employed.

The policy implications from the results are threefold. The first is to recognize that when undertaking any form of monetary policy, conventional or otherwise, central banks must learn more about the precise ways in which their policies are transmitted into the wider economy. The second implication is that when estimating models of this kind, the underlying structure of institutions in place can drive any effects of unconventional monetary policy, and as such, requires explicit consideration, shown by the heterogeneous responses across countries. The third, and potentially most important implication is that results may substantially differ depending on the

model, data, identification and method of analysis used. Policy-makers must be increasingly aware of the need to conduct a robust analysis and examine the results of an entire, credible model set, as opposed to one model alone, as more credible models are developed. Further research in this sub-discipline may involve model averaging or other incorporation of likelihoods, or to incorporate further methods of response analysis, such as considering the joint distribution of admissible models as per Inoue and Kilian (2013).

Other avenues for potential research might focus on how housing markets are affected by different types of changes in the composition of central bank balance sheets. It would be an interesting advancement to incorporate methodology from the recent factor (and factor error correction) literature as suitable data for macro-panel construction becomes available in a wider range of economies, or to consider time-varying parameters. However, the main implication of this study is that more research is required to focus upon the transmission of unconventional policies to other assets, with a particular regard to real estate markets, in addition to the already considered bond yields, output and prices.

PREFACE TO CHAPTER 5

In this chapter we take a computational approach to forecasting a macroeconometric model of housing markets across six original datasets with large cross-sectional dimensions. We compare a large number of models which vary by the choice of factors, 'observable endogenous variables' and the number of lags in addition to classical and modern (factor based) specifications. We utilize various optimal model selection and model averaging techniques, comparing them against classical benchmarks. Within a 'pseudo real-time' out-of-sample forecasting context, results show that the approximate BMA method is the best weighting and selection technique, generating forecasts able to outperform the automated univariate benchmark of Hyndman and Khandakar (2008) upwards of 58% of the time. However, the average forecast error is lower in magnitude over all recursions and countries for the benchmark compared with all models selections and combinations for all variables. We also provide a tertiary examination of forecast error increasing as a function of the underlying variance of the series being forecast, as in Chapter 2.

This chapter focuses might be described by Frank Diebold (Diebold (2012)) and others as a classic 'horse race' scenario. We revolve our baseline model around one similar in composition to that seen in Chapter 4, and we combine two competing approaches (factor extraction and model combination) to high dimensional forecasting as discussed in the preface to Chapter 2. Ultimately, we find similar results to Castle *et al.* (2013): that in some races, it is the standard AR workhorse model which is fastest.

CHAPTER 5

HOUSING MARKET FORECASTS WITH FACTOR COMBINATIONS

In this paper we take a computational approach to forecasting a macroeconometric model of housing markets across six original datasets with large cross-sectional dimensions. We compare a large number of models which vary by the choice of factors, 'observable endogenous variables' and the number of lags in addition to classical and modern (factor based) specifications. We utilize various optimal model selection and model averaging techniques, comparing them against classical benchmarks. Within a 'pseudo real-time' out of sample forecasting context, results show that the approximate BMA method is the best weighting and selection technique, generating forecasts able to outperform the automated univariate benchmark of Hyndman and Khandakar (2008) upwards of 58% of the time. However, the average forecast error is lower in magnitude over all recursions and countries for the benchmark compared with all models selections and combinations for all variables. We also provide a tertiary examination of forecast error increasing as a function of the underlying variance of the series being forecast, as in Chapter 2.

Keywords: House Prices, Forecasting, Factor Error Correction Models, FAVARs

JEL Classification: C53, C55, R30

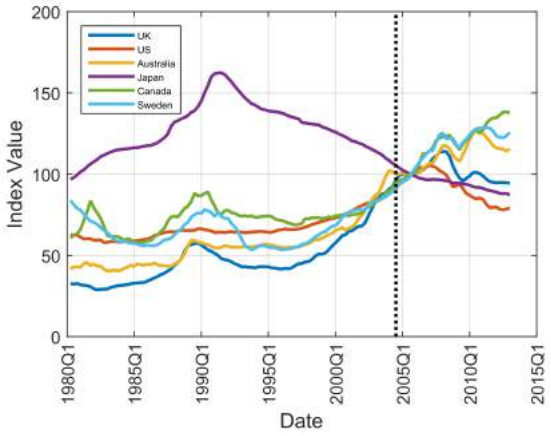
5.1 Introduction

Considering the recently verified importance of housing markets on an institutional, international scale, it may be thought of as surprising that such little space in academic journals has been devoted to forecasting housing markets. In this paper we combine various strands of the econometric, forecasting and real estate literature. In particular, vector autoregressive (VAR) models have been a predominant feature of modern forecasting since their inception during the 1980's, and our approach overcomes the deficit of most classical empirical analyses which only focus upon a small number of variables. That deficit is that the 'classical' VAR model typically includes between three and ten variables, potentially creating an omitted variable bias, with serious consequences for forecasting. Even when considering an approach which overcomes this, then the choice of which combination of either variables or factors to combine is itself based upon subjective empirical methodologies, and this is something which we attempt to overcome in our approach.

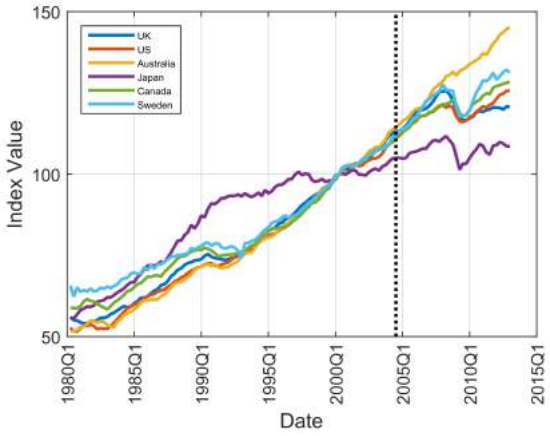
While the integration of large datasets into the econometric profession has been growing due to increased computational power and improved algorithmic efficiency, such analysis is an 'exception rather than the rule' (Bernanke *et al.*, 2005, p.1), despite factor-based approaches appearing to consistently generate more accurate forecasts through their use of large cross-sections. However, when confronted with uncertainty over which model to use in a forecasting context, pooling over all forecasts may seem like a natural way to deal with the issue. An increasing amount of literature on model averaging shows favorable results compared with a single econometric specification. We utilize such techniques to forecast four housing market variables which have recently been shown to influence consumer spending, financial markets and the macroeconomy more generally. We do this across a range of major developed economies, in contrast to previous studies primarily focused solely upon either the U.S. aggregate, or a regional U.S. housing market with a small model set (see Section 5.2). Unlike most other studies, our sample period includes the recent crisis period. Figure 5.1 shows time plots of the variables across six markets to be forecast, and from this, we can observe the heterogeneous nature of our exercise: pseudo real-time forecasting a range of different series during periods of increased volatility (where the out of sample begins in 2004:Q4-2013:Q3).

Figure 5.1: Time plots of the four variables to be forecast

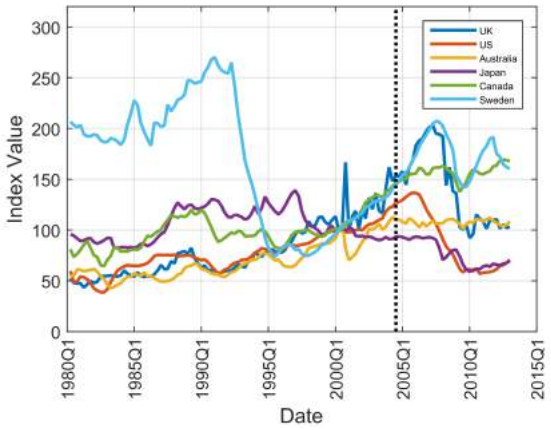
5.1.1 Real House Prices



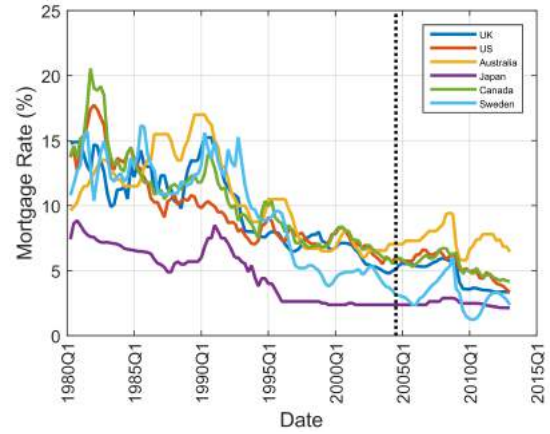
5.1.2 Real Residential Investment



5.1.3 Real Income



5.1.4 Mortgage Rates



Source: Shaded regions represent NBER recessions, vertical black dash represents final quarter of our recursive in-sample window.

In particular, the methods used in this paper provide a way of incorporating a large number of available time series, reconciling the growing availability of data while keeping the framework itself empirically tractable. The restricted and unrestricted 'large-dimensional approximate' factor models which we implement permit the idiosyncratic errors to be weakly serially and cross-sectionally correlated with large cross-sectional (N) and temporal (T) dimensions. We model the co-variability of the series in our large datasets in terms of a small number of latent factors. We then consider the relationship between these factors and our 'observed' variables. Of particular interest, we consider recent developments in the forecasting literature, estimating a large number of Factor Error Correction Model (FECM\FECMc), as proposed in Banerjee and Marcellino (2008) and implemented in Banerjee *et al.* (2014), as well as the 'two-step' variety

of unrestricted factor augmented vector autoregressive models (FAVAR) as used by Bernanke *et al.* (2005). Stationary I(0) factors are extracted as per Stock and Watson (2002a), and non-stationary I(1) factors are extracted using the technique of Bai (2004), as is common in the existing factor literature.

Particular inspiration for the model-averaging routines in this paper comes from Bai and Ng (2008), Bai and Ng (2009) and Cheng and Hansen (2015) who discuss the situation whereby the conventional ordering of factors by their importance in driving the co-variability of many predictors may not be consistent with the forecasting power for a specific series of interest. As a result, we employ model specification techniques[1] to determine which combination of factors (and which lag lengths) should be used in the models in addition to various weightings[2] of forecasts from the set of candidate models. Until recently, the effect of generated regressors on model selection and combination had not been previously investigated, in contrast to inference on the coefficients, where studies as early as Pagan (1984) show that generated regressors affect the sampling distribution. Importantly, Cheng and Hansen (2015) show, in a contrast to other results in the literature, that the generated regressor issue can be safely ignored when considering factor combination, without restrictions on N and T.

Given only the recent advent of factors and their applicability to modern macroeconometrics, and while there has been much research which considers factor models alone, there have been a relatively small number of studies which analyze the forecast performance of models which jointly consider both factors and variables. There are even fewer papers which provide empirical support for such a joint inclusion (Castle *et al.*, 2013, p.17), motivating our study of the 'small macroeconometric model'. This may appear surprising, considering the intuition that factor augmented models are able to overcome one of the primary criticisms of the classical VAR approach to macroeconomics, in that factors are able to conserve degrees of freedom while being able to span huge information sets typically available to practitioners.

In particular, early notable work along this dimension (Bernanke and Boivin (2003)) had considerable success with factors in improving the estimation of the policy reaction functions of the Federal Reserve. Such a modeling practice is entirely rationalizable given the assumption

[1] In particular: the IPC_2 for determining the appropriate number of factors, and then the SIC, AIC, sum of squared errors and $\bar{R}^2$ for determining the number of lags and predictors from our model set.

[2] We employ approximate BIC and a uniform model average method.

that central banks face information constraints similar to econometricians, who may be unable to directly observe various economic phenomena. In particular, the problem lies in the omission of key variables - a problem which has for example been used to explain anomalies such as the price puzzle. However, other macroeconomic applications such as Banerjee *et al.* (2005), which compares single indicators with factor models, do not find improvements from factors for euro aggregates, and, using similar methodologies, nor do Schumacher and Dreger (2004).

While there is existing research which considers the performance of either factor or classical macroeconometric models, very few consider both in the context of housing forecasts, or combine and weight in a combinatorial approach. However, by combining a large number of variables condensed into factors as well as explicitly incorporating key 'observed' variables decided upon by theory in a cointegrating (i.e. - FECM) context is an especially promising avenue for housing market forecasting given the characteristics of the data generating processes at hand. Following the discussion of the literature in Section 5.2, we then briefly discuss the large datasets used for factor extraction[3] in Section 5.3. Then, in Section 5.4 we outline the methodology of the model set to be estimated, as well as providing an outline of our factor extraction, model selection and model averaging techniques. Following this, we present all of the results in Section 5.5, and finally, Section 5.6 concludes.

5.2 Existing literature

While the literature in the broader subject area is expansive, we attempt to give a short précis of the studies relevant to our approach. Early studies include Gau (1984) and Hamilton and Schwab (1985), which take a forecasting approach to determining market efficiency using data on commercial income generating markets in Vancouver and national residential markets in the United States respectively. However, these studies are typically constrained by small sample sizes and relatively outdated methodologies. The classic 'bad forecast' is that of Mankiw and Weil (1989), who forecast a real house price decline of 47% based upon the aging of the baby boom ('baby bust') generation and a falling demand for housing. While this appears to be more

[3]An optional Appendix accompanies this paper which provides additional information on the series for factor extraction, as is conventional in this literature, and is available on request.

of an exercise in forecasting demographics, human capital and economic prosperity, the most direct critique comes from Hendershott (1991), who claims that their equation only fits data from the 1950's and 1960's, and that even between 1970-1987, the in-sample forecast error is unacceptably large. In a final example of the early literature (a paper which is subsequently explicitly mentioned as the motivation for several other academic works on the subject), Case and Shiller (1990) show how markets are forecastable using a repeat sales transaction index over a short time horizon due to serial dependence, citing an inability to make excess returns because of high transaction costs.

Patrick *et al.* (2000) compare ARIMA and spectral methods and their ability to forecast the turning points of US, UK and Australian securitized property markets. The authors find that while exponential models perform best during periods incorporating a steady trend, both ARIMA and spectral regression are capable of predicting turning points. In related work, Crawford and Fratantoni (2003) consider ARIMA, GARCH and regime switching models in five MSAs between 1979:Q1-2001:Q4, with out-of-sample forecasts generated over a range of different periods. The paper explicitly rationalizes the decision to consider only univariate methods due to the inherent difficulties in obtaining regional data. The authors conclude that while regime switching models perform best in-sample, with the ability to categorize boom and bust cycles, they only performed best (minimum RMSE) in three of the fifteen out-of-sample forecasts. However, the performance of the switching model in this paper is severely restricted by the fact that the authors only estimate it as a simple autoregressive specification due to concerns about the number of estimable parameters involved.

One particular paper of interest is that of Guirguis *et al.* (2004). The authors examine and compare six (VECM, AR, GARCH, Kalman Filter random walk, Kalman Filter autoregressive process, and exponential smoothing) models for the U.S housing market which are estimated between 1975-1985, with forecasts generated from 1985-1998. The comparison of the out-of-sample forecasts indicates that those generated by the Kalman Filter with an AR representation and the rolling GARCH models outperform all other specifications considered. However, the validity of the forecasting exercise is called into question through the claim that they 'include the error correction term calculated from the whole sample period extending from 1975:01 to 1997:2' (p.45), granting the model more information than would have been available at the time

of a hypothetical ex-ante out of sample forecast.

Turning to models which consider a small number of explanatory variables for housing markets (which subsequently inform our choice of 'observed' variables), Brown *et al.* (1997) apply the Kalman Filter to U.K. aggregate house prices over a boom period (1989:Q3-1992:Q2), having assumed that housing markets are structurally unstable. Such an approach allows two of the coefficients (nominal user cost of housing and expected gain on housing) to vary with time. Both static and dynamic forecasts are generated and compared against three fixed parameter models, which it outperforms, providing initial support for time varying parameter models in a house price forecasting context. In a similar study, Hall *et al.* (1997) use the Markov-switching regression proposed by Goldfeld and Quandt (1973) which includes four variables (real house prices, real income, owner occupied stock and mortgage rate). The system is a two state Markov process (stable and unstable) with transition probabilities dependent on how far out of equilibrium the system is. The model is successful in identifying the boom periods of UK housing of 1971-1974, 1977-1979 and 1986-1989, confirming parameter instability due to economic fluctuations and the degree of deviation from a long-run equilibrium. Using a similar set of dependents, Gattini and Hiebert (2010) include euro area house prices, housing investment, real disposable income per capita and a mixed maturity measure of the real interest rate. The variable choice is entirely rationalization given economic theory and various potential specifications of the data generating process for house prices. The authors find that 'the reduced form model tracks closely turning points in house prices when examining out-of-sample one and two-step ahead forecasts' (Gattini and Hiebert, 2010, p.47). While the paper presents results robust to specification tests, it only considers a single model with one specification for one economic area, considering only one and two step (quarters) ahead forecasts.

Rapach and Strauss (2007) consider state level forecasts with 25 potential predictors within an autoregressive distributed lag (ARDL) framework. The authors find that useful information is contained in the state level housing price/income ratio, unemployment, and national inflation rate. Various forecast combination methods are also considered with some degree of success. Rapach and Strauss (2009) follows in much the same trail, finding that across the twenty largest U.S. states, different housing markets can be better predicted with different information sets. In the first of the 'large dataset' type house price forecast papers which we consider, Gupta

(2013) uses 145 (monthly) variables within a dynamic factor and Bayesian shrinkage framework to forecast house price growth for the four census regions of the U.S. and the aggregate U.S. market as a whole. Using the standard Minnesota prior as well as additional priors which restrict the sum of the coefficients, the paper compares forecasts over an out-of-sample horizon of 1995:M1-2009:M3 based on an in-sample of 1968:M2-1994:M12 against a classical random walk and VAR model. The paper concludes that one of the factor models outperform in at least four of the five specifications. In the same vein, Gupta *et al.* (2011b) implement several Bayesian and classical models for 20 U.S. states, comparing Bayesian shrinkage and FAVAR models (with spatial or causality priors), showing that factor-augmented models outperform.

Exploring the subject in even more detail, Das *et al.* (2010) develops FAVAR and large BVAR models using a dataset of 126 monthly series for forecasting annualized real house price growth rates across nine census divisions of the United States. The results again indicate a preference for not only factor models, but show that large-scale models outperform their smaller counterparts - a result we are especially interested in. Considering potentially the most complete of all the studies on the subject, Gupta *et al.* (2011a) consider the US real house price index (and its turning point in 2006:Q2) using not only large BVAR and FAVARs, but also factor augmented AR, Bayesian-shrinkage models and the DSGE model from Iacoviello and Neri (2010). The study concludes that while the small-scale Bayesian version of the model outperforms other models, it is only the DSGE model which is able to forecast the turning point with any accuracy.

Das *et al.* (2009) consider forecasting the recent downturn in the South African housing market. Using 268 quarterly series to forecast small, medium and large-segment housing, the authors again compare large BVAR, classical and Bayesian versions of univariate and multivariate (vector) autoregression models with a DFM model. Based on the RMSE, the result is that the large BVAR not only outperforms, but also manages to predict the downturn in the middle-segment. Turning to the metropolitan areas of South Africa, Gupta and Das (2008) estimate a spatial BVAR based on First-Order Spatial Contiguity and Random Walk Averaging priors. The results in this study show that the larger BVAR models outperform for all markets except that for large scale dwellings.

Bork and Møller (2012) use a common factor approach based on 122 quarterly time series for U.S. house prices. They find that a three-factor model can not only explain about 50% of one-

quarter in-sample forecasting regressions but is also able to consistently outperform their AR benchmark at all horizons. The authors also document the substantial out-of-sample predictive power by comparing their model against AR benchmarks and computationally intensive forecast combination methods.

Moving away from the realm of house price forecasting, of particular interest to our work at hand is Banerjee and Marcellino (2006), who compare the augmented (with leading indicators) univariate approach of Cecchetti *et al.* (2000), the automated model selection procedure developed by Hendry and Doornik (1999), and a dynamic factor model based upon the factor extraction technique of Stock and Watson (2002a). The paper compares the performance of models which include factors against those which use a combination of leading indicators against a single leading indicator which are then themselves compared against a standard autoregressive benchmark, finding that factor models are 'on the whole disappointing' for forecasting performance, in comparison to the result by Stock and Watson (2002b). Interestingly, Banerjee *et al.* (2005) uses a similar model, finding factor models perform differently depending on the underlying datasets, with results varying across country. A common finding is that gains from factor models tend to be larger for nominal series (see, for example, Bernanke *et al.* (2005)).

In particular, some authors such as Boivin and Ng (2006) argue that it may not be optimal to always use all available data when constructing principle components. However, Bernanke and Boivin (2003) find that the forecast performance of their model is improved when the dataset is enlarged from 78 to 215 variables. In addition to this, the standard approach calculates factors independent of the target variable. For example, Bai and Ng (2008) calculate factors (using the quadratic principal components method) from 'targeted predictors' (based on soft-thresholding). Their main finding is that for forecasting inflation over all horizons, there are gains from using fewer, but more informative predictors, commenting that holding the set of predictors fixed may be unnecessarily restrictive.

One study of particular importance is that of Eickmeier and Ziegler (2008), who conduct a meta-study of 46 forecasting exercises involving dynamic factor models which are compared to a benchmark. The merit in this study comes from the fact that it is often difficult to draw conclusions across methods due to differences in data and implementation (e.g. different benchmarks). The authors find mixed results for factor model forecasts, with out-performance

in some instances (e.g. across variables, or economic region) but not others. Studies such as this reinforce the importance of taking a robust approach to the data: estimating several models for many variables in a cross-country setting.

Forecast combination methods have increasingly been used (since the seminal article of Bates and Granger (1969)) to improve forecasting performance in more conventional applications such as GDP and currency markets, but also with meteorological data, wilderness area use and the forecasting of football games. Clemen (1989) provides an extensive review of the early combination literature featuring over 200 items ('an explosion in the number of articles on the combination of forecasts'), and a thorough over-view of this literature is far beyond the scope of this paper. Hendry and Clements (2004) provide various rationales for the success of forecast pooling (alternatively known as model averaging in a forecast context). Their arguments are predicated on the notion that if individual forecasts are based upon different subsets of information, or if they are differently biased, a combination may improve the accuracy, or offset the impact of structural breaks (through a reduction of the variance). Considering specific empirical successes from a policy making point of view, papers such as Kapetanios *et al.* (2008a) and Kapetanios *et al.* (2008b) indicate that while individual central bank models have difficulty in beating a simple benchmark AR model, combined forecasts frequently outperform. However, one commonly held notion is that of the 'forecast combination puzzle', whereby simple combinations ($\frac{1}{N}$) of point forecasts are repeatedly found to do best, which we subsequently examine in Section 5.4 as one of our model weighting procedures.

5.3 Data

We utilize six new large datasets (dependent on availability), containing data on the United Kingdom, the United States, Australia, Japan, Canada and Sweden. In order to keep our 'observed' endogenous variables as consistent as possible, we utilize real house prices from Mack and Martínez-García (2013). In addition to this, we take real gross domestic production and interest rates on mortgages from Oxford Economics, and real residential investment from the OECD. For all datasets, $N > T$, ranging from $N = 162$ (Sweden) to $N = 324$ (US). All data used for factor extraction (which excludes our 'observable' variables) come from a range of national

and international sources, such as central banks, official statistical agencies and other third party data providers, dependent upon availability. All series were collected from Datastream, and an optional appendix containing the applicable Datastream identifiers is available upon request.

There is presently no definitive guide in the literature as to what data should go into factor analysis. The macroeconomic panels are 'put together' by the researcher, and the factors are inherently sample dependent by construction. However, the structure of our datasets follow the existing literature and are roughly categorized into groups such as monetary, banking, prices, foreign exchange rates, government, consumption, investment and other variables directly relevant to house price movements. While econometric theory might postulate that the size of the cross-sectional dimension positively influences the quality of the forecasts, data quality/reliability (i.e. - source) remains a primary consideration.

The data is automatically transformed using a variant of the MacroTrans procedure described in Appendix C. This procedure uses the X-13 algorithm from the Census Bureau (not making any transformation if no seasonality is detected). Having checked for NA observations, it then takes logarithms if possible (no non-positive elements of any series in any dataset), unless the series is a 'rate' (all elements under 100 in magnitude). The series are then standardized (subtract by mean, divide by standard deviation). For the I(0) factors (discussed below), all series are either all first differenced (if in logarithms), or a percentage change if not in logarithms. While the use of 'real-time' data is increasingly accounted for in papers such as Clark and McCracken (2010) and Faust and Wright (2009), it remains the exception rather than the norm. While other research utilizes sources such as the ALFRED real-time database, international availability of real-time sources remains limited, and we prefer to instead utilize a 'pseudo-real time' forecasting exercise in order to exploit a cross-country dimension.

5.3.1 Number of factors

The plots of all estimated factors can be seen in Figures 5.2 and 5.3. One essential modelling decision is regarding the number of factors to include in a model, or the maximum permitted within a determination routine such as that of Bai and Ng (2002). We choose a maximum of four Bai (2004) style factors (which are supplemented with a cyclical factor), and a maximum of

five Stock and Watson (2002a) style factors. There are a multitude of reasons for these choices. First, our choices relate to and are exactly equivalent to the choices discussed in Remark 2 and Remark 3 of Chen *et al.* (2014). The number of factors also relate directly to the choices made by Banerjee *et al.* (2014), and our cross-country datasets are made to be equivilent to the one used in that paper. We further confirm the validity of these choices in Figure 5.4 and Figure 5.5 which show the decreasing percent of variance explained, and Cattle Scree plots respectively. Analysis of these figures is further discussed in Section 5.4.1. Trials of large values in the IPC_2 criteria of Bai and Ng (2002) discussed in Section 5.4.6 also provide further justification for our choice of the maximum number of factors. For three of our models, we then recursively estimate the number of factors using the IPC_2 itself.

5.4 Methodology

In this section we detail the range of different model classes considered, the model selection/-model averaging techniques and the approach taken for the extraction of stationary and non-stationary factors. The exact approach taken to model estimation, specification and forecast generation is the focus of Section 5.4.6. Regarding the time dimension, $t = 1, \dots, T$ denotes the initial in-sample 'calibration' period; that is, in all six of our datasets: 1980:Q1-2004:Q3. The recursive procedure assumes models estimated over $t = 1, \dots, \{T+r-1\}, \dots, \{T+R-1\}$ where $r = 1, \dots, R$, and $R = 28$. Point forecasts for $H = \{1, 2, \dots, 8\}$ are then recursively generated for the horizons $t = \{T + r - 1 + H\}$ to $t = \{T + R - 1 + H\}$ from 2004:Q4 onwards.

5.4.1 Factor extraction

Recalling that not the true factors, their loadings nor idiosyncratic errors are directly observable, for stationary factors, we consider the static principle components method of Stock and Watson (2002a) which is fast, easy to implement, and, given our sample size, performs at an equivalent level to dynamic principle components, with both assuming the series are composed of two mutually orthogonal (common and idiosyncratic) components. However, a range of research shows that the forecasts produced using various estimators of factors are highly collinear,

producing similar R^2 values. Based upon simulation results, it is claimed that: "With respect to static versus dynamic factor estimates ... the estimator per se is not a choice of first-order importance" (Bernanke *et al.*, 2005, p.132). Stock and Watson (2002a) show consistency of the factors if the errors are i.) stationary, ii.) have non-trivial loadings and iii.) have idiosyncratic errors with weak serial and cross-sectional correlation.

In the notation that follows, country level subscripts are dropped for notational simplicity. Factors are extracted recursively[4], using the entire cross-section of variables (N) in our datasets excluding the four 'observed' endogenous variables (as is common in the literature; see, for example, Koop (2013)) for each economic area. For the I(0) routine, the variables are transformed into stationary variables which have zero mean and unit variance, as discussed above. The maximum number of common factors extracted is assumed to be $k_{max} \ll N$, with each factor(z) drawn from each of the datasets, denoted by $F_{z,t}$ where z=1,...,k$_{max}$\i_k$_{max}$ (where i_k$_{max}$ is the maximum number of I(1) factors). Specific factor loadings drive the cross-sectional co-movement of the variables. Formally: allow $\underset{}{X_i}$ to be a $T \times 1$ vector of observations for variable i, allowing F_t^0, Λ_t^0 to denote the true common factors:

$$\underset{T\times 1}{\underline{X_i}} = \underset{T\times r}{F_t^0} \quad \underset{r\times 1}{\lambda_i^0} + \underset{T\times 1}{\underline{\xi_i}} \tag{5.1}$$

where $\underline{X_i} = (X_{i1}, X_{i2}, \ldots, X_{iT})'$, $F^0 = (F_1^0, F_2^0, \ldots, F_T^0)' \in \mathbb{R}^{T\times r}$ and $\underline{\xi_i} = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iT})$ and r is the assumed 'true' number of factors. We take $k_{max} = 5$ for our I(0) routine, and $k_{max} = 4$ (i_k_{max}) for our I(1) routine (discussed below). For our 'macro-panel' of variables:

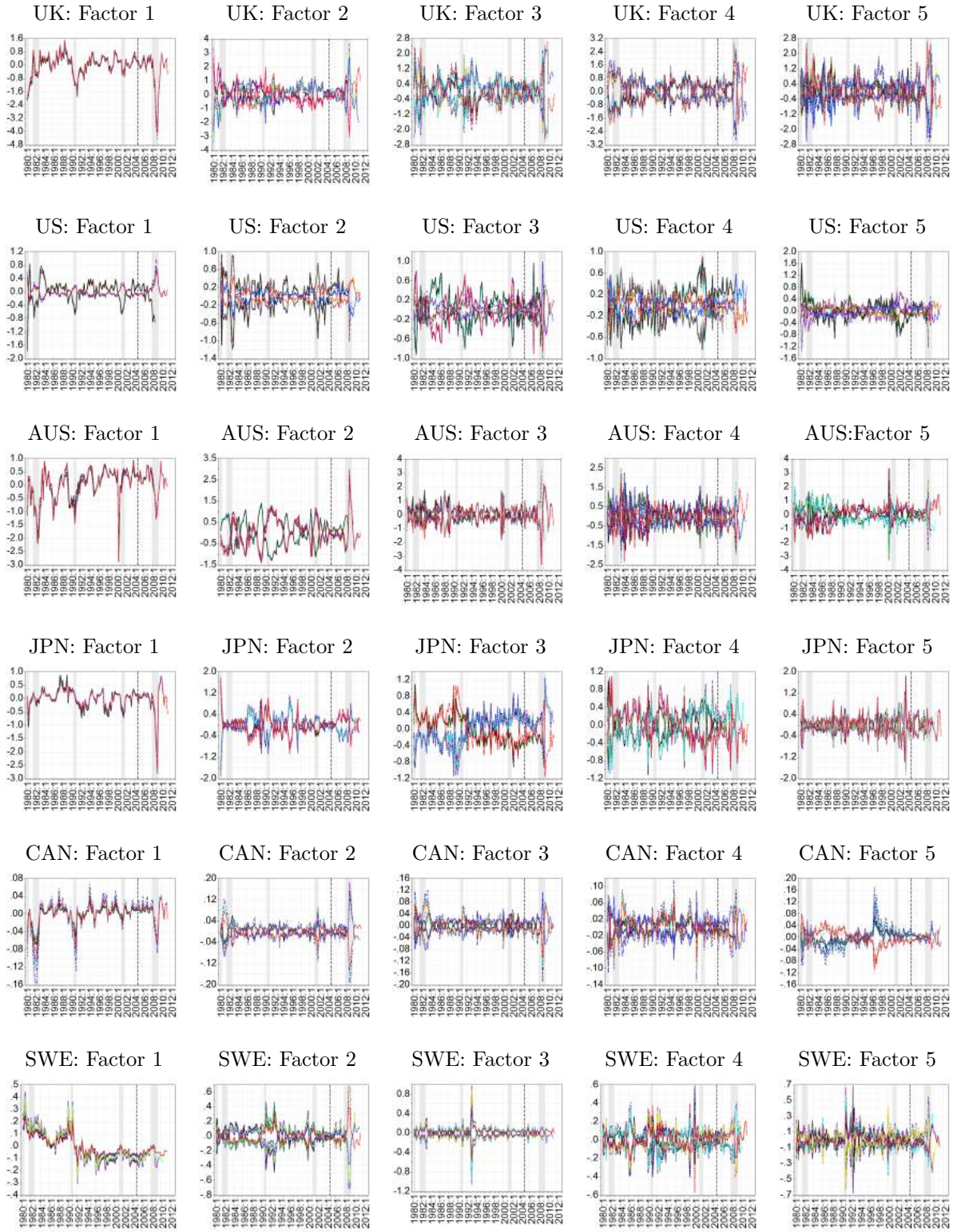
$$\underset{T\times N}{X} = \underset{T\times r}{F^0} \quad \underset{r\times N}{\Lambda^{0\prime}} + \underset{T\times N}{\xi} \tag{5.2}$$

We estimate the common factors non-parametrically using asymptotic principal components, obtained by solving the optimization problem:

$$V(k) = \min_{\Lambda, F^{k_{max}}} (N \times T)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{i,t} - \lambda_i^{k_{max}} F_t^{k_{max}})^2 \tag{5.3}$$

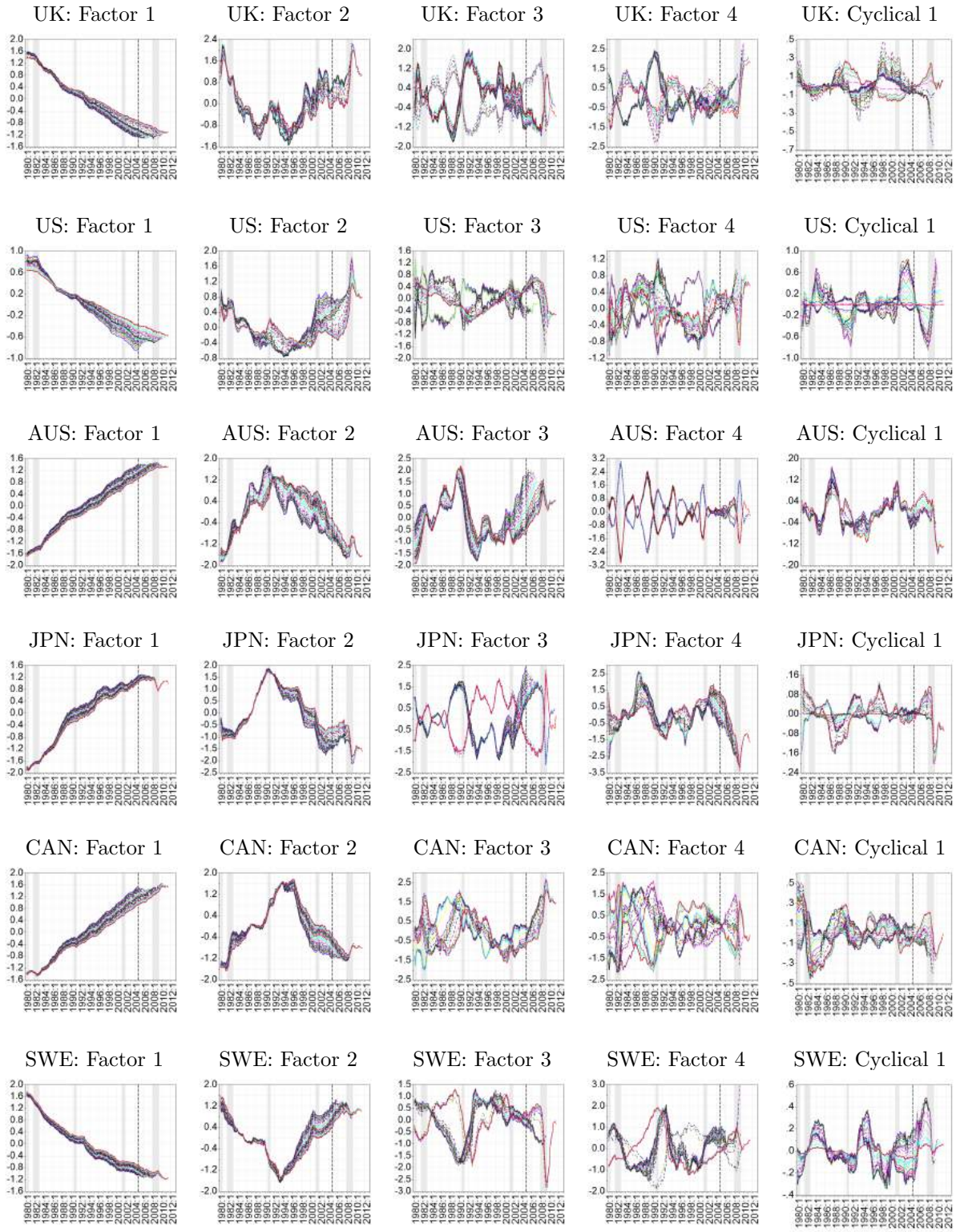
[4]We re-write and adapt the Gauss routine used by Banerjee *et al.* (2014) (available at the homepage of Igor Masten at www.econlab.si/igor.masten) for the I(0), cyclical and I(1) cases where $N > T$.

Figure 5.2: Stationary factors (recursively extracted)



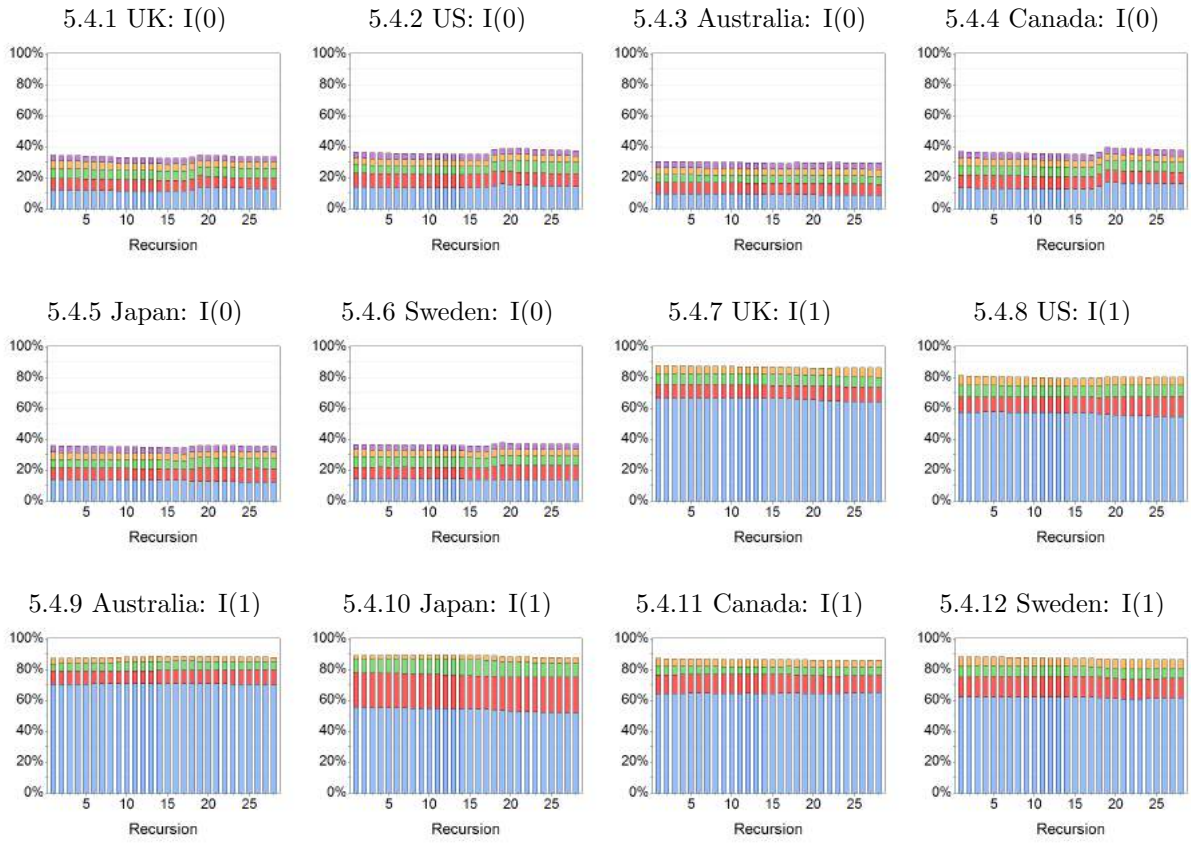
Shaded regions represent NBER recessions, vertical black dash represents final quarter of our recursive in-sample window.

Figure 5.3: Non-stationary and cyclical factors (recursively extracted)



Shaded regions represent NBER recessions, vertical black dash represents final quarter of our recursive in-sample window.

Figure 5.4: Percent of variance explained by each factor



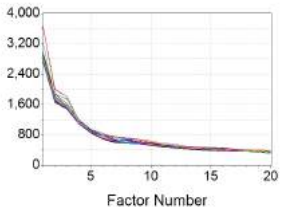
The factors chosen to solve the minimization problem (typically, in order to explain a certain proportion of the variance) can be represented as $F_t = (F_{1,t}, \ldots, F_{k_{max},t})$, a $T \times k_{max}$ matrix[5] with corresponding loadings $\Lambda = (\Lambda_{1,t}, \ldots, \Lambda_{k_{max},t})$, which is $k_{max} \times N$. At this point, note that a model with additional factors (e.g. $k_{max} + 1$) can fit no worse than a model with less (k_{max}), but efficiency is lost as more loadings are estimated. To econometrically identify the model, we apply the standard normalization that $\frac{F^{k_{max}'} F^{k_{max}}}{T} = I_{k_{max}}$.

For the I(0) factors, the estimated factor matrix is $\hat{F}^{I(0)} = \sqrt{T}\hat{P}$ where $\hat{P}$ is a $T \times k_{max}$ matrix of eigenvectors corresponding to the k_{max} largest eigenvalues of XX'. The corresponding matrix of factor loadings is $\hat{\Lambda}^{k'_{max}} = \frac{\hat{F}^{I(0)'} X}{T}$. Note that while this solution is not unique, it is a computationally less costly alternative when (as is our case) $N > T$. As one may expect, the theory of principal components estimation requires modification when applied to integrated and cointegrated variables. For the case where $N > T$, for

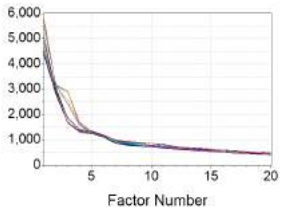
[5]For I(0) factors, the dimension shrinks to $T-1$ due to the first differences taken to induce stationarity.

Figure 5.5: Scree plots

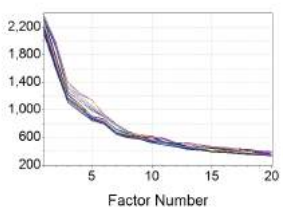
5.5.1 UK: I(0)



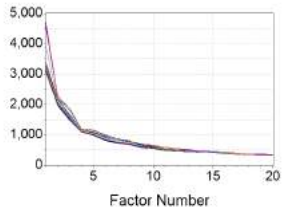
5.5.2 US: I(0)



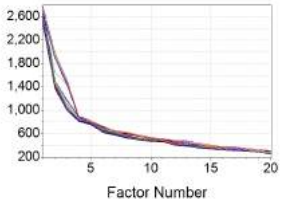
5.5.3 Australia: I(0)



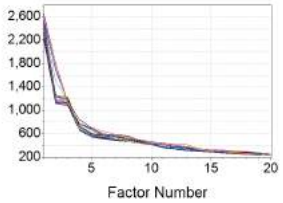
5.5.4 Canada: I(0)



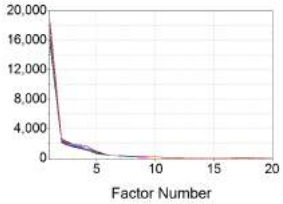
5.5.5 Japan: I(0)



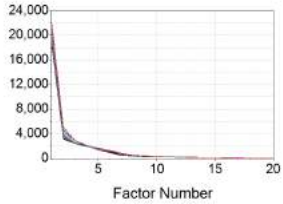
5.5.6 Sweden: I(0)



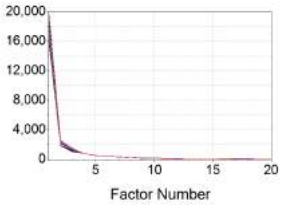
5.5.7 UK: I(1)



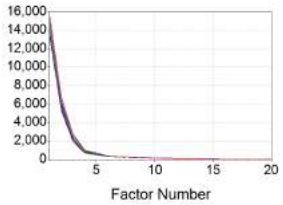
5.5.8 US: I(1)



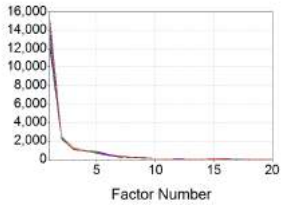
5.5.9 Australia: I(1)



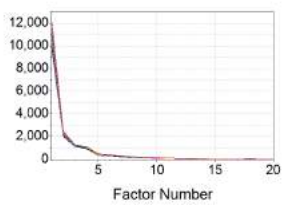
5.5.10 Japan: I(1)



5.5.11 Canada: I(1)



5.5.12 Sweden: I(1)



the I(1) methodology proposed by Bai (2004), each column of our matrix of variables X is again mean zero with unit variance, but this time is not transformed to contain only stationary variables. That is, it contains both stationary and non-stationary variables as per the 'raw' datasets described in Section 5.3. Banerjee *et al.* (2014) find no difference between a measure which excludes all stationary variables in their I(1) routine, and we deem it favorable to preserve the rich cross sectional dimension, neglecting to discard the I(0) variables in I(1) factor extraction[6]. The factors and loadings are extracted recursively as: $\hat{F}^{I(1)} = T\hat{P}$ and $\hat{\Lambda} = \frac{\hat{F}^{I(1)^T}X}{T^2}$ respectively.

The percent of variance explained by each of the I(0) and I(1) factors can be seen in Figure 5.4, showing an average cumulative variance explained across the k_{max} I(0) factors at around 35-60%, and for the i_k_{max} I(1) factors, upwards of 80%, completely in accordance with the literature. In addition to this, our choice of five I(0) factors and four

[6]Banerjee *et al.* (2005) shows an irrelevance between modelling price series as I(1) or I(2).

I(1) factors appears entirely justifiable considering the Cattle Scree plots in Figure 5.5 which depict the regions in which additional factors ceases to explain a sufficient amount of variance (the 'kink'). For the idiosyncratic cyclical factor, we extract one factor using the same method as the I(0) routine above, but instead of extracting eigenvectors from XX^T, we extract one cyclical factor from χ=$(XX^T - F^{I(1)}\Lambda^{I(1)})$. These additional I(0) factors may be thought to represent 'common cycles' (and this is how Banerjee *et al.* (2014) define the 'c' in FECMc), as opposed to those represented by the I(1) factors, and are also generated recursively. While it would be possible to give interpretation to our factors as in papers such as Bork and Møller (2012) who regress each factor with each component series, plotting against series category, such an exercise is not the primary consideration of the combinatorial nature of our forecasting algorithm[7].

5.4.2 Classical models

In order to compare our factor based methodology with classical 'observed' models, we estimate three benchmarks. The first of which is a univariate model which is estimated based on the algorithm of Hyndman and Khandakar (2008). While not the same step-wise approach, it estimates models in the possible parameter space of an ARIMA(p,d,q) of the form:

$$\Phi(B)(1 - B^d)Y_t = c + \theta(B)\varepsilon_t \tag{5.4}$$

where $\{\varepsilon_t\}$ is a white noise process with mean zero and variance σ^2, B is a backshift operator and Φ and θ are polynomials of order p and q. We discount any models where Φ and θ have no roots for —z— ¡ 1. We search the model set where $p = 1, \ldots 4$, $d = 0, 1, 2$ and $q = 1, \ldots, 4$, first running a KPSS test, and then estimating the ARMA(p,q) model on the suitably transformed data (at the 5% level). The resulting model is the one which satisfies the above condition and minimizes the Schwarz Information Criterion (SIC) in the interest of parsimony, and, like all of the following models, the specifications are determined and

[7]The majority of the code used is a variation on that used to estimate factors in Appendix B.

estimated at each recursive step for each country. Our endogenous variable (Y_t) is the (log of) real house price, real residential investment, real income, or the mortgage rate in basis points. This is termed 'Model 1' in the analysis in Section 5.5

Naturally, we also estimate corresponding symmetric vector autoregression and vector error correction models as per Eq. 5.5 and Eq. 5.6:

$$\Delta \mathbf{Z_t} = C + \Phi(L)\Delta \mathbf{Z_{t-1}} + \varepsilon_{z,t} \tag{5.5}$$

$$\Delta \mathbf{Z_t} = \alpha(\delta' \mathbf{Z_{t-1}} + \rho_0) + \Phi(L)\Delta \mathbf{Z_{t-1}} + \alpha \perp \gamma_0 + \varepsilon_{z,t} \tag{5.6}$$

We denote by $\Phi(L)$ a standard conformable lag polynomial of finite order, with the order determined recursively by the SIC and the cointegrating rank (the size of α and β is $r \times k$, where k is the number of endogenous variables and r is the rank) determined recursively by the Johansen Trace Test at the 5% level (as are the FECM/FECMc models in Section 5.4.3.1) where $\mathbf{Z_t}$ is a vector of all 'observable' endogenous variables (real house prices, real GDP, real residential investment, mortgage rate) similar to Gattini and Hiebert (2010) and Hall *et al.* (1997), discussed in Section 5.2.[8] Our deterministic specification (C in Eq. 5.5, ρ_0 and γ_0 in Eq. 5.6) is based on the assumption that the level data have linear trends but the cointegrating equations have only intercepts. These are termed 'Model 2' and 'Model 3' in Section 5.5.

5.4.3 Factor augmented models

Turning to factor models, we estimate our FAVARs as the 'two-step' variant of Bernanke *et al.* (2005)[9], where the factors are extracted by the methods of Stock and Watson (2002a), prior to the estimation of the FAVAR, as discussed above. It is duly noted that it would instead be possible to undertake the exercise as a 'direct' multi-step forecast from

[8] $\mathbf{Z_t}$ is commonly given the notation of $\mathbf{Z_t} = [Y_{1,t}, Y_{2,t}, Y_{3,t}, Y_{4,t}]$ in the literature, where $Y_{i,t}$ represent each of our four endogenous variables as per Equation 5.4.

[9] This is claimed to produce 'more plausible' structural responses than the computationally intensive Bayesian method based upon Gibbs sampling.

regressing one of the observed elements of $\mathbf{Z_t}$ on the others and the various factors $\hat{F}_t$ and their corresponding lags. However, we instead use system methodology in conjunction with the one-step ahead forecasting equation, iterating the model forward. Our justification is based on Marcellino *et al.* (2006) who give particular empirical validation over longer forecast horizons. In addition to this, it is claimed: 'when there is more than one factor, vector-autoregressive forecasts of the factors should be considered, not univariate autoregressive forecasts' (Bernanke *et al.*, 2005, p.125).

We thusly model the joint dynamics of our FAVAR system as:

$$\begin{bmatrix} \Delta \mathbf{Z_t} \\ \hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(0)}} \end{bmatrix} = C + \Phi(L) \begin{bmatrix} \Delta \mathbf{Z_{t-1}} \\ \hat{\mathbf{F}}_{\mathbf{t-1}}^{I(0)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{z,t} \\ \varepsilon_{f,t} \end{bmatrix} \tag{5.7}$$

where the $\hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(0)}}$ represents the stacked factors (the number of which is determined recursively by the IPC_2, discussed below), which are extracted from suitably transformed (i.e. I(0)) data. If the relative coefficients of Φ are zero, there exists a direct mapping onto our earlier models (that is to say that Eq. 5.5 is nested in Eq. 5.7), allowing us to analyze the contributions of $\hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(0)}}$ (i.e. the difference between forecast errors). Conversely, if the true system is indeed a FAVAR, omission of factors will lead to biased coefficient estimates in the VAR framework, and poorer forecasts. These are termed 'Model 4' in Section 5.5.

5.4.3.1 FECM: Factor error correction models

While a full derivation of the model, from VAR to ECM to FECM, is provided in Banerjee and Marcellino (2008) (with Monte Carlo\empirical forecasting applications in Banerjee *et al.* (2014)) we merely show the final models to be estimated, following similar notation[10]. In particular, from the Granger representation theorem, there exists an error correction

[10]Our equations are similar in nature to equations nine and ten in Banerjee and Marcellino (2008) and seven and eight in Banerjee *et al.* (2014) respectively.

of the form ('Model 5'):

$$\begin{bmatrix} \Delta \mathbf{Z_t} \\ \Delta \hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(1)}} \end{bmatrix} = \alpha(\delta' \begin{bmatrix} \mathbf{Z_{t-1}} \\ \hat{\mathbf{F}}_{\mathbf{t-1}}^{\mathbf{I(1)}} \end{bmatrix} + \rho_0) + \Phi(L) \begin{bmatrix} \Delta \mathbf{Z_{t-1}} \\ \Delta \hat{\mathbf{F}}_{\mathbf{t-1}}^{\mathbf{I(1)}} \end{bmatrix} + \alpha \perp \gamma_0 + \begin{bmatrix} \varepsilon_{z,t} \\ \varepsilon_{f,t} \end{bmatrix} \tag{5.8}$$

whereby the I(1) factors ($\hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(1)}}$) are extracted from the dataset using the methodology of Bai (2004). When this model is augmented with additional lags of the differenced dependent variables, we have what is defined as an FECM by Banerjee and Marcellino (2008). From a practical point of view, one issue of particular importance regards the number of cointegrating vectors to estimate. Following Banerjee *et al.* (2014), we determine this recursively through the Johansen Trace Test (at the 5% level), although it would be possible to estimate all models with all viable r. As discussed in Section 5.4.5, and as per Banerjee *et al.* (2014), we also extract single idiosyncratic cyclical components, which, when augmented with additional lag lengths, form the FECMc ('Model 6'):

$$\begin{bmatrix} \Delta \mathbf{Z_t} \\ \Delta \hat{\mathbf{F}}_{\mathbf{t}}^{\mathbf{I(1)}} \\ \Delta \hat{C}_t^{I(0)} \end{bmatrix} = \alpha(\delta' \begin{bmatrix} \mathbf{Z_{t-1}} \\ \hat{\mathbf{F}}_{\mathbf{t-1}}^{\mathbf{I(1)}} \\ \hat{C}_{t-1}^{I(0)} \end{bmatrix} + \rho_0) + \Phi(L) \begin{bmatrix} \mathbf{\Delta Z_{t-1}} \\ \Delta \hat{\mathbf{F}}_{\mathbf{t-1}}^{\mathbf{I(1)}} \\ \Delta \hat{C}_{t-1}^{I(0)} \end{bmatrix} + \alpha \perp \gamma_0 + \begin{bmatrix} \varepsilon_{z,t} \\ \varepsilon_{f,t} \\ \varepsilon_{c,t} \end{bmatrix} \tag{5.9}$$

5.4.4 Candidate models

Equations 5.4 to 5.9 above are used to generate the first six of the twelve forecasts which we evaluate. The second set of six come from selecting four of, and weighting over the set of candidate models which are defined as follows. The largest model in our set of candidate models (for the FAVAR - where k_{max}=5 and the FECMc - where i_k_{max}=4 augmented with one additional cyclical factor), is the one which includes all endogenous 'observed' variables and the maximum number of factors for both FAVAR and FECMc style models. Within this large model, we estimate all possible subsets of endogenous 'observable variables' with all possible subsets of factor combinations, with one of the

variables stacked in $\mathbf{Z_t}$ always remaining as an endogenous variable along with at least one of the factors. That is; any given model may include all factors or variables, a subset of factors or a subset of variables. Only four possible 'observed' variables are chosen (instead of a larger number, such as the ten variables used in Iacoviello and Neri (2010), as per Gupta *et al.* (2011a)) in order to follow the advice of Cheng and Hansen (2015) so that the total number of models is practically and computationally feasible.

As in Cheng and Hansen (2015), our strategy is to treat the lag and factor structure as unknown, selecting and weighting across the model set. Given that there are three supplementary 'observable' variables for each one of our four endogenous variables, and five possible factors (whereby we exclude a null subset of factors which would render us equivalent to Eq. 5.5 or Eq. 5.6), we estimate $2^3 \times (2^5 - 1) \times 4$ models for both the FAVAR and FECMc, with $k = \{1, 2, 3, 4\}$, giving a total of 3,968 factor models[11] for each of the FAVAR and FECMc model classes, generated at each of 28 recursive periods across six datasets, for a total of 1,333,248 ($2 \times 28 \times 3968 \times 6$) candidate factor models[12]. This can be thought of as our total set of candidate models.

5.4.5 Estimation techniques

We initially estimate our sets of automated ARIMA model specifications with maximum likelihood (BFGS optimization algorithm), and VAR and FAVAR models with ordinary least squares. For the VECM and FECM models (using the same procedure for both), we use a two-step method. The estimate of $\hat{\delta}$ of the cointegrating matrix in Eq. 5.6 and Eq. 5.9 is computed by maximum likelihood in the first step. This is then substituted in as the true δ with all other parameters estimated in the second stage of the least squares estimation (Lütkepohl, 2006, p.301) for α and Φ. We determine the size of α and δ as

[11] It would be possible to estimate a smaller number of models than is done in the accompanying code files, which take into account the ordering of the variables in the loops within the script, although our approach loops through each variable one by one, ordering it first, in order to make the program as accessible as possible for replication purposes.

[12] If any models cannot be estimated (which may happen due to singularity in a small number of calculations of the error correction term), they are not candidates for model selection and are subsequently assigned a weight of zero in the following routines.

per the Johansen Trace Test (5%) recursively where applicable, and separately for each different combination of observed variables and factors.

5.4.6 Forecasting approach

The forecasting approach which we take is as follows. We utilize six datasets described in Section 5.3. For each of these datasets, we attempt to create a 'pseudo-real time' forecasting environment through a recursive procedure. Each of our forecasts are generated through an iterative system-based procedure as opposed to a dynamic multistep framework. We begin by estimating our models 'in-sample' up until 2004Q3 and then generate multi-step point forecasts which are evaluated at horizons of $1, 2, \ldots, 8$ periods ahead ($h = 1, \ldots, 8$). The models are then re-estimated, recursively moving one step ahead (that is, for example where $r = 2$, 2004Q4: $t = (T + r - 1)$) with forecasts being generated between 2005Q1 (that is, a point forecast for $T + r - 1 + h$) up to 2006Q4, and so on. This procedure is repeated until 28 recursions ($r = 1, 2, \ldots, 28(R)$) are complete. We do this for each of our six 'pre-determined' models, and for our entire set of candidate models as discussed in Section 5.4.4.

In general, if we let $w = ((w(1), ..., w(M))'$ denote the weight vector, requiring that the weights sum to one: $\sum_i^M w_i = 1$ and in our case, constrained to be nonnegative ($0 < w(i) < 1$), model selection is akin to a combination routine where $w(i) = 1$ and $w_j = 0 \; \forall \; i \neq j$. More generally, combination involves selecting the weight vector w from $\mathcal{H}^M$, the unit simplex in $\mathbb{R}^M$. For the p and q of the ARIMA(p,d,q) model, and k of the VAR and VECM models ('Model 1' to 'Model 3'), we utilize the SIC:

$$\text{SIC}_i^m = \frac{-2 \times ln(\mathcal{L}_i^m)}{n_i^m} + \frac{p_i^m \times ln(n_i^m)}{n_i^m} \tag{5.10}$$

where p_i^m represents the number of estimated coefficients for model m, n_i^m the number of observations for model m, and $\mathcal{L}_i^m$ the log likelihood of model m, all for variable i. For 'Model 2' and 'Model 3', we use system log likelihoods.

Following this, we recursively select the number of I(0) and I(1) factors for 'optimal' FAVAR, FECM and FECMc models using the IPC_2 of Bai and Ng (2002) as per Banerjee *et al.* (2014), due to the fact that unlike the PC criterion, the IPC_2 is not contingent on k_{max} or i_k_{max}. The procedure is as follows, where we choose k to minimize IPC_2 criteria recursively (although IPC_1 and IPC_3 criterion are also generated and determined as a robustness check), and our maximum number of permitted factors is as per Section 5.4.1. First, we estimate the sum of squared residuals divided by $N \times T$ when k factors are estimated:

$$V(k, \hat{F}_t) = \min_{\Lambda} \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{(i,t)} - \lambda_i^{k'} \hat{F}_t^k)^2 \tag{5.11}$$

As per Bai and Ng (2002), denote $C_{N,T}^2 = min\{N, T\}$, and then the IPC_2 is defined for k factors as:

$$IPC_2(k) = ln(V(k, \hat{F}_t^k)) + (k \times \frac{(N+T)}{(N \times T)} \times lnC_{N,T}^2) \tag{5.12}$$

where we choose the number of factors (k) which minimizes IPC_2[13]. Having used the SIC to select the ARIMA specification ('Model 1'), and the lags for each of the VAR ('Model 2') and VECM ('Model 3') models, we then use it to determine the number of lags in our IPC_2 determined FAVAR/FECM/FECMc models ('Model 4' - 'Model 6'). The number of factors and the number of lags is re-specified on a recursive basis each period for each country for each model.

5.4.6.1 Selecting across the total set of candidate models

For the four forecasts which are generated from 'selected' models from our candidate set ('Model 7' - 'Model 10'), two are system based, and two are equation based. The four selection techniques which we use are as follows. 'Model 7' and 'Model 8' take an approach extremely common to the literature, in that out of the 992 FAVAR and 992 FECMs available per variable per recursion per country, they choose the model with the lowest system SIC and AIC for the variable of interest respectively. Where the SIC is

[13]The majority of the code used to calculate the Bai and Ng (2002) style criteria in this section was also featured in Appendix B.

calculated as per Equation 5.10, the AIC is calculated as $\text{AIC}_i^m = \frac{-2\times ln(\mathcal{L}_i^m)}{n_i^m} + \frac{2\times p_i^m}{n_i^m}$. 'Model 9' takes the form of the model which maximizes the $\bar{R_i^m}^2$ at an equation level, where $\bar{R}_i^{m,2} = 1 - (1 - R_i^{m,2})\frac{n_i^m - 1}{n_i^m - p_i^m}$, $R_i^m = 1 - \frac{\varepsilon_{i,t}^{m,\top}\varepsilon_{i,t}^m}{(Yi,t-\bar{Y}_{i,t})^\top \times (Yi,t-\bar{Y}_{i,t})}$ and $\bar{Y}_{i,t} = \sum_{t=1}^{T}\frac{Y_{i,t}}{n_i^m}$, where subscript i in $Y_{i,t}$ and $\varepsilon_{i,t}^m$ refers to the equation (endogenous variable) of interest in the system. For Model '10', we take the model which minimizes the sum of squared residuals ($SSR = \varepsilon_{i,t}^{m,\top}\varepsilon_{i,t}^m$), and, unlike the other models estimated, this model makes no penalization for additional estimable parameters.

5.4.6.2 Weighting across the total set of candidate models

Following the selection and identification of our three classical models (ARIMA/VAR/VECM), our three factor models (FAVAR/FECM/FECMc) models and our four 'model selection' models (minimum AIC/SIC/SSR, maximum $\bar{R}^2$), we utilize two forecasts created from weighting across the entire set of candidate models. One is created with approximate Bayesian Model Averaging (BMA) weights, and one with a simple naïve weighted average. In forecast combinations generally, a large number of individual models are combined (denoted as ^{c}, below) to produce a weighted forecast: $\hat{Y}_{i,t+h|t}^c$ from different forecasts ($\hat{Y}_{i,t+h|t}^m$, with each model taking an ^{m} superscript). Weights for each individual model are denoted as $\hat{w}_{i,t}^m$. The combination is thus calculated as:

$$\hat{Y}_{i,t+h|t}^c = \sum_{i=1}^{M} \hat{w}_{i,t}^m \hat{Y}_{i,t+h|t}^m \tag{5.13}$$

Having calculated the $\text{SIC}_{i,t}^m$ for each model(m) at each recursion (t) from Eq. 5.10 for each variable (i), we 'weight in evidence' in favor of the model m being the best approximating model out of M total models. We construct 'exponential' weights as is common in the literature such as Cheng and Hansen (2015) based on the original work on Bayesian Model Averaging of Min and Zellner (1992):

$$\hat{w}_{i,t}^m = \frac{exp(-\frac{1}{2}SIC_{m,t})}{\sum_{m=1}^{M} exp(-\frac{1}{2}SIC_{m,t})} \tag{5.14}$$

where $M = 1,984$ for each of our dependent variables for each country for each recursion. For the SIC case, assuming diffuse priors and equal model prior probabilities, Eq. 5.14 can be interpreted as approximate BMA weights. This combination forecast is denoted as 'Model 11' in the ensuing analysis. A cursory, unreported analysis shows that models with fewer parameters are assigned higher weights than the fully parameterized model, and this is no surprise given the well documented properties of the information criterion being used. For the final forecast to be evaluated, we use uniform weights only ($\hat{w}_{i,t}^{m} = \frac{1}{1984}$): a simple naïve weighted average, termed 'Model 12'.

Models are re-estimated and re-specified for 28 recursions (separately for each country) up until 2011Q3, at each recursion generating multi-step forecasts, with the last forecast occurring at the final recursion (for h=8) in 2013:Q3. While we mainly consider only either the average of the absolute forecast error, or the number of times a specific forecast error is lower in magnitude than its rivals, we also consider how forecast error responds to underlying volatility, as discussed as appropriate in Section 5.5.

5.5 Results

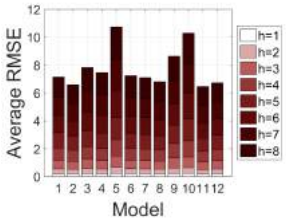
In Table 5.1, we recap the 12 forecasts generated which are described above and the forecast errors from which are plotted in Figures 5.7 - 5.9 and discussed in Sections 5.5.1 to 5.5.5.

5.5.1 Forecast error

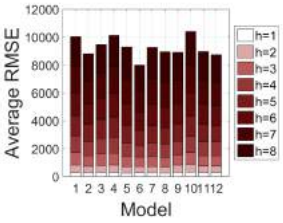
In Figures 5.6-5.8 (below), we consider different aggregations of the absolute forecast error. In Figure 5.6, we display average forecast error per model per variable per country. This is akin to the standard mean squared forecast error (MSFE) metric, where – in the code files – for each variable, at each recursion, we calculate the average one to eight period ahead forecast, and then average this over each iteration. With each segment of each bar representing a specific forecast error for a specific horizon, the total height of the bar is

Figure 5.6: Average forecast errors

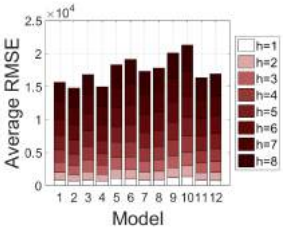
5.6.1 UK: Prices



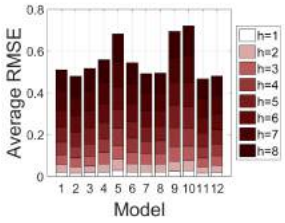
5.6.2 UK: Income



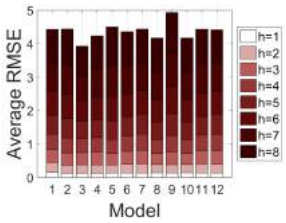
5.6.3 UK: Investment



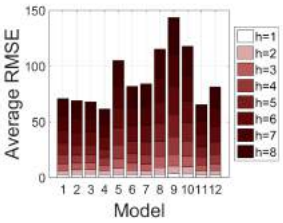
5.6.4 UK: Mortgages



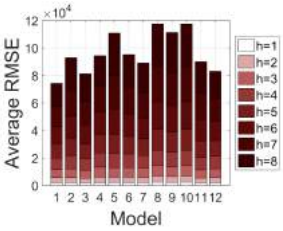
5.6.5 US: Prices



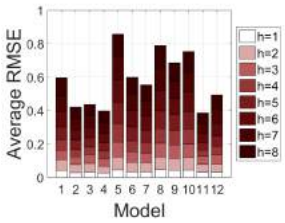
5.6.6 US: Income



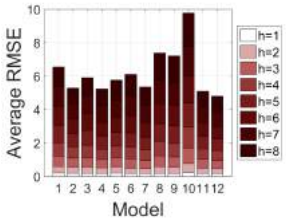
5.6.7 US: Investment



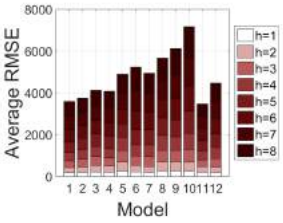
5.6.8 US: Mortgage



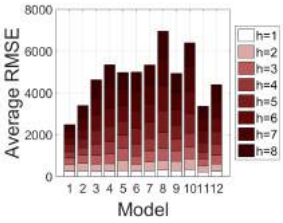
5.6.9 AUS: Prices



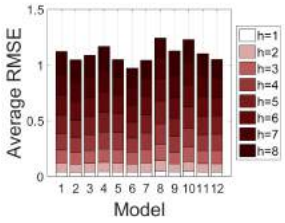
5.6.10 AUS: Income



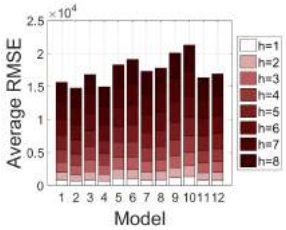
5.6.11 AUS: Investment



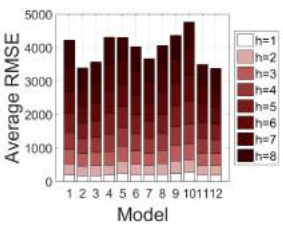
5.6.12 AUS: Mortgage Rate



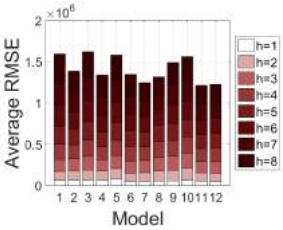
5.6.13 JPN: Prices



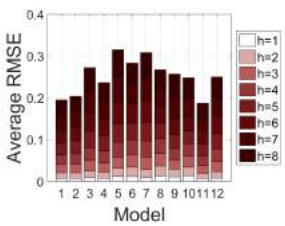
5.6.14 JPN: Income



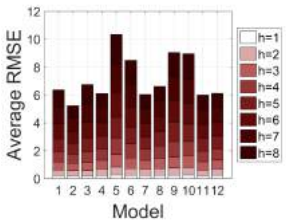
5.6.15 JPN: Investment



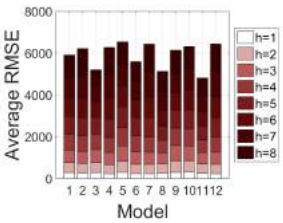
5.6.16 JPN: Mortgage



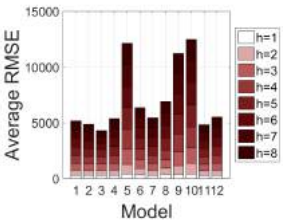
5.6.17 CAN: Prices



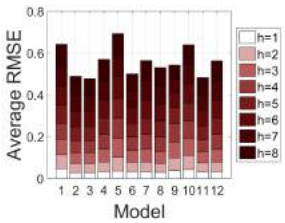
5.6.18 CAN: Income



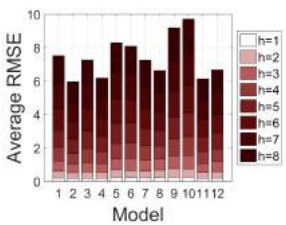
5.6.19 CAN: Investment



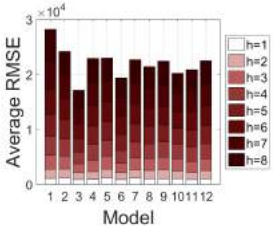
5.6.20 CAN: Mortgage



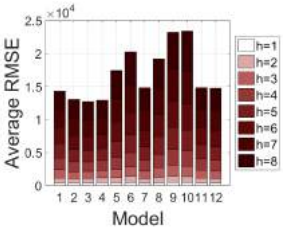
5.6.21 SWE: Prices



5.6.22 SWE: Income



5.6.23 SWE: Investment



5.6.24 SWE: Mortgage

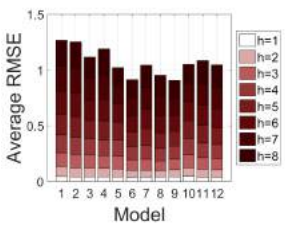
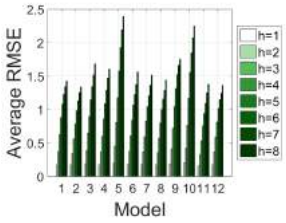
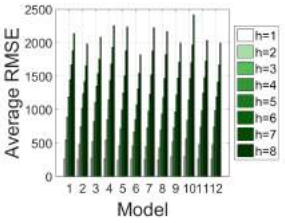


Figure 5.7: Errors across horizons

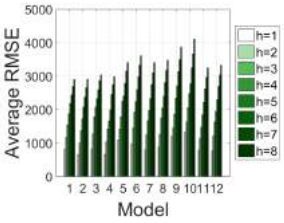
5.7.1 UK: House Prices



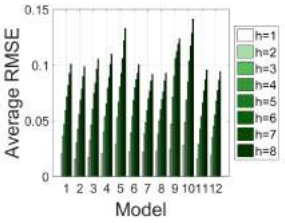
5.7.2 UK: Income



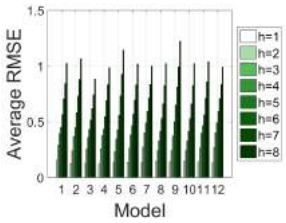
5.7.3 UK: Investment



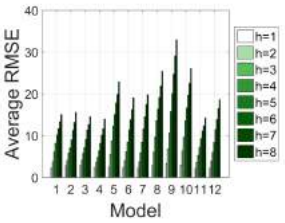
5.7.4 UK: Mortgages



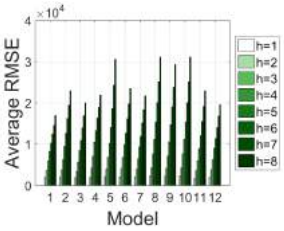
5.7.5 US: House Prices



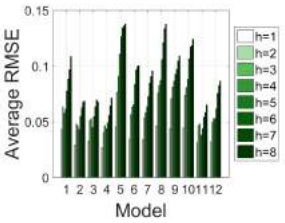
5.7.6 US: Income



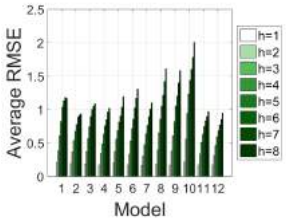
5.7.7 US: Investment



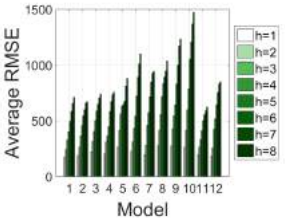
5.7.8 US: Mortgage



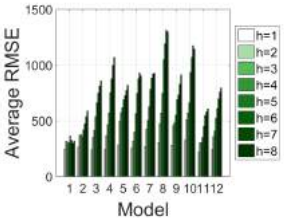
5.7.9 AUS: House Prices



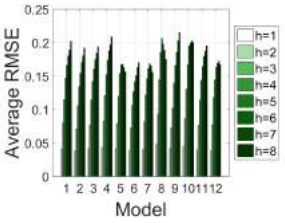
5.7.10 AUS: Income



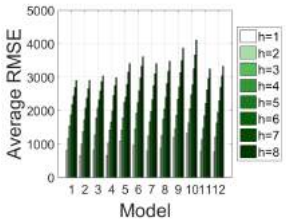
5.7.11 AUS: Investment



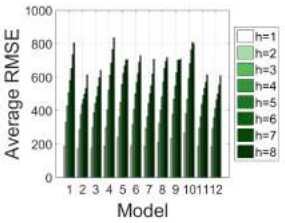
5.7.12 AUS: Mortgage Rate



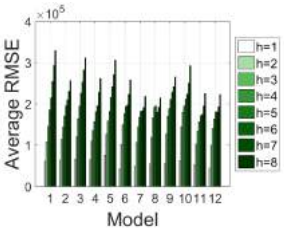
5.7.13 JPN: House Prices



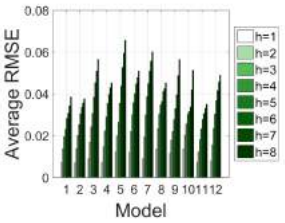
5.7.14 JPN: Income



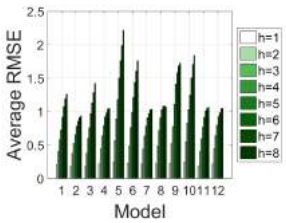
5.7.15 JPN: Investment



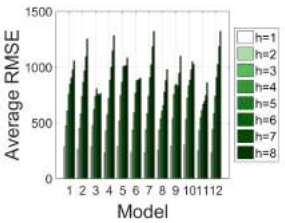
5.7.16 JPN: Mortgage



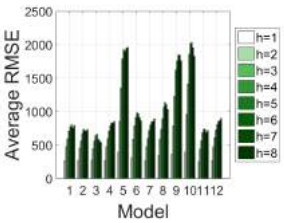
5.7.17 CAN: House Prices



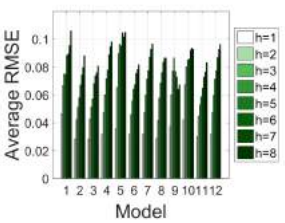
5.7.18 CAN: Income



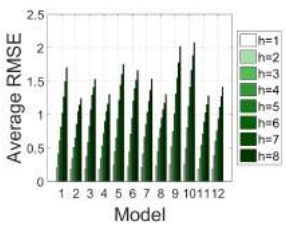
5.7.19 CAN: Investment



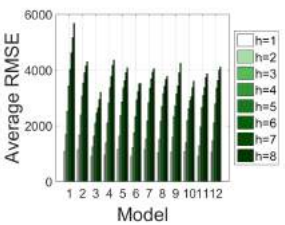
5.7.20 CAN: Mortgage



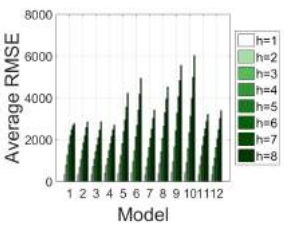
5.7.21 SWE: House Prices



5.7.22 SWE: Income



5.7.23 SWE: Investment



5.7.24 SWE: Mortgage

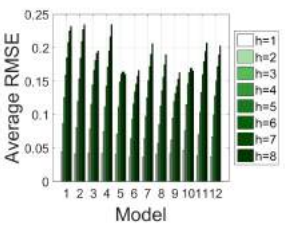
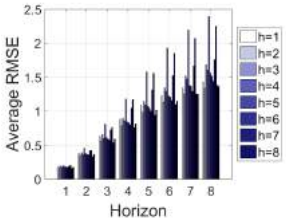
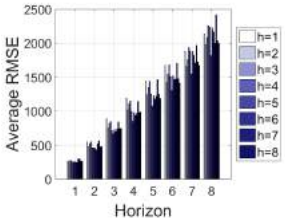


Figure 5.8: Average forecast errors

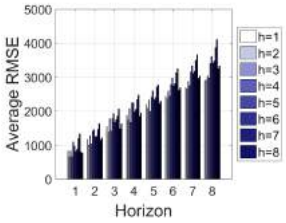
5.8.1 UK: Prices



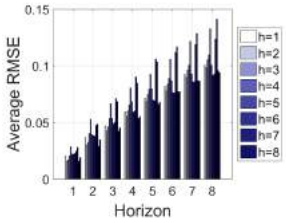
5.8.2 UK: Income



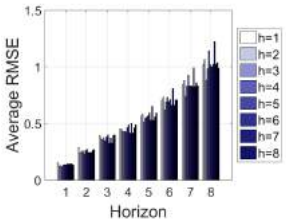
5.8.3 UK: Investment



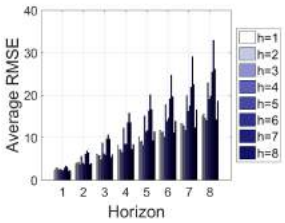
5.8.4 UK: Mortgages



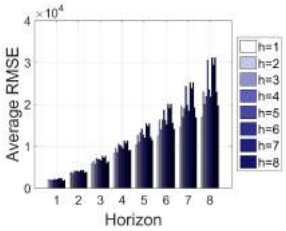
5.8.5 US: Prices



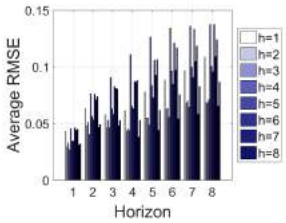
5.8.6 US: Income



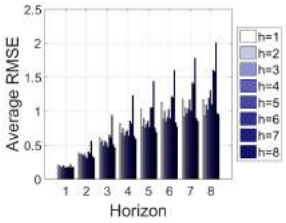
5.8.7 US: Investment



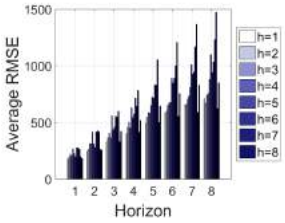
5.8.8 US: Mortgage



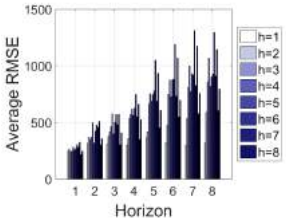
5.8.9 AUS: Prices



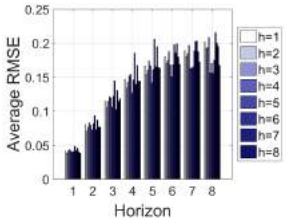
5.8.10 AUS: Income



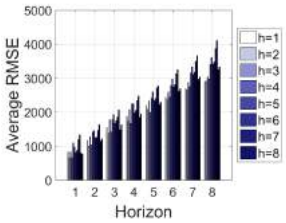
5.8.11 AUS: Investment



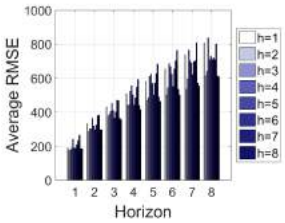
5.8.12 AUS: Mortgage Rate



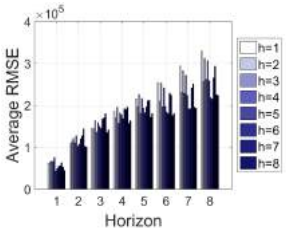
5.8.13 JPN: Prices



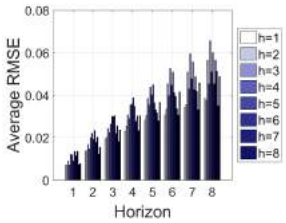
5.8.14 JPN: Income



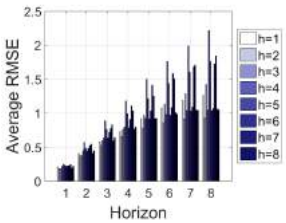
5.8.15 JPN: Investment



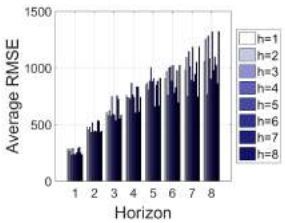
5.8.16 JPN: Mortgage



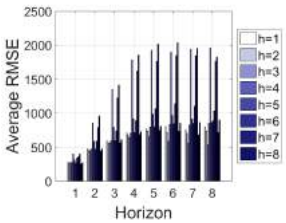
5.8.17 CAN: Prices



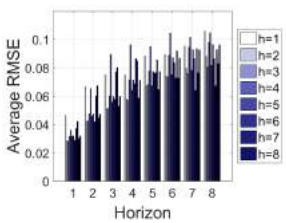
5.8.18 CAN: Income



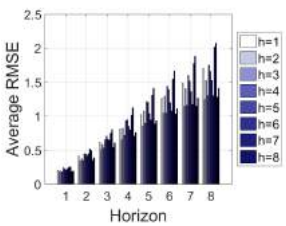
5.8.19 CAN: Investment



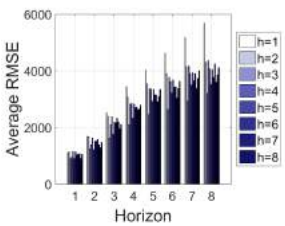
5.8.20 CAN: Mortgage



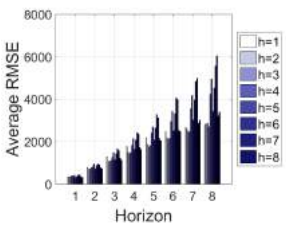
5.8.21 SWE: Prices



5.8.22 SWE: Income



5.8.23 SWE: Investment



5.8.24 SWE: Mortgage

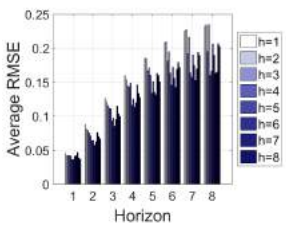


Table 5.1: Twelve models for evaluation

Model	Type	Classical	Factor	Candidate Set	Weighted
Model 1	ARIMA	✓	×	×	×
Model 2	VAR	✓	×	×	×
Model 3	VECM	✓	×	×	×
Model 4	FAVAR	×	✓	×	×
Model 5	FECM	×	✓	×	×
Model 6	FECMc	×	✓	×	×
Model 7	Min SIC	×	✓	✓	×
Model 8	Min AIC	×	✓	✓	×
Model 9	Max $\bar{R}^2$	×	✓	✓	×
Model 10	Min SSR	×	✓	✓	×
Model 11	BMA	×	✓	✓	✓
Model 12	Naïve	×	✓	✓	✓

calculated as $\frac{\sum_{r=1}^{28}\sum_{h=1}^{8}(|Y^{g}_{i,t+r+h}-\hat{Y}^{m,g}_{i,t+r+h}|)}{8\times 28}$, where $\hat{Y}^{m,g}_{i,t+r+h}$ represents the forecast for each of the dependent variables (i) in each country (g) for each model (m), and $Y^{g}_{i,t+r+h}$ the realized value. Considering the average performance of specific models, forecasts based on weighted combinations perform consistently well. While no model is able to systematically dominate for any country or variable, the average forecast error of weighted models is consistently low. In particular, the approximate BMA method ('Model 11') of forecast combination generated consistently favorable forecasts, generating lower forecast errors than its uniform counterpart ('Model 12') in 16 of the 24 plots in Figure 5.6. Neither of these two models ever generate the largest average forecast error. 'Model 10', based on minimizing the sum of squared residuals, with no penalization for over-parameterization, appears to perform worst under this metric, producing the highest average forecast error in 10 of the 24 panels. Classical models ('Model 1'-'Model 3') appear to also perform consistently well, generating the largest forecast error only 4 out of 24 times between them.

In Figure 5.7, we decompose the elements of the stacked bar chart in Figure 5.6 into bars of their own, which allows us to better consider how the forecast error rises in magnitude across each of the horizons for each model. In this figure, the height of the bars is simply measured as the average forecast error for each horizon: $\frac{\sum_{r=1}^{28}(|Y^{g}_{i,t+r+h}-\hat{Y}^{m,g}_{i,t+r+h}|)}{28}$.

As may be predicted by econometric theory, one recurring observation is that forecast error is increasing in the horizon - average forecast error is several times larger in magnitude at the two year horizon than the one quarter horizon, for each model for each variable for each country. This is true apart from a couple of notable exceptions such as for the U.S. mortgage rate, and residential investment in Australia. Again, as in Figure 5.6, we are generally able to witness the poor performances of 'Model 9' and 'Model 10' - the lesser used methods of model selection across the candidate set.

The final method of viewing this metric (average forecast error) is to consider each horizon on the x-axis, with each bar decomposed into each specific model, and is shown in Figure 5.8. The higher bars again represent higher (less favorable) forecast errors. Again, an identical result is apparent, in that not only is the forecast error almost monotonically increasing in the horizon, but the peaks of the groups of bars occur in the Model Selection region ('Model 6' - 'Model 10') of the plots. From this figure, this observation becomes more apparent at longer horizons, with higher average absolute forecast error values occurring (with respect to the forecasts generated by either the weighting, or classical models) at longer horizons.

5.5.2 The best performing model

The number of times one of the twelve forecasts has the lowest average eight-step forecast error across all recursions and all countries is shown in Table 5.2, with the final column averaging over the previous four. Note that there are 168 competing horizons for each. An important comment relates to the fact that while it may have been expected that each column sums to a hundred, a feature of the code is such that when two model selection techniques select the same model, they are each accredited with having the best forecast, rather than sharing the attribute. The results strongly suggest that if the forecasting objective is to generate the absolute lowest forecast error with no regard for the magnitude of the loss, then the classical models (notably - the Hyndman and Khandakar (2008) ARIMA selection algorithm) may be the best way to achieve this on aggregate.

If we form a prior expectation that each of the models is equally likely to generate the lowest forecast, they would be expected to acheive the best forecast 8.333% of the time each. Out of sixteen possible entries in the table, our method of model selection over the candidate set, it is only able to achieve this (14 times the best model) twice, while the classical models achieve this value 11 of 12 possible times. Interestingly, we can see that on average, the FECMc performs best more times than the FECM, which subsequently performs best more times than the FAVAR model ('Model 6', 'Model 5' and 'Model 4' respectively).

Table 5.2: The percent of times a model performs best

	House Prices	Income	Investment	Mortgage Rate	Average Percent	Ranking
Model 1	11.31%	14.286%	19.643%	13.69%	14.732%	1
Model 2	11.905%	9.524%	10.714%	8.929%	10.268%	4
Model 3	9.524%	13.69%	14.881%	4.762%	10.714%	3
Model 4	8.929%	3.571%	8.929%	10.119%	7.887%	7
Model 5	5.357%	8.929%	7.143%	10.714%	8.036%	6
Model 6	8.333%	14.286%	4.762%	16.667%	11.012%	2
Model 7	4.167%	5.952%	7.143%	7.143%	6.101%	12
Model 8	13.095%	7.143%	7.143%	7.738%	8.78%	5
Model 9	7.738%	6.548%	4.762%	5.952%	6.25%	10
Model 10	8.333%	4.762%	4.762%	7.143%	6.25%	10
Model 11	6.548%	10.714%	5.952%	5.357%	7.143%	9
Model 12	10.119%	2.976%	8.333%	7.738%	7.292%	8

5.5.3 Beating the ARIMA benchmark

One of the most important binary output metrics is the ability of our classical, factor, or candidate model set to outperform a standard univariate benchmark ('Model 1'). While we might not be concerned with consistently generating the absolute lowest forecast error, we might be most concerned with choosing a forecast which is able to most consistently beat an established, recognized benchmark. The results in this section therefore concur with those in Section 5.5.1, as we can see in Table 5.3: the model most able to beat the ARIMA is the approximate BMA model at 58.036% of the time on average. We are again able to see a strong performance of 'Model 12', the naïvely weighted model, although it is again unable to perform as admirably as the BMA under this criteria.

Table 5.3: Outperforming the Hyndman and Khandakar (2008) algorithm

	House Prices	Income	Investment	Mortgage Rate	Average Percent	Ranking
Model 2	62.5 %	48.214%	52.381%	58.929%	55.506%	2
Model 3	54.762%	54.762%	43.452%	51.19%	51.042%	5
Model 4	60.119%	44.048%	48.214%	56.548%	52.232%	4
Model 5	42.262%	45.833%	33.929%	44.048%	41.518%	9
Model 6	51.19%	51.786%	35.714%	54.762%	48.363%	7
Model 7	58.333%	39.881%	42.262%	54.762%	48.810%	6
Model 8	53.571%	44.643%	32.738%	50 %	45.238%	8
Model 9	42.262%	43.452%	30.952%	44.643%	40.327%	10
Model 10	41.071%	39.881%	26.786%	45.833%	38.393%	11
Model 11	63.69%	61.905%	47.024%	59.524%	58.036%	1
Model 12	63.69%	45.238%	44.048%	57.738%	52.679%	3

5.5.4 Forecast error as a ratio against the benchmark

In the final part of analysis of this type, Table 5.4 shows the average forecast errors of each variable for each model as a ratio of the corresponding forecast errors for that variable for 'Model 1' - the Hyndman and Khandakar (2008) algorithm. The results highlight the fact that while lower eight step errors might be generated (compared to the ARIMA model) more consistently by 'Model 11' and 'Model 12', overall, the ARIMA model might be preferable if the objective is to reliably generate a forecast with consistently reliable performance. As above, some models (notably 'Model 9' and 'Model 10') have an average forecast error double that of the benchmark, as indicated by error ratios greater than two. The final column of this table provides a ranking of the models on the basis of the average ratio (which is a more appropriate metric than in the subsections above due to the less binary nature of outcomes). This simple ranking process is again able to highlight the favorable performance of the weighted models, and the less favorable selection method of the less common techniques used to generate Model '9' and Model '10'.

5.5.5 Forecast error and underlying variance

In Figure 5.9, we plot a simple scatter plot of average forecast error (across recursion) for each variable in each country, against the underlying volatility (standard error) of each series against the entire sample, as done in Bork and Møller (2015) and Chapter 2.

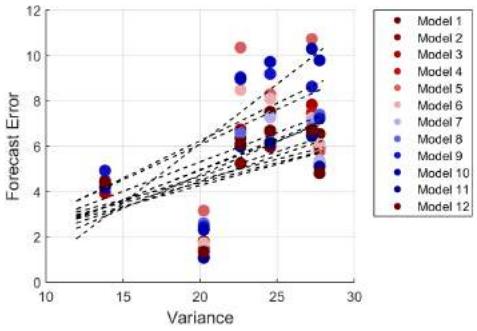
Table 5.4: A ratio to the Hyndman and Khandakar (2008) algorithm

	House Prices	Income	Investment	Mortgage Rate	Average Ratio	Model Ranking
Model 2	1.0718	1.2599	1.2525	1.0762	1.1651	2
Model 3	1.3233	1.3838	1.4479	1.1455	1.3251	5
Model 4	1.1643	1.3995	1.4245	1.1912	1.2949	3
Model 5	1.9057	1.8094	2.1554	1.6164	1.8717	9
Model 6	1.4829	1.6025	1.7243	1.3674	1.5443	7
Model 7	1.2325	1.6171	1.5833	1.3739	1.4517	6
Model 8	1.5168	1.6743	1.9138	1.463	1.642	8
Model 9	1.8385	2.0253	2.1531	1.4827	1.8749	10
Model 10	1.8976	2.1197	2.3019	1.6545	1.9935	11
Model 11	1.1074	1.1631	1.2745	1.0325	1.1444	1
Model 12	1.1259	1.4342	1.4027	1.243	1.3014	4

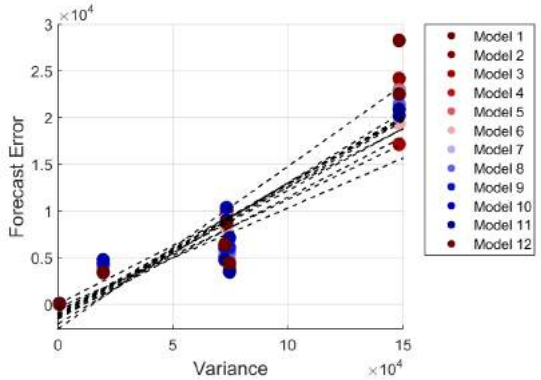
Our results concur with these two aforementioned papers and the pre-existing economic theory, that more volatile series are difficult forecast (higher forecast errors), as shown by the strong correlations between error and standard deviation.

Figure 5.9: Forecast error and variance

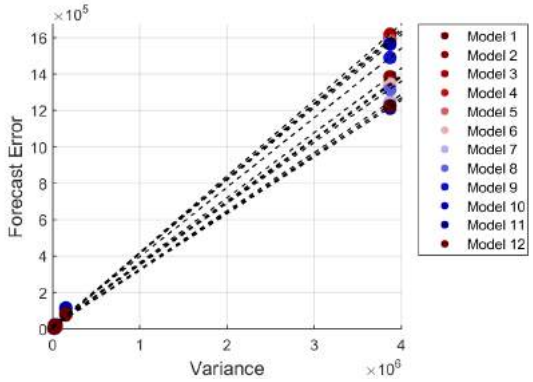
5.9.1 UK: House Prices



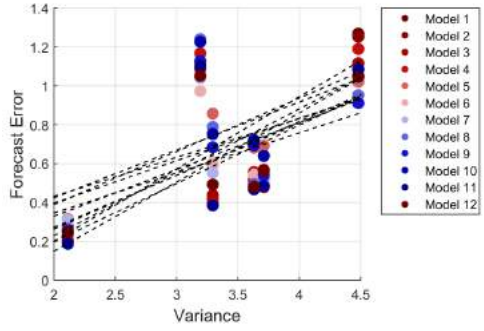
5.9.2 UK: Income



5.9.3 UK: Investment



5.9.4 UK: Mortgages



5.6 Conclusion

In this paper we develop a forecasting algorithm to forecast the typical variables used in a small macroeconometric model of housing markets over a range of different developed countries using six original large macro-panel datasets. Extracting a maximum of five stationary and four non-stationary and one stationary cyclical factor, we estimate all possible combinations of factors and key 'observable' across large sets of FAVAR and restricted FECM models, comparing the results of classical autoregressive models against a range of model selection and model weighting algorithms.

The results show the ability of an approximate BMA method of forecast combination across our large model sets is able to outperform our univariate benchmark more than 58% of the time. However, although the average ratio of forecast errors between the ARIMA model and the approximate BMA forecasts (while this method again performs best under this criteria) is still above 1, indicating that on average, depending on the loss function of the forecaster involved, the algorithm of Hyndman and Khandakar (2008) may remain the preferred model. Other results include the increasing of forecast error in the horizon, and fact that some of the model selection routines can generate a very large forecast error when a poor model is selected (especially enforcing the importance of parsimony).

While 'Model 11' and 'Model 12' provide reasonable evidence of the ability of factor based models to perform well, there are two major reasons why models 4-10 do not perform better. The first of which is potentially due to the quality of the underlying data which is utilized for factor extraction. While datasets such as the one compiled from US aggregates are typical in the literature, few studies have focused on Sweden, Canada or Australia due to a shortage of the series necessary to generate factors which fully capture common unobserved forces. Recent literature such as Chen *et al.* (2014) has also focused on time variance of factor loadings (parameter shifts which are not local to zero). Further research would draw inspiration from earlier papers such as Banerjee *et al.* (2008) who claim that it is the instability of factor loadings which may cause factor-based forecasts to fail to outperform classical models.

Other avenues for further research in this field of factor combinations might include the development of more real-time big datasets for forecasting purposes and then to estimate and compare similar specification/combination methodologies with the ones presented in this paper. Another possibility would be to instead estimate the models as 'direct' rather than 'iterated' forecasts, or to consider other information criteria, such as those promoted in the recent selection and weighting literature (such as Cheng and Hansen (2015)). Other interesting recent developments in the field could lead to applications of time varying parameter factor (such as TVP-FAVAR) models being applied to house price forecasting, with the potential for a further reduction in forecast error.

CHAPTER 6

CONCLUSION

We have taken a computational approach to existing econometric questions. In every chapter, we have attempted to highlight the difference between this approach and a benchmark 'classical' model, showcasing the danger of 'eyeball econometrics' across large model sets. Our primary intention has been to document the dangers of failing to utilize modelling frameworks which make the best use of large, freely accessible datasets and ever increasing computational power. In light of this work, there are a number of important conclusions to be drawn and policy implications to be noted. First of all, it is important to recall that all routines are run locally, and that they provide great insight into the variance of either forecast performance, coefficients or impulse responses across specifications, identifications and combinations. With all replication files readily available, there are only minimal barriers to interested readers either developing similar routines or advancing the ones outlined here using the replication files as a baseline. The programs are written with openly available, 'off the shelf' software with the intention of being easily adapted for use on local desktop computers, as opposed to high performance computers and clusters which were deliberately avoided (in to encourage replication). However, to maintain a tractable running time, parallel processing toolboxes and vectorization remains essential.

From a policy perspective, the results show the importance of providing further training and development opportunities to all research economists doing applied work. Internal and external training needs to be provided at a more advanced level in order to reduce the reliance on 'menu-driven' software, especially given the uncertainties regarding the data generating processes of such a large proportion of economic variables. From an academic perspective, at an undergraduate level, interdisciplinary relationships between computer science and economics need to be fostered in the form of joint honors programs (in the UK). At a postgraduate level, it is necessary to progressively develop taught level courses which focus on not only computational economics (currently revolved around agent based and DSGE models), but courses which focus on applications of statistical programming and data science to econometric problems.

It remains critical for economic stability that public and privately funded research

agendas continue to support econometric software development specifically focused on model selection and weighting. For example, the Economic and Social Research Council has a long history of funding software development in the social sciences. In the private sector, recent developments can be seen in popular econometric software in the form of some basic model averaging algorithms which go some small way to meeting the recognized end-user demand for such tools.[1] However, further work is required from all stakeholders, be they academic, policy practitioners or software developers themselves, to bring big-data tools focused on model specification, model selection, and model averaging to the fore.

Future work from the author of this thesis will focus on the intersection between econometrics and data science. One particularly promising project currently in its early stages is a direct result of Chapters 2 and, to a lesser extent Chapter 5. It focuses upon the development of user-friendly, front-end application which is able to dynamically interact with the ALFRED database of real-time data provided by the Federal Reserve. The software will then pair this accessibility with a suite of big data models which are dynamically benchmarked against a range of univariate models, outputting forecasts for any series available on ALFRED, and then, later extended to other central banks as more vintages become available.

It is hoped that if nothing else, this thesis highlights the importance of model identification, specification and combination in economics. While in several situations, simpler, more parsimonious models may perform most favorably under a range of criteria, technical and computational concerns are no longer a barrier to investigating situations where larger, more complex models may be optimal. This is especially relevant in the field of macroeconometrics, where identifying particular effects have, up until recently, been difficult to get a handle on.

[1] A notable example is the development of the Automatic ARIMA forecasting algorithm by Gareth Thomas, implemented into the recently released EViews 9.

APPENDIX A

A REPLICATION OF DUBIN (1998)

A.1 Introduction

In this appendix we replicate six of the two dimensional correlograms shown in Dubin (1998). Due to the fact that Chapter 3 is programmed exclusively in MATLAB, we provide MATLAB code in this Appendix, although EViews code is available on request to show the viability of programming spatial econometrics in any vector language, and in Section A.5 we provide the short code itself.

A.2 Distance matrix

We utilize the exact same simulated locations of each observation as per the original paper which can be seen in Equation A.1 (which is akin, albeit *significantly* smaller, to our distance matrices utilized in in Chapter 3), or in Table 1 of Dubin (1998).

$$D = \begin{bmatrix} 0 & 2.8 & 5.64 & 3.45 & 3.43 & 3.76 & 1.53 & 3.14 & 5.67 & 4.42 \\ 2.8 & 0 & 7.7 & 6.11 & 2.06 & 6.01 & 4.32 & 0.55 & 8.45 & 2.32 \\ 5.64 & 7.7 & 0 & 5.82 & 9 & 7.26 & 4.61 & 7.71 & 4.14 & 7.82 \\ 3.45 & 6.11 & 5.82 & 0 & 5.93 & 1.52 & 2.33 & 6.52 & 3.33 & 7.86 \\ 3.43 & 2.06 & 9 & 5.93 & 0 & 5.31 & 4.84 & 2.55 & 8.84 & 4.26 \\ 3.76 & 6.01 & 7.26 & 1.52 & 5.31 & 0 & 3.2 & 6.5 & 4.76 & 8.05 \\ 1.53 & 4.32 & 4.61 & 2.33 & 4.84 & 3.2 & 0 & 4.62 & 4.14 & 5.72 \\ 3.14 & 0.55 & 7.71 & 6.52 & 2.55 & 6.5 & 4.62 & 0 & 8.72 & 1.77 \\ 5.67 & 8.45 & 4.14 & 3.33 & 8.84 & 4.76 & 4.14 & 8.72 & 0 & 9.61 \\ 4.42 & 2.32 & 7.82 & 7.86 & 4.26 & 8.05 & 5.72 & 1.77 & 9.61 & 0 \end{bmatrix} \tag{A.1}$$

A.3 The model and matrices used

From A.1, we calculate three matrices as found in Dubin (1998). While Dubin (1998) calculates others in addition, such as nearest neighbour and maximum distance, we instead focus on the two specifications for inverse distances only. That is, similar to Section 3.3, we create three spatial weighting matrices (W) with elements specified as $W_{i,j} = 1/D_{i,j}^{P}$, where P=$\{1,2,3\}$.[1] As in Chapter 3 we are interested in a correlogram formed from the variance-covariance matrix, defined as:

$$V = \sigma^2(I - \lambda W)^{-1}(I - \lambda W^{\top})^{-1} \tag{A.2}$$

$$Corr_{i,j} = \frac{V_{i,j}}{\sqrt{(V_{i,i}, V_{j,j}}} \tag{A.3}$$

Where W is substituted for three combinations of P and two of $\lambda = \{0.33, 0.67\}$

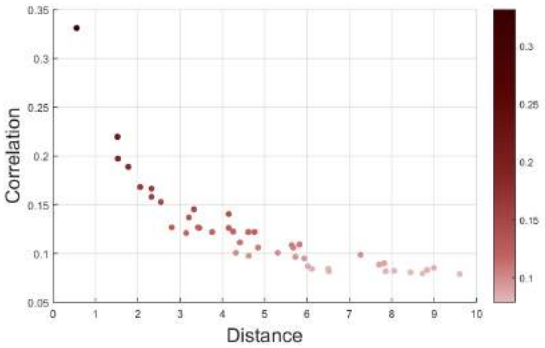
A.4 Six outputs which are identical to Dubin (1998)

The first three panels (shown in blue) of Figure A.1 correspond to (and replicate exactly) B1-B3 on page 315 of Dubin (1998), and the latter three panels (shown in red) relate to Figures A1-A3 on page 314.

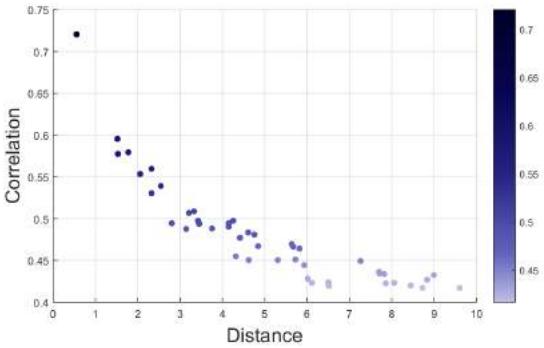
[1] This is using the notation of Dubin (1998), where P is equivalent to α in Chapter 3.

Figure A.1: Replication of Figures A1-B3 on page 314-315 of Dubin (1998)

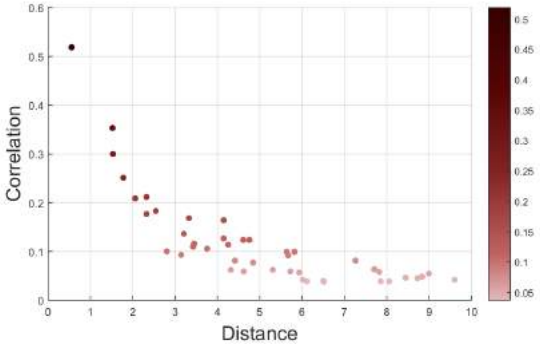
A.1.1 $\lambda = 0.33$, p = 1



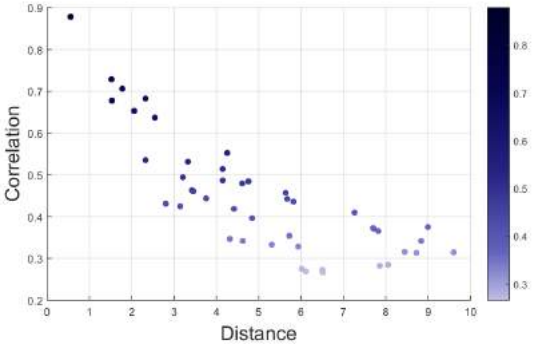
A.1.2 $\lambda = 0.67$, p = 1



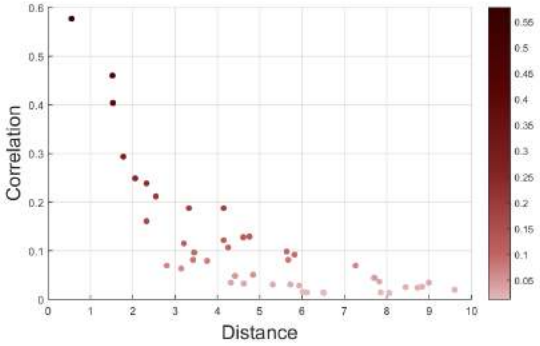
A.1.3 $\lambda = 0.33$, p = 2



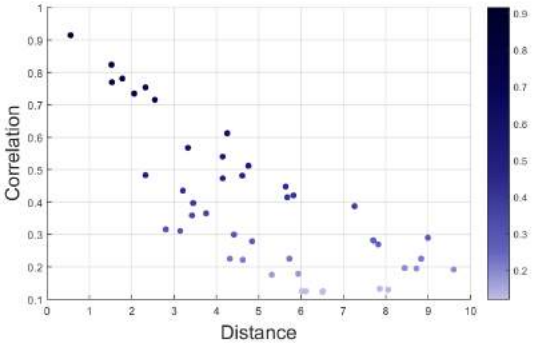
A.1.4 $\lambda = 0.67$, p = 2



A.1.5 $\lambda = 0.33$, p = 3



A.1.6 $\lambda = 0.67$, p = 3



A.5 The replication code

```
cd C:\<path>;
Distance=load(['durbindistance.txt'],'-ascii');
[nc nr]=size(Distance);
number=0
for l=1:2;
    for p=1:3;
        number=number+1
        for a = 1:nc;
            for b=a:nc;
                if a~=b;
                   weight(a,b)=1/(Distance(a,b)^p);
                   weight(b,a)=weight(a,b);
                end;
            end;
        end;
        for a=1:nc;
            rowsum=sum(weight(a,:));
            for b=1:nc;
                if a~=b;
                    weight(a,b)=weight(a,b)/rowsum;
                end
            end
        end
        if l==1;
            vmatrix=inv(eye(nr)-(0.33*weight))*inv(eye(nr)-...
            (0.33*transpose(weight)));
        else
            vmatrix=inv(eye(nr)-(0.67*weight))*inv(eye(nr)-...
            (0.67*transpose(weight)));
        end
        for a=1:nr
            for b=a:nr
                cmatrix(a,b)=vmatrix(a,b)/...
                ((vmatrix(a,a)*vmatrix(b,b))^0.5);
            end
        end
        c=0;
        for a=1:nr
            for b=a:nr
                if a~=b
                    c=c+1;
                    output(c,2)=cmatrix(a,b);
                    output(c,1)=Distance(a,b);
                end
             end
        end
        h=scatter(output(:,1),output(:,2),400,output(:,2),'.')
        colormap(copper)
        set(get(colorbar,'ylabel'),'string','','fontsize',18)
        axis square
        xlabel('Distance','Fontsize',24);
        ylabel('Correlation','Fontsize',24);
        set(gcf,'color','w')
        grid on
        saveas(h,sprintf('spatial_dubin_%d.png',number));
        close all
    end
end
```

APPENDIX B

A GUIDE TO THE STATFACT EVIEWS ADD-IN

For a direct link to the EViews .aipz installation file, please see

http://www.eviews.com/Addins/StatFact.aip.

For a link to the forum post discussing this add-in can, please see

http://forums.eviews.com/viewtopic.php?f=23&t=12047&p=42131#p42131.

This work is gratefully reproduced under license number 3676441294414 granted by Springer Link, where it is published as: Rahal, C. (2015), 'A Guide to the StatFact EViews Add-in', *Computational Economics*, pp.1-6, doi::10.1007/s10614-015-9507-6.

B.1 Introduction

This work is a response to the increasing acceptance of factor models in time series econometrics as a viable method of dimension reduction in the 'big data' paradigm. In particular, a noticeable demand exists in the econometric community for a front-end interface which makes available some of the routines discussed within prominent papers in the factor literature from recent years. This add-in aims to fill that gap by providing options on how to transform data and subsequently extract stationary factors in the style of Stock and Watson (2002a). The add-in then calculates six information criteria (PC1, PC2, PC3, IPC1, IPC2, IPC3, AIC3, BIC3) based on the seminal paper of Bai and Ng (2002), which guide the choice of factors typically used in the second stage of empirical applications. Such applications are predicated on the work (in a forecasting context for instance) not only of Stock and Watson (2002a), but Bernanke *et al.* (2005) in a vector autoregressive framework, which is then generalized to an error correction context in Banerjee *et al.* (2014). Following this, the routine outputs a series of (optional) graphics to allow further analysis of the extracted factors. It is hoped that this add-in will contribute to the field by allowing an easily accessible method of computing the factors, made possible by the increasing availability of 'macro-panel' style datasets, especially within a developing economy context.

B.2 Econometric procedure

The common factors are estimated by the method of asymptotic principal components (the maximum number of which is $min\{N, T\}$). The estimated factor matrix ($T \times k$, where k is the number of factors), $\tilde{F}^k$, is $\sqrt{T}$ times the eigenvectors which correspond to the k largest eigenvalues of the $T \times T$ matrix XX' when $T < N$. Alternatively, when $T > N$, the factor matrix is calculated as $\bar{F}^k = \frac{X\bar{\Lambda}^k}{N}$, whereas the loadings matrix ($\bar{\Lambda}^k$) is calculated as $\sqrt{N}$ multiplied by the eigenvectors corresponding to the k largest eigenvalues of $X'X$. Please see Stock and Watson (2002a) for more details. The series

which form the group which is then converted into matrix $X'X$ or XX' are also optionally first differenced and/or standardized (mean subtracted and divided by series standard deviation), as per the Gauss replication files of Igor Maston associated with Banerjee *et al.* (2014). Six criteria proposed by Bai and Ng (2002) are optionally calculated. Their general form is: $PC(k) = V(k, \hat{F}^k) + kg(N, T)$, where $V(k, \hat{F}^k)$ is the sum of squared residuals from estimating k factors (in a dynamic factor model representation), $g(N, T)$ is a necessary penalty function, and we are interested in estimating the most suitable dimension ($r \leq kmax$) of $\hat{F}$.

B.3 Use of the software

The .aipz associated with the add-in is a self-installing file (compatible with versions of EViews 7.1 or later) which can be accessed either through the EViews website[1] or directly through the client itself. The add-in can be executed through both the command line and a graphical user interface (GUI — through suitable group objects), and it is possible to edit the .prg to run as a standalone script through the program interface to be compatible with earlier versions of the statistical package. The command line syntax is:

group_object.statfact(options) *factornames* #

and the various options are described below and summarized in Table B.1. Note that *'group_object'* refers to the name of the input, a group object, and the output is denoted as *factornames* which is a user-defined prefix to create series objects called *factornames_1*,...,*factornames_#*, where # refers to the maximum number of factors for consideration ('*kmax*' in the literature). Given the nature of the procedure, the group cannot be defined in a panel workfile (our examples below use both a dated (regular frequency) and undated workfile structure. The number of elements in *group_object* must naturally be less than # (which must in turn be a positive number, which is automatically rounded to an integer), with elements containing no missing values.

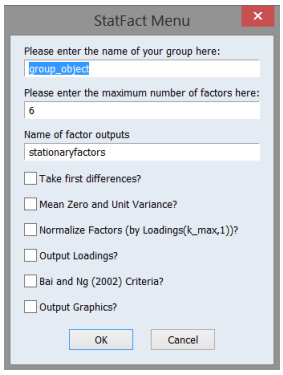
[1] Available at http://www.eviews.com/Addins/StatFact.aipz.

Table B.1: A summary of the input options

Option Syntax	Optional Output Created
ipc	Reports the information criteria outlined based on Bai and Ng (2002).
d	Takes first difference of all the series in the group if required.
m	Standardizes the series: subtracts the mean and divides by st. dev.
l	Outputs the loadings, denoted as *loading_1* ,. . .,*loading_#*.
n	Normalizes factors\loadings by element (#,1) of the loadings matrix.
g	Creates four graphical outputs, as described below.

The GUI is an analogous method of specifying each of the six options in Table B.1 in addition to # (*kmax*) and the name of the input and output group and series names respectively, and can be seen in Figure B.1. The ordering of the operations is such that the add-in first de-means and divides by the standard deviation (if selected), then takes first differences of all series in the group. It then calculates the factors and loadings (if the latter option is selected) and then from these factors, the Bai and Ng (2002) criteria are calculated to create an outputs table (as seen in Table B.2 in Section B.4.2). The final operation is to calculate the graphs, shown in Figure B.2, which show the factors and loadings themselves, a scree plot (eigenvalues against factor number) and the percentage of variance explained (eigenvalues divided by sum of all eigenvalues plotted against corresponding factor).

Figure B.1: The GUI representation of the command line syntax



B.4 Illustrative examples

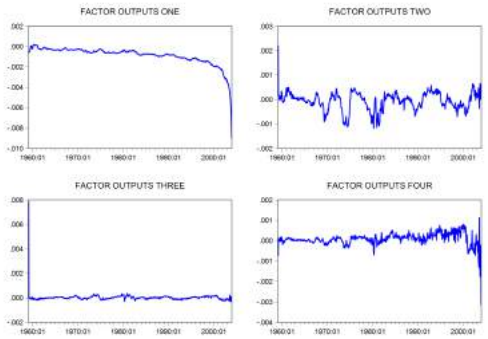
For both of our empirical examples, the datasets are provided in the accompanying .zip file which also contains two (very) short programs to perform the procedures featured in Sections B.4.1 and B.4.2.

B.4.1 Stock and Watson (2005)

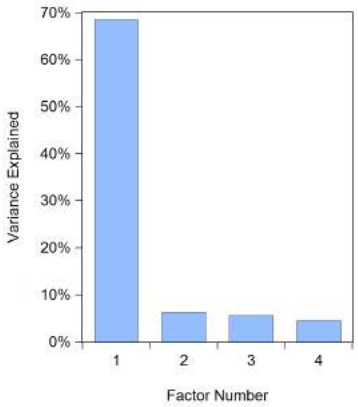
The first example takes the dataset made available from the replication files of Banerjee *et al.* (2014) which is a replica of the data used in Stock and Watson (2005), containing 132 monthly macroeconomic series from 1959M1 to 2003M12. Our script imports the data and 'T-Codes' separately, and (after adjusting for outliers,) transforms the series as applicable. The graphical outputs can be seen in Figure B.2 for $kmax = 6$.

Figure B.2: Graphical output from Stock and Watson (2005) data

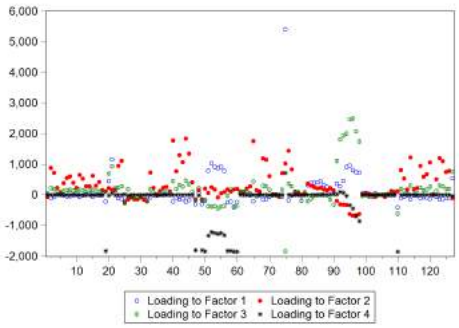
B.2.1 Factors



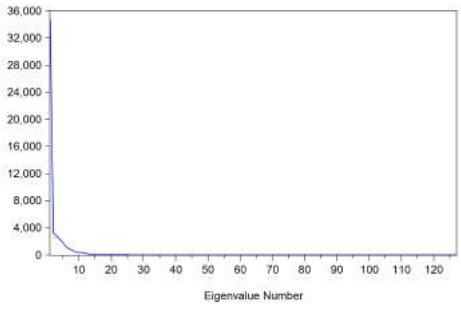
B.2.2 Percent Variance Explained



B.2.3 Loadings



B.2.4 Scree Plot: Stationary Variable



B.4.2 Hurlin dataset

As a second example (this time using an undated workfile), we use the dataset provided by Christophe Hurlin in his Bai and Ng (2002) replication .m files. In this example, we focus on presenting the IPC output of our code, which can be seen to match with the corresponding MATLAB outputs. The dataset is a cross-section of 25×40 dimension containing real GDP growth rates from the OECD (GPDVD). The data is in volume at 2000 PPP, in USD, rebased to 100 in 2000, measured between 1963 and 2003, and is centred.[2]

Table B.2: Bai and Ng (2002) output for the Hurlin dataset

	PC1	PC2	PC3	IPC1	IPC2	IPC3	AIC3	BIC3
6.0000	0.0004	0.0004	0.0004	−7.4583	−7.2690	−7.7518	0.0003	0.0007
5.0000	0.0004	0.0004	0.0004	−7.4437	−7.2859	−7.6883	0.0004	0.0007
4.0000	0.0004	0.0005	0.0004	−7.4352	−7.3090	−7.6308	0.0004	0.0006
3.0000	0.0005	0.0005	0.0004	−7.4181	−7.3234	−7.5648	0.0004	0.0006
2.0000	0.0005	0.0005	0.0005	−7.4031	−7.3400	−7.5009	0.0005	0.0006
1.0000	0.0005	0.0006	0.0005	−7.4049	−7.3734	−7.4538	0.0005	0.0006
Optimal	6.0000	6.0000	6.0000	6.0000	1.0000	6.0000	6.0000	1.0000

B.5 Conclusion and further work

This appendix detailed the development, the execution of, and examples related to stationary factor extraction as part of the recently published StatFact add-in. The software is under constant revision and extension, and users can easily request any modifications or report bugs as part of a dynamic development process. The availability of an easy to use GUI will hopefully further encourage researchers to consider this method of dimension reduction as part of their modelling strategy as larger macro-panels are created with the increased reporting and collection of the underlying series. Given the prevalent use of EViews as a teaching tool at undergraduate, postgraduate and executive level, it is hoped that the examples documented within this appendix and the supplementary mate-

[2]The data for Section B.4.2 are available as part of the replication files of this guide, and .m files at http://www.runmycode.org/coder/view/151.

rial might act as an introductory guide for interactive educational purposes. An obvious extension from a software development perspective would be to consider the dynamic factor approach, as discussed in papers such as Forni *et al.* (2005).

APPENDIX C

A GUIDE TO THE MACROTRANS EVIEWS ADD-IN

For a direct link to the EViews .aipz installation file, please see

http://www.eviews.com/Addins/MacroTrans.aipz.

For a link to the forum post discussing this add-in can, please see

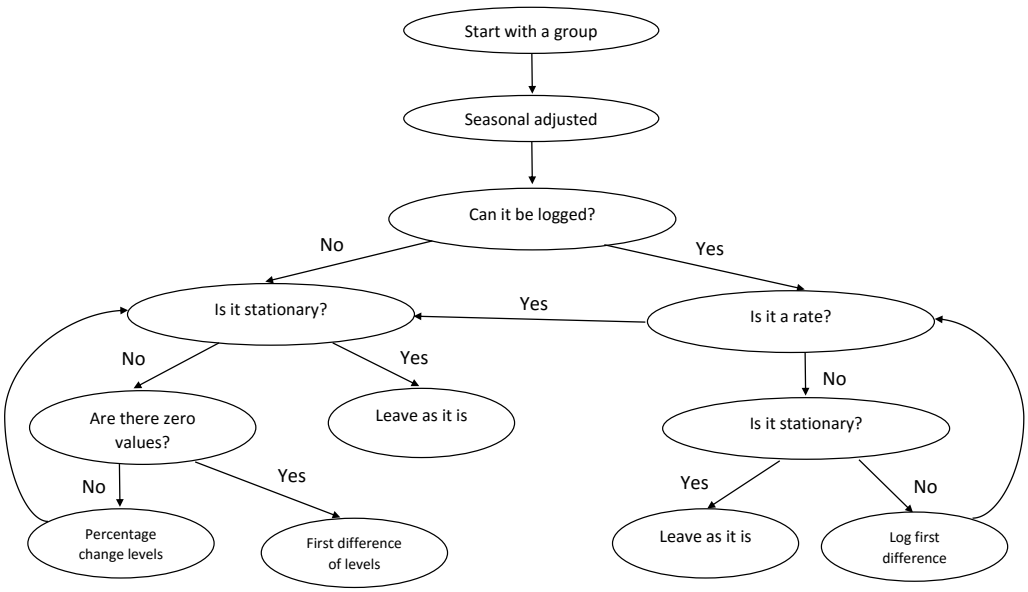
http://forums.eviews.com/viewtopic.php?f=23&t=12381.

C.1 Introduction

This is a short guide to the MacroTrans add-in. The add-in operates on group objects from either the command line or group proc, automatically adjusting group members (series) one at a time in convention with the prominent macroeconometric literature. Designed for quarterly or monthly workfiles, it provides options on seasonal adjustment, the inclusion of interest/discount rates, a choice of unit root tests, and logarithmic transformations. It is designed to pre-processes series for 'big data' modelling (i.e. factors/LBVARs/combination algorithms), although it can also be used for smaller, 'medium' sized models.

The add-in expands on the **GroupX12** add-in, and takes **group** objects only and operates on them either through the command line or group proc. It takes each series and systematically transforms them ready for macroeconometric modelling, consistent with the 'T-Codes'/transformations used in prominent macro modelling papers (and,

Figure C.1: The schematic of transformations



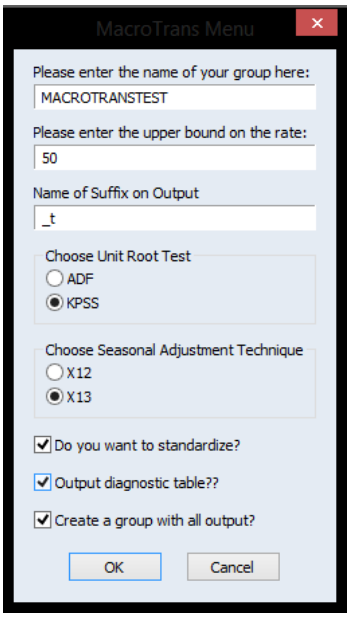
notably the factor literature, such as Stock and Watson (2005)). Details of the procedure can be seen in Figure C.1.

The add-in takes a group object, and if required, checks for and seasonally adjusts each of the series using either the X-12 or X-13 (EViews 8 or later) algorithms developed by the United States Census Bureau. If the series can be logged (no negative or zero values), *and* it is not solely bounded between zero and the rate max (i.e. - it is not already an interest rate or bond yield), then it is logged. It then undertakes tests for stationarity, using either the KPSS or ADF tests. The maximum number of differencing is set to 2 (to change this, set !maxdiffs in macrotrans.prg), and the significance level on unit root tests is set to 5% (change `!sig` to 0,1,2 and `!sigval` to 0.01, 0.05 and 0.1 for 1%, 5% and 10% respectively). If the series is in logarithms and it is nonstationary, it is differenced, if it is not, and percentages can be taken (no zero values in constituent series), then they are, or, if they cannot, then the series is differenced.

C.2 The command line and GUI

The options are as follows, and can be seen in Figure C.2 and Table C.1 below. It is necessary to specify a suffix for the transformed output series (with the command line option `suffix=''_<yoursuffix>''`, or in the edit box in the GUI, defaults to '_t'), and specify whether you would like all these series to be placed in a group together (include the term 'groupout' in the command line options). Of critical importance is the inclusion of a positive `rate=<yourrate>` option. This is to allow for the existence of things like bond yields or discount rates which should not be in logarithms (and in further developments of this add-in, may be subject to fractional integration tests). Other options include the choice of seasonal unit root test as mentioned above, with the inclusion of the options `seas=''X13''` or `seas=''X12''` and the choice of unit root test is introduced through the options `uroot=''adf''` or `uroot=''kpss''`. Finally, the inclusion of the option **standardize** standardizes the output series (subtracts the mean, divides by standard

Figure C.2: The GUI representation of the command line syntax



deviation), and the use of the option **tableout** creates a diagnostic table which shows the percent of series which are seasonally adjusted (as X-12 and X-13 sometimes encounter errors due to erroneous features of the series), the percent of series which have at least one unit root, the percent of series which are made stationary through a percentage change, and the percent of series which are logged. The table will be called *macrotrans_table01* if this name is not already taken. Note that the group can be untitled, the add-in cannot be run on anything other than monthly or quarterly dated workfiles, that the constituent series cannot have any missing values, that the specified rate must be positive.

Table C.1: A summary of the input options

Option Syntax	Optional Output Created
rate=X	Sets upper bound for interest rate/yield series.
suffix="_yoursuffix"	Declares the suffix for output series
unit="yourtest"	Specifies unit root test ('adf' or 'kpss')
seas="yourseas"	Specifies seasonal adjustment ('X12' or 'X13')
standardize	Standardizes all output
tableout	Creates an output diagnostic table
groupout	Creates an output group with all transformed series

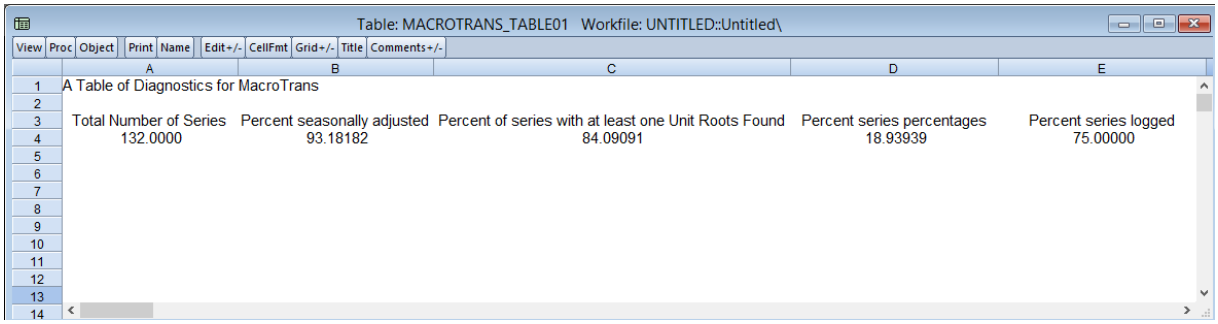
C.3 An example using the Stock and Watson (2005)

As an example, we take the dataset used in Stock and Watson (2005), made available in the replication files of Banerjee *et al.* (2014). We form a group of the 132 series and call it *macrotranstestgroup*. There are a few commented out lines of the *macrotrans.prg* which is dedicated to importing this data and creating the group. Using the command line interface, we type:

```
macrotranstestgroup.macrotrans(rate=50,unit="kpss",seas="X13",
suffix="_t",tableout,groupout)
```

As described above, this will give us transformed output series of the original name with a new suffix *_t* (while retaining the original series also), an output group called macrotranstestgroup_t, the upper bound on rates set to 50, unit root tests done using the KPSS test, seasonal adjustment completed using the X-13 algorithm, and a diagnostics table called macrotrans_table01 which can be seen in Figure C.3, below:

Figure C.3: An example using the Stock and Watson (2005) dataset



C.4 Conclusion and further work

If the command line syntax is used and does not have a specification for seasonal adjustment or unit root tests, it defaults to KPSS/X-13, although the code is easily modifiable to omit both procedures altogether (change `!urootnumber` and `!seasonalnumber` on lines

64 and 65 of the .prg respectively). As mentioned above, an obvious extension of the test is to incorporate different unit root or seasonal adjustment tests.

APPENDIX

BIBLIOGRAPHY

Adelman, I. (2007) The research for the paper on the dynamics of the Klein-Goldberger model. **Journal of Economic and Social Measurement**, 32: 29–33

Adler, J. (2010) **R in a Nutshell: A Desktop Quick Reference**. O'Reilly Media. URL https://books.google.co.uk/books?id=05-fSBScR_4C.

Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle." In Petrov, B.N. and Csaki, F. (eds.) **Second International Symposium on Information Theory**. Akadémiai Kiado. pp. 267–281

Anselin, L. (1988) **Spatial Econometrics: Methods and Models**. Studies in Operational Regional Science, Springer Netherlands. URL https://books.google.co.uk/books?id=G47tCAAAQBAJ.

Antipov, E. and Pokryshevskaya, E. (2010) Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. MPRA Paper 27645, University Library of Munich, Germany. URL https://ideas.repec.org/p/pra/mprapa/27645.html.

Aoki, K., Proudman, J. and Vlieghe, G. (2002) Houses as collateral: has the link between house prices and consumption in the UK changed? **Economic Policy Review**, (May): 163–177

Aoki, K., Proudman, J. and Vlieghe, G. (2004) House prices, consumption, and monetary policy: a financial accelerator approach. **Journal of Financial Intermediation**, 13 (4): 414–435. URL http://ideas.repec.org/a/eee/jfinin/v13y2004i4p414-435.html.

Assenmacher-Wesche, K. and Gerlach, S. (2008) Monetary policy, asset prices and macroeconomic conditions: A panel-VAR study. Working Paper Research 149, National Bank of Belgium.

Bai, J. (2004) Estimating cross-section common stochastic trends in nonstationary panel data. **Journal of Econometrics**, 122 (1): 137–183. URL http://ideas.repec.org/a/eee/econom/v122y2004i1p137-183.html.

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

Bai, J. and Ng, S. (2008) Forecasting economic time series using targeted predictors. **Journal of Econometrics**, 146 (2): 304–317. URL http://ideas.repec.org/a/eee/econom/v146y2008i2p304-317.html.

Bai, J. and Ng, S. (2009) Boosting diffusion indices. **Journal of Applied Econometrics**, 24 (4): 607–629. URL http://ideas.repec.org/a/jae/japmet/v24y2009i4p607-629.html.

Banbura, M., Giannone, D. and Reichlin, L. (2010) Large Bayesian vector auto regressions. **Journal of Applied Econometrics**, 25 (1): 71–92. URL http://ideas.repec.org/a/jae/japmet/v25y2010i1p71-92.html.

Banerjee, A. and Marcellino, M. (2006) Are there any reliable leading indicators for US inflation and GDP growth? **International Journal of Forecasting**, 22 (1): 137–151. URL http://ideas.repec.org/a/eee/intfor/v22y2006i1p137-151.html.

Banerjee, A. and Marcellino, M. (2008) Factor-augmented error correction models. CEPR Discussion Papers 6707, C.E.P.R. Discussion Papers. URL http://ideas.repec.org/p/cpr/ceprdp/6707.html.

Banerjee, A., Marcellino, M. and Masten, I. (2005) Leading indicators for euro-area inflation and GDP growth. **Oxford Bulletin of Economics and Statistics**, 67 (s1): 785–813. URL http://ideas.repec.org/a/bla/obuest/v67y2005is1p785-813.html.

Banerjee, A., Marcellino, M. and Masten, I. (2008) Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. (334). URL http://EconPapers.repec.org/RePEc:igi:igierp:334.

Banerjee, A., Marcellino, M. and Masten, I. (2014) Forecasting with factor-augmented error correction models. **International Journal of Forecasting**, 30 (3): 589–612

Banternghansa, C. and McCracken, M.W. (2011) Real-time forecast averaging with ALFRED. **Review**, (Jan): 49–66. URL http://ideas.repec.org/a/fip/fedlrv/y2011ijanp49-66nv.93no.1.html.

Barry, R.P. and Pace, R.K. (1999) Monte Carlo estimates of the log determinant of large sparse matrices. **Linear Algebra and its Applications**, 289 (13): 41 – 54

Bates, J. and Granger, C. (1969) The combination of forecasts. **Operations Research Quarterly**, 20: 451–468

Baumeister, C. and Kilian, L. (2011) Real-time forecasts of the real price of oil. **Journal of Business & Economic Statistics**, 30 (2): 326–336. URL http://ideas.repec.org/a/taf/jnlbes/v30y2011i2p326-336.html.

Bean, L.H. (1929) A simplified method of graphic curvilinear correlation. **Journal of the American Statistical Association**, 24 (168): pp. 386–397. URL http://www.jstor.org/stable/2277084.

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

Bernanke, B.S. and Boivin, J. (2003) Monetary policy in a data-rich environment. **Journal of Monetary Economics**, 50 (3): 525–546. URL http://ideas.repec.org/a/eee/moneco/v50y2003i3p525-546.html.

Bhattacharjee, A. and Jensen-Butler, C. (2013) Estimation of the spatial weights matrix under structural constraints. **Regional Science and Urban Economics**, 43 (4): 617 – 634. URL http://www.sciencedirect.com/science/article/pii/S0166046213000288.

Boeckx, J., Dossche, M. and Peersman, G. (2014) Effectiveness and transmission of the ECB's balance sheet policies. Tech. rep.

Bogdon, A.S. and Can, A. (1997) Indicators of local housing affordability: Comparative and spatial approaches. **Real Estate Economics**, 25 (1): 43–80. URL http://dx.doi.org/10.1111/1540-6229.00707.

Boivin, J. and Ng, S. (2006) Are more data always better for factor analysis? **Journal of Econometrics**, 132 (1): 169–194. URL http://ideas.repec.org/a/eee/econom/v132y2006i1p169-194.html.

Bork, L. and Møller, S.V. (2012) Housing price forecastability: A factor analysis. CREATES Research Papers 2012-27, School of Economics and Management, University of Aarhus. URL http://ideas.repec.org/p/aah/create/2012-27.html.

Bork, L. and Møller, S.V. (2015) Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. **International Journal of Forecasting**, 31 (1): 63 – 78. URL http://www.sciencedirect.com/science/article/pii/S0169207014000958.

Bracke, P. (2014) Micro evidence from a matched dataset in Central London. **Real Estate Economics, Forthcoming**.

Brown, J.P., Song, H. and McGillivray, A. (1997) Forecasting UK house prices: A time varying coefficient approach. **Economic Modelling**, 14 (4): 529–548. URL http://ideas.repec.org/a/eee/ecmode/v14y1997i4p529-548.html.

Bruyne, K.D. and Hove, J.V. (2013) Explaining the spatial variation in housing prices: an economic geography approach. **Applied Economics**, 45 (13): 1673–1689. URL http://ideas.repec.org/a/taf/applec/45y2013i13p1673-1689.html.

Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model selection: An integral part of inference. **Biometrics**, 53 (2): 603–618. URL http://dx.doi.org/10.2307/2533961.

Buiter, W.H. (2010) Housing wealth isn't wealth. **Economics - The Open-Access, Open-Assessment E-Journal**, 4: 1–29. URL http://ideas.repec.org/a/zbw/ifweej/201022.html.

Burridge, P. (1981) Testing for a common factor in a spatial autoregression model. **Environment and Planning A**, 13 (7): 795–800. URL http://epn.sagepub.com/content/13/7/795.abstract.

Can, A. (1990) The measurement of neighborhood dynamics in urban house prices. **Economic Geography**, 66 (3): pp. 254–272. URL http://www.jstor.org/stable/143400.

Can, A. (1992) Specification and estimation of hedonic housing price models. **Regional Science and Urban Economics**, 22 (3): 453 – 474. URL http://www.sciencedirect.com/science/article/pii/0166046292900394. Special Issue Space and Applied Econometrics.

Can, A. and Megbolugbe, I. (1997) Spatial dependence and house price index construction. **The Journal of Real Estate Finance and Economics**, 14 (1-2): 203–22. URL http://EconPapers.repec.org/RePEc:kap:jrefec:v:14:y:1997:i:1-2:p:203-22.

Carstensen, K., Hlsewig, O. and Wollmershuser, T. (2009) Monetary policy transmission and house prices: European cross-country evidence. Tech. rep.

Case, K.E., Quigley, J.M. and Shiller, R.J. (2005) Comparing wealth effects: The stock market versus the housing market. **The B.E. Journal of Macroeconomics**, 0 (1): 1

Case, K.E., Quigley, J.M. and Shiller, R.J. (2013) Wealth effects revisited: 1975-2012. Working Paper 18667, National Bureau of Economic Research. URL http://www.nber.org/papers/w18667.

Case, K.E. and Shiller, R.J. (1990) Forecasting prices and excess returns in the housing market. **Real Estate Economics**, 18 (3): 253–273. URL http://dx.doi.org/10.1111/1540-6229.00521.

Castle, J.L., Clements, M.P. and Hendry, D.F. (2013) Forecasting by factors, by variables, by both or neither? **Journal of Econometrics**, 177 (2): 305–319. URL http://ideas.repec.org/a/eee/econom/v177y2013i2p305-319.html.

Cecchetti, S.G., Chu, R.S. and Steindel, C. (2000) The unreliability of inflation indicators. **Current Issues in Economics and Finance**, (Apr). URL http://ideas.repec.org/a/fip/fednci/y2000iaprnv.6no.4.html.

Chen, L., Dolado, J.J. and Gonzalo, J. (2014) Detecting big structural breaks in large factor models. **Journal of Econometrics**, 180 (1): 30–48. URL http://ideas.repec.org/a/eee/econom/v180y2014i1p30-48.html.

Cheng, X. and Hansen, B.E. (2015) Forecasting with factor-augmented regression: A frequentist model averaging approach. **Journal of Econometrics**, 186 (2): 280 – 293. URL http://www.sciencedirect.com/science/article/pii/S0304407615000342. High Dimensional Problems in Econometrics.

Chevillon, G. and Hendry, D.F. (2005) Non-parametric direct multi-step estimation for forecasting economic processes. **International Journal of Forecasting**, 21 (2): 201–218. URL http://ideas.repec.org/a/eee/intfor/v21y2005i2p201-218.html.

Chow, G.C. and Lin, A.l. (1971) Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. **The Review of Economics and Statistics**, 53 (4): 372–75

Clark, T.E. and McCracken, M.W. (2010) Averaging forecasts from VARs with uncertain instabilities. **Journal of Applied Econometrics**, 25 (1): 5–29. URL http://ideas.repec.org/a/jae/japmet/v25y2010i1p5-29.html.

Clemen, R.T. (1989) Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, 5 (4): 559–583. URL http://ideas.repec.org/a/eee/intfor/v5y1989i4p559-583.html.

Cliff, A. and Ord, K. (1969) Spatial autocorrelation. Pion, 1st Edition.

Court, A. (1939) **Hedonic Price Indexes with Automotive Examples**. URL http://books.google.co.uk/books?id=NCoVtwAACAAJ.

Crawford, G.W. and Fratantoni, M.C. (2003) Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. **Real Estate Economics**, 31 (2): 223–243

Crawford, G.W. and Rosenblatt, E. (1995) Efficient mortgage default option exercise: Evidence from loss severity. **Journal of Real Estate Research**, 10 (5): 543–556. URL http://ideas.repec.org/a/jre/issued/v10n51995p543-556.html.

Das, S., Gupta, R. and Kabundi, A. (2009) Could we have predicted the recent downturn in the South African housing market? **Journal of Housing Economics**, 18 (4): 325–335

Das, S., Gupta, R. and Kabundi, A. (2010) The blessing of dimensionality in forecasting real house price growth in the nine census divisions of the US. 19 (1), Journal of Housing Research.

Davis, M. and Heathcote, J. (2001) Housing and the business cycle. Tech. rep.

Diebold, F.X. (2012) Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. Working Paper 18391, National Bureau of Economic Research. URL http://www.nber.org/papers/w18391.

Doornik, J.A., Hendry, D.F. and Shephard, N. (2002) Computationally-intensive econometrics using a distributed matrix-programming language. (2001-W22). URL http://ideas.repec.org/p/nuf/econwp/0122.html.

Dubé, J. and Legros, D. (2013a) Dealing with spatial data pooled over time in statistical models. **Letters in Spatial and Resource Sciences**, 6 (1): 1–18. URL http://dx.doi.org/10.1007/s12076-012-0082-3.

Dubé, J. and Legros, D. (2013b) A spatio-temporal measure of spatial dependence: An example using real estate data. **Papers in Regional Science**, 92 (1): 19–30. URL http://ideas.repec.org/a/bla/presci/v92y2013i1p19-30.html.

Dubé, J. and Legros, D. (2014a) Spatial econometrics and spatial data pooled over time: Towards an adapted modeling approach. **Journal of Real Estate Literature**, 22 (1): 103–125

Dubé, J. and Legros, D. (2014b) Spatial econometrics and the hedonic pricing model: what about the temporal dimension? **Journal of Property Research**, 31 (4): 333–359. URL http://ideas.repec.org/a/taf/jpropr/v31y2014i4p333-359.html.

Dubé, J. and Legros, D. (2015) Modeling spatial data pooled over time: Schematic representation and monte carlo evidences. **Theoretical Economics Letters**, 5: 132–154

Dubin, R.A. (1998) Spatial autocorrelation: A primer. **Journal of Housing Economics**, 7 (4): 304–327. URL http://EconPapers.repec.org/RePEc:eee:jhouse:v:7:y:1998:i:4:p:304-327.

Eickmeier, S. and Ziegler, C. (2008) How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. **Journal of Forecasting**, 27 (3): 237–265. URL http://ideas.repec.org/a/jof/jforec/v27y2008i3p237-265.html.

Elbourne, A. (2008) The UK housing market and the monetary policy transmission mechanism: An SVAR approach. **Journal of Housing Economics**, 17 (1): 65–87

Faust, J. (1998) The robustness of identified VAR conclusions about money. **Carnegie-Rochester Conference Series on Public Policy**, 49: 207 – 244. URL http://www.sciencedirect.com/science/article/pii/S0167223199000093.

Faust, J. and Wright, J.H. (2009) Comparing Greenbook and reduced form forecasts using a large realtime dataset. **Journal of Business & Economic Statistics**, 27 (4): 468–479. URL http://ideas.repec.org/a/bes/jnlbes/v27i4y2009p468-479.html.

Fik, T.J., Ling, D.C. and Mulligan, G.F. (2003) Modeling spatial variation in housing prices: A variable interaction approach. **Real Estate Economics**, 31 (4): 623–646. URL http://dx.doi.org/10.1046/j.1080-8620.2003.00079.x.

Fletcher, M., Mangan, J. and Raeburn, E. (2004) Comparing hedonic models for estimating and forecasting house prices. **Property Management**, 22 (3): 189–200. URL http://www.emeraldinsight.com/doi/abs/10.1108/02637470410544986.

Forni, M., Hallin, M., Lippi, M. *et al.* (2005) The generalized dynamic factor model: One-sided estimation and forecasting. **Journal of the American Statistical Association**, 100 (471): pp. 830–840. URL http://www.jstor.org/stable/27590616.

Foster, C. and Van Order, R. (1985) An option based model of mortgage default. **Housing Finance Review**, 3:4: 351–372

Fry, R. and Pagan, A. (2011) Sign restrictions in structural vector autoregressions: A critical review. **Journal of Economic Literature**, 49 (4): 938–60

Fuster, A. and Willen, P.S. (2010) $1.25 Trillion is still real money: some facts about the effects of the Federal Reserves mortgage market investments. Tech. rep.

Gabriel, S. and Lutz, C. (2014) The impact of unconventional monetary policy on real estate markets. **Available at SSRN: http://ssrn.com/abstract=2493873**.

Gambacorta, L., Hofmann, B. and Peersman, G. (2014) The effectiveness of unconventional monetary policy at the zero lower bound: a cross-country analysis. **Journal of Money, Credit and Banking**, 46 (4): 615–642

Garratt, A., Koop, G. and Vahey, S. (2008) Forecasting substantial data revisions in the presence of model uncertainty. **Economic Journal**, 118 (530): 1128–1144. URL http://ideas.repec.org/a/ecj/econjl/v118y2008i530p1128-1144.html.

Garriga, C., Kydland, F.E. and Sustek, R. (2013) Mortgages and monetary policy. NBER Working Papers 19744, National Bureau of Economic Research, Inc.

Gattini, L. and Hiebert, P. (2010) Forecasting and assessing euro area house prices through the lens of key fundamentals. Working Paper Series 1249, European Central Bank. URL http://ideas.repec.org/p/ecb/ecbwps/20101249.html.

Gau, G.W. (1984) Weak form tests of the efficiency of real estate investment markets. **The Financial Review**, 19 (4): 301–20. URL http://ideas.repec.org/a/bla/finrev/v19y1984i4p301-20.html.

Giuliodori, M. (2005) The role of house prices in the monetary transmission mechanism across european countries. **Scottish Journal of Political Economy**, 52 (4): 519–543. URL http://dx.doi.org/10.1111/j.1467-9485.2005.00354.x.

Goldfeld, S.M. and Quandt, R.E. (1973) A Markov model for switching regressions. **Journal of Econometrics**, 1 (1): 3–15. URL http://ideas.repec.org/a/eee/econom/v1y1973i1p3-15.html.

Goodhart, C. and Hofmann, B. (2008) House prices, money, credit, and the macroeconomy. **Oxford Review of Economic Policy**, 24 (1): 180–205

Granger, C. and Jeon, Y. (2004) Forecasting performance of information criteria with many macro series. **Journal of Applied Statistics**, 31 (10): 1227–1240. URL http://dx.doi.org/10.1080/0266476042000285495.

Guirguis, H., Giannikos, C. and Anderson, R. (2004) The U.S. housing market: Asset pricing forecasts using time varying coefficients. **The Journal of Real Estate Finance and Economics**, 30 (1): 33–53

Gupta, R. (2013) Forecasting house prices for the four census regions and the aggregate us economy in a data-rich environment. Tech. Rep. 33.

Gupta, R. and Das, S. (2008) Spatial Bayesian methods of forecasting house prices in six metropolitan areas of South Africa. **South African Journal of Economics**, 76 (2): 298–313

Gupta, R. and Kabundi, A. (2010) The effect of monetary policy on house price inflation. **Journal of Economic Studies**, 37 (6): 616–626. URL http://dx.doi.org/10.1108/01443581011086657.

Gupta, R., Kabundi, A. and Miller, S.M. (2011a) Forecasting the US real house price index: Structural and non-structural models with and without fundamentals. **Economic Modelling**, 28 (4): 2013–2021

Gupta, R., Kabundi, A. and Miller, S.M. (2011b) Using large data sets to forecast housing prices: A case study of twenty US states. 20(2), Journal of Housing Research.

Hall, S., Psaradakis, Z. and Sola, M. (1997) Switching error-correction models of house prices in the United Kingdom. **Economic Modelling**, 14 (4): 517–527

Hamilton, B.W. and Schwab, R.M. (1985) Expected appreciation in urban housing markets. **Journal of Urban Economics**, 18 (1): 103–118. URL http://ideas.repec.org/a/eee/juecon/v18y1985i1p103-118.html.

Harrison, D. and Rubinfeld, D.L. (1978) Hedonic housing prices and the demand for clean air. **Journal of Environmental Economics and Management**, 5 (1): 81 – 102. URL http://www.sciencedirect.com/science/article/pii/0095069678900062.

Hendershott, P.H. (1991) Are real house prices likely to decline by 47 percent? **Regional Science and Urban Economics**, 21 (4): 553 – 563. URL http://www.sciencedirect.com/science/article/pii/016604629190019J.

Hendry, D.F. and Clements, M.P. (2004) Pooling of forecasts. **Econometrics Journal**, 7 (1): 1–31. URL http://ideas.repec.org/a/ect/emjrnl/v7y2004i1p1-31.html.

Hendry, D.F. and Doornik, J.A. (1999) The impact of computational tools on time-series econometrics. **In T. Coppock (Ed.), Information Technology and Scholarship**, p. 257269

Hordijk, L. (1979) Problems in estimating econometric relations in space. **Papers of the Regional Science Association**, 42 (1): 99–115. URL http://dx.doi.org/10.1007/BF01935148.

Huang, B., Wu, B. and Barry, M. (2010) Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. **International Journal of Geographical Information Science**, 24 (3): 383–401. URL http://www.tandfonline.com/doi/abs/10.1080/13658810802672469.

Hyndman, R.J. and Khandakar, Y. (2008) Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, 27 (3): 1–22. URL http://www.jstatsoft.org/v27/i03.

Iacoviello, M. (2002) House prices and business cycles in Europe: a VAR analysis. Boston College Working Papers in Economics 540, Boston College Department of Economics.

Iacoviello, M. (2005) House prices, borrowing constraints, and monetary policy in the business cycle. **The American Economic Review**, 95 (3): 739–764

Iacoviello, M. and Minetti, R. (2008) The credit channel of monetary policy: Evidence from the housing market. **Journal of Macroeconomics**, 30 (1): 69–96

Iacoviello, M. and Neri, S. (2010) Housing market spillovers: Evidence from an estimated DSGE model. **American Economic Journal: Macroeconomics**, 2 (2): 125–64. URL http://ideas.repec.org/a/aea/aejmac/v2y2010i2p125-64.html.

Inoue, A. and Kilian, L. (2013) Inference on impulse response functions in structural VAR models. **Journal of Econometrics**, 177 (1): 1 – 13. URL http://www.sciencedirect.com/science/article/pii/S0304407613001310.

Ismail, S. (2005) **Hedonic Modelling of Housing Markets Using Geographical Information System (GIS) and Spatial Statistics: A Case Study of Glasgow, Scotland**. Aberdeen University. URL http://books.google.co.uk/books?id=SavkcQAACAAJ.

Jackson, J.R. (1979) Intraurban variation in the price of housing. **Journal of Urban Economics**, 6 (4): 464 – 479. URL http://www.sciencedirect.com/science/article/pii/0094119079900251.

Jarocinski, M. and Smets, F. (2008) House prices and the stance of monetary policy. Working Paper Series 891, European Central Bank.

Joyce, M.A.S., Lasaosa, A., Stevens, I. *et al.* (2011) The financial market impact of quantitative easing in the United Kingdom. **International Journal of Central Banking**, 7 (3): 113–161

Kapetanios, G., Labhard, V. and Price, S. (2008a) Forecast combination and the Bank of England's suite of statistical forecasting models. **Economic Modelling**, 25 (4): 772–792. URL http://ideas.repec.org/a/eee/ecmode/v25y2008i4p772-792.html.

Kapetanios, G., Labhard, V. and Price, S. (2008b) Forecasting using Bayesian and information-theoretic model averaging: An application to UK inflation. **Journal of Business & Economic Statistics**, 26: 33–41. URL http://ideas.repec.org/a/bes/jnlbes/v26y2008p33-41.html.

Kim, S. and Roubini, N. (2000) Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach. **Journal of Monetary Economics**, 45 (3): 561–586. URL http://ideas.repec.org/a/eee/moneco/v45y2000i3p561-586.html.

Koop, G.M. (2013) Forecasting with medium and large Bayesian VARS. Tech. Rep. 2. URL http://dx.doi.org/10.1002/jae.1270.

Lancaster, K.J. (1966) A new approach to consumer theory. **Journal of Political Economy**, 74. URL http://EconPapers.repec.org/RePEc:ucp:jpolec:v:74:y:1966:p:132.

Leamer, E.E. (2007) Housing IS the business cycle. Working Paper 13428, National Bureau of Economic Research. URL http://www.nber.org/papers/w13428.

LeSage, J. and Pace, R. (2004) Models for spatially dependent missing data. **The Journal of Real Estate Finance and Economics**, 29 (2): 233–254. URL http://dx.doi.org/10.1023/B%3AREAL.0000035312.82241.e4.

LeSage, J. and Pace, R. (2009) **Introduction to Spatial Econometrics**. Statistics: A Series of Textbooks and Monographs, Taylor & Francis. URL http://books.google.co.uk/books?id=EKiKXcgL-D4C.

LeSage, J.P. and Pace, R.K. (2014) The biggest myth in spatial econometrics. **Econometrics**, 2 (4): 217. URL http://www.mdpi.com/2225-1146/2/4/217.

Litterman, R.B. (1983) A random walk, Markov model for the distribution of time series. Tech. rep.

Lütkepohl, H. (1985) Comparison of criteria for estimating the order of a vector autoregressive process. **Journal of Time Series Analysis**, 6 (1): 35–52. URL http://dx.doi.org/10.1111/j.1467-9892.1985.tb00396.x.

Lütkepohl, H. (2006) **New introduction to multiple time series analysis**. New York. URL http://books.google.co.uk/books?id=muorJ6FHIiEC.

Mack, A. and Martínez-García, E. (2013) A cross-country quarterly database of real house prices: a methodological note. Tech. rep.

Malpezzi, S. (2008) **Hedonic Pricing Models: A Selective and Applied Review**. Blackwell Science Ltd. pp. 67–89. URL http://dx.doi.org/10.1002/9780470690680.ch5.

Mankiw, N.G. and Weil, D.N. (1989) The baby boom, the baby bust, and the housing market. Tech. Rep. 2. URL http://ideas.repec.org/a/eee/regeco/v19y1989i2p235-258.html.

Marcellino, M., Favero, C.A. and Neglia, F. (2005) Principal components at work: the empirical analysis of monetary policy with large data sets. **Journal of Applied Econometrics**, 20 (5): 603–620. URL http://ideas.repec.org/a/jae/japmet/v20y2005i5p603-620.html.

Marcellino, M., Stock, J.H. and Watson, M.W. (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. **Journal of Econometrics**, 135 (1-2): 499–526. URL http://ideas.repec.org/a/eee/econom/v135y2006i1-2p499-526.html.

Medel, C.A. and Salgado, S.C. (2012) Does BIC Estimate and Forecast Better Than AIC? Working Papers Central Bank of Chile 679, Central Bank of Chile.

Meen, G. (2001) **Modelling Spatial Housing Markets: Theory, Analysis, and Policy**. Advances in Urban and Regional Economics, Springer US. URL http://books.google.com.au/books?id=QT-ySsgeF40C.

Miles, D. (1994) **Housing, financial markets and the wider economy**. Series in financial economics and quantitative analysis, J. Wiley. URL http://books.google.co.uk/books?id=wXORAAAAIAAJ.

Min, C. and Zellner, A. (1992) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Working papers, California Irvine - School of Social Sciences. URL http://EconPapers.repec.org/RePEc:fth:calirv:90-92-23.

Nappi-Choulet, I. and Maury, T.P. (2011) A spatial and temporal autoregressive local estimation for the Paris housing market. **Journal of Regional Science**, 51 (4): 732–750. URL http://ideas.repec.org/a/bla/jregsc/v51y2011i4p732-750.html.

Pace, R., Barry, R., Clapp, J. *et al.* (1998) Spatiotemporal autoregressive models of neighborhood effects. **The Journal of Real Estate Finance and Economics**, 17 (1): 15–33. URL http://dx.doi.org/10.1023/A%3A1007799028599.

Pace, R. and LeSage, J.P. (2008) A spatial Hausman test. **Economics Letters**, 101 (3): 282–284. URL http://ideas.repec.org/a/eee/ecolet/v101y2008i3p282-284.html.

Pace, R.K., Barry, R., Gilley, O.W. *et al.* (2000) A method for spatial-temporal forecasting with an application to real estate prices. **International Journal of Forecasting, Elsevier**, 16 (2): 229–246. URL http://ideas.repec.org/a/eee/intfor/v16y2000i2p229-246.html.

Pace, R.K. and Gilley, O.W. (1997) Using the spatial configuration of the data to improve estimation. **The Journal of Real Estate Finance and Economics, Springer**, 14 (3): 333–40. URL http://ideas.repec.org/a/kap/jrefec/v14y1997i3p333-40.html.

Paelinck, J. and Klaassen, L. (1979) **Spatial Econometrics**. Studies in Spatial Analysis, Saxon House. URL http://books.google.co.uk/books?id=mVTaAAAAMAAJ.

Pagan, A. (1984) Econometric issues in the analysis of regressions with generated regressors. **International Economic Review**, 25 (1): 221–47. URL http://ideas.repec.org/a/ier/iecrev/v25y1984i1p221-47.html.

Park, B. and Bae, J.K. (2015) Using machine learning algorithms for housing price prediction. **Expert Syst. Appl.**, 42 (6): 2928–2934. URL http://dx.doi.org/10.1016/j.eswa.2014.11.040.

Patrick, J., Okunev, J., Ellis, C. *et al.* (2000) Comparing univariate forecasting techniques in property markets. **Journal of Real Estate Portfolio Management**, 3: 283–306

Pavlov, A.D. (2000) Space-varying regression coefficients: A semi-parametric approach applied to real estate markets. **Real Estate Economics**, 28 (2): 249–283. URL http://dx.doi.org/10.1111/1540-6229.00801.

Peersman, G. (2011) Macroeconomic effects of unconventional monetary policy in the euro area. Tech. rep.

Pesaran, H. and Timmermann, A. (2005) Real-Time Econometrics. **Econometric Theory**, 21 (01): 212–231. URL http://ideas.repec.org/a/cup/etheor/v21y2005i01p212-231_05.html.

Pesaran, M. and Smith, R. (1995) Estimating long-run relationships from dynamic heterogeneous panels. **Journal of Econometrics**, 68 (1): 79–113

Phillips, P. (1986) ET interview with Professor T.W. Anderson. **Econometric Theory**, (2): 249288

Qu, X. and Lee, L.f. (2015) Estimating a spatial autoregressive model with an endogenous spatial weight matrix. **Journal of Econometrics**, 184 (2): 209–232. URL http://EconPapers.repec.org/RePEc:eee:econom:v:184:y:2015:i:2:p:209-232.

Raftery, A.E., Karny, M. and Ettler, P. (2010) Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. **Technometrics**, 52 (1): 52–66. PMID: 20607102.

Rapach, D.E. and Strauss, J.K. (2007) Forecasting real housing price growth in the eighth district states. **Regional Economic Development**, (Nov): 33–42. URL http://ideas.repec.org/a/fip/fedlrd/y2007inovp33-42nv.3no.2.html.

Rapach, D.E. and Strauss, J.K. (2009) Differences in housing price forecastability across US states. **International Journal of Forecasting**, 25 (2): 351–372. URL http://ideas.repec.org/a/eee/intfor/v25y2009i2p351-372.html.

Ren, Y. and Yuan, Y. (2014) Why the housing sector leads the whole economy: The importance of collateral constraints and news shocks. **The Journal of Real Estate Finance and Economics**, 48 (2): 323–341. URL http://dx.doi.org/10.1007/s11146-012-9389-5.

Rosen, S. (1974) Hedonic prices and implicit markets: Product differentiation in pure competition. **Journal of Political Economy**, 82 (1): 34–55. URL http://ideas.repec.org/a/ucp/jpolec/v82y1974i1p34-55.html.

Rubio, M. (2014) Housing-market heterogeneity in a monetary union. **Journal of International Money and Finance**, 40: 163 – 184. URL http://www.sciencedirect.com/science/article/pii/S0261560613000867.

Rudebusch, G. and Svensson, L.E. (1999) "Policy Rules for Inflation Targeting." In **Monetary Policy Rules**. NBER Chapters, National Bureau of Economic Research, Inc. pp. 203–262. URL http://ideas.repec.org/h/nbr/nberch/7417.html.

Sá, F., Towbin, P. and Wieladek, T. (2011) Low interest rates and housing booms: the role of capital inflows, monetary policy and financial innovation. Tech. rep.

Samuelson, P. and Nordhaus, W. (2009) **Economics**. The McGraw-Hill series economics, McGraw-Hill Higher Education. URL https://books.google.co.uk/books?id=eS5ZAAAAYAAJ.

Schumacher, C. and Dreger, C. (2004) Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models? **Journal of Economics and Statistics**, 224 (6): 731–750. URL http://ideas.repec.org/a/jns/jbstat/v224y2004i6p731-750.html.

Shiller, R. (2005) **Irrational Exuberance**. Broadway Books, 1st Edition, Currency/Doubleday. URL http://books.google.co.uk/books?id=x1PaZY_KtBEC.

Shiller, R. (2008) **The subprime solution: How today's global financial crisis happened, and what to do about it**. Princeton University Press, 1st Edition. URL http://books.google.co.uk/books?id=aj0kmAEACAAJ.

Sims, C.A. (1980) Macroeconomics and reality. **Econometrica**, 48 (1): 1–48. URL http://ideas.repec.org/a/ecm/emetrp/v48y1980i1p1-48.html.

Sims, C.A., Stock, J.H. and Watson, M.W. (1990) Inference in linear time series models with some unit roots. **Econometrica**, 58 (1): 113–44. URL http://ideas.repec.org/a/ecm/emetrp/v58y1990i1p113-44.html.

Smets, F. and Peersman, G. (2001) The monetary transmission mechanism in the euro area: more evidence from VAR analysis. Working Paper Series 091, European Central Bank.

Smith, T. and Wu, P. (2009) A spatio-temporal model of housing prices based on individual sales transactions over time. **Journal of Geographical Systems**, 11 (4): 333–355. URL http://dx.doi.org/10.1007/s10109-009-0085-9.

Stock, J. and Watson, M. (2002a) Forecasting using principal components from a large number of predictors. **Journal of the American Statistical Association**, 97: 1167–1179. URL http://ideas.repec.org/a/bes/jnlasa/v97y2002mdecemberp1167-1179.html.

Stock, J.H. and Watson, M. (2003) Forecasting output and inflation: The role of asset prices. **Journal of Economic Literature**, 41 (3): 788–829. URL http://ideas.repec.org/a/aea/jeclit/v41y2003i3p788-829.html.

Stock, J.H. and Watson, M.W. (1998) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. NBER Working Papers 6607, National Bureau of Economic Research, Inc. URL http://ideas.repec.org/p/nbr/nberwo/6607.html.

Stock, J.H. and Watson, M.W. (2002b) Macroeconomic forecasting using diffusion indexes. **Journal of Business & Economic Statistics**, 20 (2): 147–62. URL http://ideas.repec.org/a/bes/jnlbes/v20y2002i2p147-62.html.

Stock, J.H. and Watson, M.W. (2005) Implications of dynamic factor models for VAR analysis. Working Paper 11467, National Bureau of Economic Research. URL http://www.nber.org/papers/w11467.

Sun, H., Tu, Y. and Yu, S.M. (2005) A spatio-temporal autoregressive model for multi-unit residential market analysis. **The Journal of Real Estate Finance and Economics**, 31 (2): 155–187. URL http://dx.doi.org/10.1007/s11146-005-1370-0.

Timmermann, A. (2006) **Forecast Combinations**, Handbook of Economic Forecasting, vol. 1. Elsevier. chap. 4, pp. 135–196. URL http://ideas.repec.org/h/eee/ecofch/1-04.html.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. **Economic Geography**, 46 (2): 234–240

Uhlig, H. (2005) What are the effects of monetary policy on output? Results from an agnostic identification procedure. **Journal of Monetary Economics**, 52 (2): 381–419

Vargas-Silva, C. (2008a) The effect of monetary policy on housing: a factor-augmented vector autoregression (FAVAR) approach. **Applied Economics Letters**, 15 (10): 749–752

Vargas-Silva, C. (2008b) Monetary policy and the US housing market: A VAR analysis imposing sign restrictions. **Journal of Macroeconomics**, 30 (3): 977–990

Wang, L. and Ready, R. (2005) Spatial econometric approaches to estimating hedonic property value models. 2005 Annual meeting, July 24-27, Providence, RI 19174, American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association). URL http://EconPapers.repec.org/RePEc:ags:aaea05:19174.

Wilhelmsson, M. (2002) Spatial models in real estate economics. **Housing, Theory and Society**, 19 (2): 92–101. URL http://dx.doi.org/10.1080/140360902760385646.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. **Journal of the American Statistical Association**, 57 (298): 348–368

Zhu, B., Füss, R. and Rottke, N. (2011) The predictive power of anisotropic spatial correlation modeling in housing prices. **The Journal of Real Estate Finance and Economics**, 42 (4): 542–565. URL http://dx.doi.org/10.1007/s11146-009-9209-8.