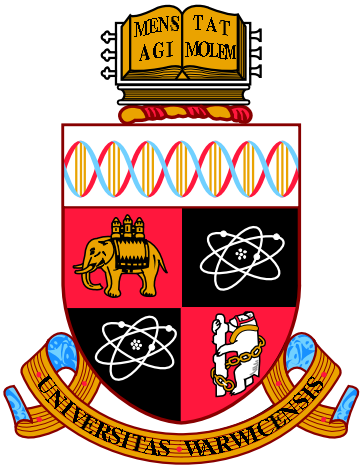

Forecasting real crude oil prices, their uncertainties and a Bayesian structural method for the world crude oil market

by

Yunyi Zhang

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Business School

Sepetember 2019

THE UNIVERSITY OF WARWICK

Contents

List of Tables v

List of Figures vii

Acknowledgments x

Declarations xi

Abstract xii

Chapter 1 Introduction 1

Chapter 2 Point Forecasts of Real Crude Oil Prices: Combinations for Baumeister & Kilian's (2015) Model Space 4

2.1 Introduction 5

2.2 A real-time data set for oil price forecasting 7

2.2.1 Revisions 10

2.2.2 Nowcasts 13

2.3 The model space and combinations 14

2.4 Empirical results 17

2.4.1 Narrow replication of the sample period 1992:01—2012:09 . 18

2.4.2 Extension: the sample period 1992:01–2016:12 for RAC and WTI measures 20

2.4.3 Extension: the Brent measure 21

2.4.4 Extension: recursive forecast accuracy of equal weighted combinations 22

2.4.5 Robustness and sensitivity . . . 23
2.4.6 Quarterly forecasts of the sample period 1992Q1—2012Q3 . 36
2.5 Conclusions . . . 41

Chapter 3 Density Forecasts of Real Crude Oil Prices Using Macro Founded Dynamic Models and Extended Evaluations 42

3.1 Introduction . . . 43
3.2 Models . . . 47
3.2.1 Bayesian estimation of SMSS for VAR with time-varying parameters and stochastic volatility . . . 47
3.2.2 Model specifications . . . 50
3.2.3 Out-of-sample forecast implementation . . . 54
3.3 The data . . . 59
3.4 Forecast evaluation . . . 61
3.4.1 In-sample diagnostic statistic . . . 61
3.4.2 Point forecast evaluation . . . 65
3.4.3 Density forecast evaluation . . . 69
3.4.4 Forecast performance measured using financial market pay-off 75
3.4.5 Probability forecasts for 'extreme high' and 'extreme low' real oil prices . . . 77
3.5 A brief discussion for the results so far . . . 88
3.6 Concluding remarks . . . 89

Chapter 4 Constrained Bayesian SVARs: Two Specifications in the World Crude Oil Market Modelling 91

4.1 Introduction . . . 92
4.2 Literature review . . . 97
4.3 The constrained Bayesian structural vector autoregression . . . 100
4.3.1 Step I: estimation of the reduced-form model and measuring its empirical fit . . . 103
4.3.2 Step II: numerically determine the starting point for the MH algorithm . . . 104
4.3.3 Step III: the MH algorithm . . . 108

4.4 Two world crude oil market specifications . . . 110
4.4.1 Specification I: identification restrictions in Kilian & Murphy (2014) . . . 111
4.4.2 Discussion: what can we learn about oil demand elasticities from C-BSVARs relative to the oil literature? . . . 116
4.4.3 Specification II: the uncertainty of the short-run lower bound for oil demand elasticity for use . . . 121
4.5 Conclusion . . . 129

Chapter 5 Conclusion **130**

Appendix A Process for Constructing CPI, World Economic Activity Index and Backcasts **133**
A.1 The U.S. consumer price index for all urban consumers (CPI) . . . 133
A.2 An index of the bulk dry cargo ocean shipping freight rates . . . 134
A.3 Backcasts . . . 136
A.3.1 The backcasts for RAC and OECD petroleum stocks . . . 137
A.3.2 The backcast for oil futures with maturities 1–24 months . . . 138

Appendix B Technical Details and Additional Results in Chapter 3 **141**
B.1 Prior and the posterior computation for ω_j^* in SMSS . . . 141
B.2 Technical details for specifications . . . 143
B.3 The extension of Baumeister & Kilian's (2015) combination into density forecasts . . . 144
B.3.1 An unrestricted globe oil market vector autoregression (VAR) 144
B.3.2 A commodity price based model . . . 144
B.3.3 An oil futures spread based model . . . 145
B.3.4 A gasoline spread based model . . . 145
B.3.5 A time varying parameter product spread model . . . 146
B.3.6 No-change or random walk model . . . 146
B.4 Convergence diagnostics . . . 147
B.5 A justification for MKLD accompanied with the extreme-forecasts-elimination process . . . 149

B.6 A realistic signalling strategy for daily crude oil futures, conditional on real crude oil price density forecasts 153
B.7 The calculation of excess returns 156
B.8 The cumulative excess returns for WTI and Brent measures in futures market 157
B.9 Probabilities for negative excess returns 160

Appendix C Technical Details and Additional Results in Chapter 4 164

C.1 Reference priors for Σ_u and B 164
C.2 Sampling from the posterior distribution 166
C.3 Instruction for A^{-1} based on the identification restrictions in Kilian & Murphy (2014) 167
C.4 The economic findings in Specification I 170
C.5 Additional results for Specification II 174
C.6 $\eta^{Supply} \leq 0.0258$ VS. $\eta^{Supply} \leq 0.5$ 178

List of Tables

2.1 Data (without subsequent revisions) resources for observations from 1991:12 to 2017:06 8

2.2 Real-time data (with subsequent revisions) resources for vintages dated 1991:12 to 2017:06 9

2.3 Descriptive statistics on data revisions 12

2.4 Real-time forecast accuracy of baseline forecast combination based on all six forecasting models for the real U.S. refiners' acquisition cost for oil imports (the evaluation period: 1992:01–2012:09) 19

2.5 Real-time forecast accuracy of baseline forecast combination based on all six forecasting models for the real WTI price (the evaluation period: 1992:01–2012:09) 19

2.6 Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real U.S. refiners' acquisition cost for oil imports 20

2.7 Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real WTI price 21

2.8 Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real Brent price 22

2.9 Real-time forecast accuracy of equal weighted forecast combination based on all six forecasting models (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12) under different revision assumptions . 25

2.10 Real-time recursive MSPE and success ratios of forecast combinations with equal weights under different nowcast assumptions (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12) 28

2.11 Historical changes in real-time recursive MSPE and success ratios of leave-one-out forecast combinations with equal weights (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12) 33
2.12 Real-time forecast accuracy of equal-weighted forecast combinations at quarterly horizons 39

3.1 A list of specifications used for this application, including their label, estimator, prior and uncertainty 53
3.2 Real-time forecast accuracy of model specifications at selected forecast monthly horizons 66
3.3 Real-time density forecast accuracy of model specifications at selected forecast monthly horizons for log(RAC) 72
3.4 Hit rate of financial market excess returns for the WTI and Brent measures, 1992:01–2016:12 77
3.5 Payout contingencies for the outcome of a symmetric fair bet 82
3.6 Payout contingencies for the outcome of an asymmetric fair bet . . . 83
3.7 Forecasting 'extreme high' and 'extreme low' real crude oil prices, 1992:01–2016:12 84
3.8 Brier Score and decomposition for 'extreme high' and 'extreme low' real crude oil prices forecasts, 1992:01–2016:12 87

4.1 Sign restrictions on impact responses in the VAR model 112
4.2 Posterior distributions of the short-run price elasticities of demand for crude oil 125

B.1 Convergence diagnostics for all specifications 148
B.2 Out-of-sample AR(1) coefficients for time-varying parameters 150
B.3 Probability of negative financial market excess returns for the WTI measure, 1992:01–2016:12 162
B.4 Probability of negative financial market excess returns for the Brent measure, 1992:01–2016:12 163

C.1 Hyperparameter for Shrinkage 166
C.2 Prior distributions for model parameters 169

List of Figures

2.1 Monthly real crude oil prices (nominal prices divided by the U.S. consumer price index (CPI)) 5

2.2 Historical recursive forecast accuracy of Baumeister & Kilian's (2015) equal weighted averages at selected horizons (recursive MSPE and success ratios) 23

2.3 Historical changes in real-time recursive MSPE ratios of leave-one-out forecast combinations with equal weights 34

2.4 Historical changes in real-time success ratios of leave-one-out forecast combinations with equal weights 35

2.5 Quarterly historical recursive forecast accuracy relative to no-change forecast for refiner's acquisition cost for oil imports: EIA forecast, Equal weighted combination of all six forecasting models excepting a gasoline spread based model and no-change forecast, and Equal weighted combination of all six forecasting models 40

3.1 Monthly real crude oil prices (nominal prices divided by the U.S. CPI) 59

3.2 Estimated weights for different VAR lag-length choices at forecast horizons 1, 3, 6, 12, 18 and 24 months 62

3.3 Estimated time-invariance probabilities for TVPsmss, TVPSVsmss, and CombineTVPSVsmss 63

3.4 The percentage of permissible draws from posterior for BAVR, SV, TVP, TVPsmss, TVPSV and CombineTVPSVsmss over the evaluation period 1992:01–2016:12 64

3.5 Historical recursive MSPE ratios at selected forecast horizons for the evaluation period 2008:03-2016:12 67

3.6 Historical Success ratios at selected forecast horizons for the evaluation period 2008:03-2016:12 68
3.7 Histogram of *pits* for all specifications at selected horizons 73
3.8 Density forecasts (1% to 99% quantiles) for log(RAC) over the period 1992:01–2016:12 based on VAR, SV and BK at selected horizons . . 74
3.9 Probability forecasts for extreme high and low crude oil prices at the 1-, 3- and 6-month horizons 79
3.10 Probability forecasts for extreme high and low crude oil prices at the 12-, 18- and 24-month horizons 80

4.1 Kernel fitted probability distribution functions 95
4.2 Structural impulse responses: 1973:02–2009:08 ($\eta^{Supply} \leq 0.0258$) . . 115
4.3 Posterior densities of the oil demand elasticity in production and use under a -0.8 lower bound 117
4.4 Convergence diagnostics for A^{-1}, $\eta^{O,Use}$ and $\eta^{O,Production}$ in Specification I 118
4.5 Convergence diagnostics for A^{-1}, $\eta^{O,Use}$ and $\eta^{O,Production}$ in Specification II 123
4.6 Posterior densities of the oil demand elasticity in production and use under an uncertain lower bound with different oil supply elasticity higher bounds, including 0.0258, 0.05, 0.1 and 0.5 124
4.7 The historical decomposition for the real crude oil price for the 1978:06–2009:08 period 127

A.1 Monthly index of the global real economic activity based on dry cargo bulk freight rates and the Baltic Dry Cargo Index (1973:01-2017:06) 136

B.1 Recursive out-of-sample AR(1) coefficients for time-varying parameters 152
B.2 Four scenarios for the calculation of $\zeta_{T,h}$ 154
B.3 Financial market cumulative excess returns for the WTI measure in the period 1992:01–2016:12 158
B.4 Financial market cumulative excess returns for the Brent measure in the period 1992:01–2016:12 159

C.1 Historical decomposition for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$ 172

C.2 Historical decompositions for specific events (in columns) of the real price of oil (the upper panels) and the change in oil inventories (the lower panels) . . . 173
C.3 Prior (*red lines*) and posterior (*black histograms*) distributions for the unknown elements in A^{-1} in the Bayesian implementation of the 4-variable baseline model ($\eta^{Supply} \leqslant 0.0258$) . . . 174
C.4 Structural impulse responses under the uncertainty of the oil demand elasticity in use for the 1973:02–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$ . . . 175
C.5 Historical decomposition under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$ . . . 176
C.6 Historical decompositions for specific events (in columns) of the real price of oil (the upper panels) and the change in oil inventories (the lower panels) under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$ . . . 177
C.7 Historical decomposition under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, comparing the models with $\eta^{Supply} \leq 0.0258$ and $\eta^{Supply} \leq 0.5$ . . . 178
C.8 Structural impulse responses under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, comparing the models with $\eta^{Supply} \leq 0.0258$ and $\eta^{Supply} \leq 0.5$ . . . 179

Acknowledgments

First and foremost, I want to express my infinite gratefulness to my supervisors: Professor Anthony Garratt and Professor Shuan Vahey, without whom I could not achieve what I have today. Their constant encouragement boost the confidence on myself and my career; their constant challenges shape my critical and incisive view on the academic work; their constant care establish a great sample for being a decent and thoughtful person, and an excellent supervisor. I also want to thank other scholars from Warwick Business School and Economics Department. Their sharp and valuable comments on some parts of my thesis make a huge difference on the completed version. The opportunities they have offered on both researching and teaching are so precious that I could not imagine if I was elsewhere.

Besides, I want to show great respect to my families, especially my parents, who are thousands miles away from here. Thanks to their visionary choice for sending me to the UK for my higher education, so that I could start an exciting and fruitful journey. Without their understanding and support both mentally and financially, I could not reach the level where I am now.

Also, a great credit should be given to my fiancée, who has been my company for the three-year's bittersweetness. She is the only one besides me who cheers me up when I was down, who helps me know myself better, who brings me endless joy, laughter and support.

Finally, I applause for my friends and team members of the Warwick Chinese Basketball team. It is they that stayed by my side all the time. They are the colour pallet to my life. They have taught me how to be a good friend, and what is a team.

Declarations

This thesis is the result of the author's original research, produced under the supervision of Professor Anthony Garratt and Professor Shaun P. Vahey at Warwick Business School, the University of Warwick, UK, between October 2015 and September 2019. All parts of the research have been composed by the author and have not been previously submitted for examination with the purpose of obtaining a higher degree.

A short version of Chapter 2 has been published as: Garratt, Anthony, Vahey, Shaun P., & Zhang, Yunyi (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, 34(3), 456–462. Chapter 2 contains a significant amount of unpublished additional materials, including a real time data set for real oil price forecasting, the forecasts under alternative revision and nowcast assumptions, as well as, the forecasts with quarterly data. These additional results not only slightly improve the forecast accuracy, but also confirm the robustness on quarterly frequency. The results are the author's original contribution.

Abstract

This thesis comprises three essays focusing on real crude oil price forecasting and structural analysis. The first essay (Chapter 2) begins by broadly reproducing Baumeister & Kilian's (2015) main economic findings, where an equal-weight combination of six econometric models outperforms a recursive mean squared predictor error weight based combination in oil-price point forecasting. The six models are an unrestricted global oil market vector autoregression, a commodity-price model, an oil-futures-spread model, a gasoline-spread model, a time-varying parameter product-spread model, and a random-walk model. I use their preferred measures of the real oil price and similar real-time variables. Remaining mindful of the importance of Brent crude oil as a global price benchmark and the divergence in oil price measures since 2010, I extend consideration to the North Sea-based measure and update the evaluation sample to 2016:12, finding that the combined forecasts for Brent crude oil are as accurate as the forecasts for other oil price measures. The extended sample employing the oil price measures adopted by Baumeister & Kilian (2015) yields similar results to those reported in their paper.

The second essay (Chapter 3) uses a Bayesian vector autoregression (BVAR) utilising time-varying parameters and stochastic volatility modelling time variation in forecasting real crude oil prices. An unrestricted global oil market vector autoregression and the equal-weight combination in Baumeister & Kilian (2015) are benchmarks. I extend the evaluation for model comparison purposes from standard statistical terms of point and density forecasts to an economic evaluation based

on which specification would be more profitable in the crude oil futures market, and the forecast likelihood of extreme high and low real crude oil prices. For the same evaluation period as in Chapter 2, 1992:01–2016:12, the empirical results offer strong support for models using stochastic volatility in real crude oil price density forecasts relative to conventional VAR. Restricting time-varying parameters and allowing stochastic volatility can increase the probability of positive excess returns through utilising daily crude oil futures data, and can improve the calibration of the extreme high and low real oil price events forecasting. In conclusion, adding stochastic volatility and using the stochastic model specification search shrinkage prior of Eisenstat et al. (2016) are both important in ensuring reliable forecasts.

Finally, in the third essay (Chapter 4) I develop a parallel Metropolis–Hastings (MH) algorithm to identify and compute Bayesian structural vector autoregressions (SVARs), which I refer to C-BSVARs. The motivation for this is the inefficiency of the traditional computation method for SVARs under certain types of identification. C-BSVARs extend Baumeister & Hamilton's (2015) method from only using sign restrictions to a broader set of identification assumptions and improve the computational efficiency relative to the traditional method for SVARs. Two specifications from the world crude oil market modelling are used to illustrate this. The first employs Kilian & Murphy's (2014) set of identification restrictions, while the second imposes an additional restriction on top of theirs — the uncertainty of a lower-bound on 'the short-run oil demand elasticity for use'. C-BSVARs dramatically improve the acceptance rate for models deemed as admissible relative to the method used in Kilian & Murphy (2014), and it can narrow the critical intervals of Kilian & Murphy's (2014) impulse response functions. The additional restriction in the second specification enables precise estimates of oil demand elasticities, which is of importance for deciding the existence of crude oil price endogeneity and the relative weights of oil demand and supply shocks for driving the fluctuation of crude oil prices. To my knowledge, only the C-BSVARs approach in the existing SVARs literature is able to impose the restriction of uncertainties for elasticities, thereby providing a novel way of identifying key structural parameters that are non-linear.

Chapter 1

Introduction

Oil price shocks have long been viewed as one of the leading candidates for explaining economic fluctuations in the U.S. and global economies, and therefore accurate forecasting of real crude oil prices is of considerable importance. In this thesis I (1) extend Baumeister & Kilian's (2015) analysis of point forecasts of real oil prices to post-oil-price slump data while incorporating the Brent measure, and then (2) turn to density forecasts of crude oil prices with a new model space considering the time variation of parameters. Further, I develop a new Bayesian algorithm which allows for world oil market transmission analysis. This is motivated by the fact that the traditional accept–reject method for structural vector autoregressions (SVARs) frequently employed in the oil literature is a computationally inefficient method when imposing restrictions, for example the dynamic sign restrictions approach employed in Kilian & Murphy (2014).

Baumeister & Kilian (2015) combine the forecasts from six empirical models to predict real oil prices, and for their 1992:01–2012:09 evaluation period they demonstrate that forecast combinations are substantially more accurate than no-change forecasts at horizons up to 18 months. Following the introduction, Chapter 2 begins by broadly reproducing their main economic findings, using their preferred measures of the real oil price and similar real-time variables. Given the importance of Brent crude oil as a global price benchmark and the divergence in oil price measures since 2010, I extend consideration to the North Sea-based measure and update the evaluation sample to 2016:12. I find that the combined forecasts for Brent crude oil are as accurate as the forecasts for other oil price measures, while the extended sample utilising the oil price measures adopted by Baumeister & Kilian (2015) yields similar results to those reported in their paper.

Baumeister & Kilian's (2015) combination method is effective for point forecasts of real crude oil prices, as confirmed in Chapter 2. However, density forecasting is important, because in general a loss function needs more information than typically given in the point forecast. Chapter 3 introduces a new model space: a Bayesian vector autoregression (BVAR) considering time-varying parameters and stochastic volatility, applied to unrestricted 4-variable world oil market VAR proposed in Kilian & Murphy (2014). This is aimed at incorporating the time variation of parameters thought to be a feature of the oil markets whilst at the same time retaining the interactions allowed for through the use of the VAR. Specifically in Chapter 3, I (1) minimise the one-step-ahead in-sample forecasts Kullback–Leibler's 'distance' whilst incorporating the elimination of out-of-sample extreme forecasts, which in turn enables the shrinkage of highly parameterised models' out-of-sample density forecasts, particularly at long-forecast horizons; and (2) extend the evaluation for model comparison purposes from standard statistical metrics for point and density forecasts to profit and event outcome based metrics which suggest specifications that would be more profitable in the crude oil futures market, and the accuracy of forecasting the likelihood of extreme high and low real crude oil prices. Over the 1992:01–2016:12 period, the empirical results offer strong evidence in favour of models using stochastic volatility in real crude oil price density forecasts as opposed to using a conventional VAR. Two specifications standout in terms of the density and probability event forecast performance: a constant parameter VAR with stochastic volatility, as well as, a specification with stochastic volatility and time-varying parameters. For the second model a majority of the parameters were restricted to be time-invariant via a stochastic model specification selection prior and a linear opinion pool combination model for the 1- to 12-month VAR lag length choices.

In contrast to Chapters 2 and 3, Chapter 4 contributes to the SVARs literature with two specifications of the world crude oil market. The 4-variable world oil market SVARs developed in Kilian & Murphy (2014) is widely utilised in the real crude oil literature. The traditional accept-reject method for SVARs, employed in Kilian & Murphy (2014), is a computationally inefficient method of imposing identification. One consequence is that typically only a low number of admissible draws are available, leading to imprecise inferences. For example, the sifting of 16 models in Kilian & Murphy (2014) required 5,000,000 posterior draws, only 150 of which were deemed admissible. Recently, an alternative method of identifying sign restrictions was proposed in Baumeister & Hamilton (2015), which does not run into the computational problem, but it is limited to sign restrictions only. In practice, however, a set of identification restrictions, including further qualitative restrictions

taking the analysis beyond the use of sign of contemporary correlation matrix, is required for precise estimates.

Therefore, Chapter 4 develops a constrained estimation of Bayesian SVARs, which I refer to as C-BSVARs. C-BSVARs use a parallel Metropolis–Hastings (MH) algorithm to identify and compute Bayesian SVARs, which extends Baumeister & Hamilton's (2015) method from sign to a set of identification restrictions. In Chapter 4 I illustrate that C-BSVARs enable a flexible identification scheme, and dramatically improve the computational efficiency relative to traditional accept-reject SVARs using two specifications of the 4-variable world oil market SVAR proposed in Kilian & Murphy (2014). The first specification employs Kilian & Murphy's (2014) exact identification, and finds results consistent with their economic findings. However, the density estimates of oil demand elasticities are too diffuse, and are sensitive to the choice of seed for random numbers. From the evidence of the convergence diagnostics, the sensitivity is highly probably caused by the non-unimodal posterior distributions of the elasticities, which may lead a standard MH sampler to sample draws within a local maximum (Waggoner et al., 2016). However, the magnitude of oil demand elasticities is of considerable importance, since this determines the existence of oil price endogeneity and the relative importance of oil demand and supply shocks in terms of influencing oil market fluctuations. The second specification, therefore, proposes an additional identification restriction — the uncertainty of a lower-bound on the short-run oil demand elasticity for use, which is randomly sampled from a truncated Student t distribution — whilst relating Kilian & Murphy's (2014) set of identification restrictions. Due to the inclusion of this additional restriction, C-BSVARs not only narrow the distribution of impulse response functions, but shift the estimates of oil demand elasticities towards zero, and has the desirable features of being insensitive to the seeds. Note also that the second specification provides a novel way of identifying key structural parameters' uncertainty, which are non-linear.

Each chapter is self-contained, as described above, and thus each chapter will provide an introduction including the specific motivation for each method, and contain the necessary information for the reader to understand the intuition of the methods proposed. Further technical details and additional empirical results are illustrated in the appendices. Chapter 2 has been published as Garratt, A., Vahey, S. P., & Zhang, Y. (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, 34(3), 456–462. Chapters 3 and 4 will be subsequently submitted to academic journals.

Chapter 2

Point Forecasts of Real Crude Oil Prices: Combinations for Baumeister & Kilian's (2015) Model Space

Abstract

Baumeister & Kilian (2015) combine forecasts from six empirical models to predict real oil prices. Over their 1992:01 to 2012:09 evaluation period, they demonstrate that forecast combinations are substantially more accurate than no-change forecasts at horizons up to 18 months. In this chapter, I begin by broadly reproducing their main economic findings, employing their preferred measures of the real oil price and similar real-time variables. Mindful of the importance of Brent crude oil as a global price benchmark and the divergence in oil price measures since 2011, I extend consideration to the North Sea based measure and update the evaluation sample to 2016:12. I find that the combined forecasts for Brent match the precision of Baumeister & Kilian's own combination forecasts (for other their preferred oil price measures). The extended sample using the oil price measures adopted by Baumeister & Kilian (2015) yields similar results to those reported in their paper. The real-time data set for this replication is available for download from the web of National Institute of Economic and Social Research (NIESR).

2.1 Introduction

There are three notable features of the oil market since 2011 the divergence between Brent crude, the U.S. refiners' acquisition cost (RAC) for crude oil imports, and the West Texas Intermediate (WTI) measures of the oil price; the convergence of the measures since 2014:08; and the rapid fall in these measures in the summer of 2014.

Figure 2.1: Monthly real crude oil prices (nominal prices divided by the U.S. consumer price index (CPI))

These characteristics are illustrated in Figure 2.1, which displays the real oil price since 2010. The divergence of crude oil price measures, with Brent oil prices persistently higher than the RAC and WTI, is most pronounced after 2011:01 and leading up to the oil price fall, after which the differential is reduced while nonetheless remaining a feature. Baumeister & Kilian (2015) argue that the WTI measure has suffered from structural instability following the restrictions on U.S. crude oil exports and transportation bottlenecks since 2011, highlighting the lack of arbitrage between WTI and Brent. The reduced divergence after 2014:06 reflects partially the repeal of the oil export ban at the end of 2015 and the boom in the U.S. tight oil production, reducing the U.S. crude oil demand from imports and stimulating exports of refined products (Kilian, 2017).

Many candidate explanations for the sharp drop in the nominal (real) oil price since 2014:07 across all three measures have been proposed in the literature.

The increased supply of oil, with the growth in North American shale oil production and the boom of U.S. crude oil exports since 2014, coupled with a reduced sensitivity to geopolitical uncertainties in Russia, Iraq, Iran, Libya and the organization of the petroleum exporting countries (OPEC) affecting the elasticity of supply; see ECB (2014). Baumeister & Kilian (2016a,c) emphasise the importance of weak demand for crude oil in particular for China and the European Union, together with depressed expectations. Baumeister & Kilian's (2015) combination provides a replicable econometric methods for the real crude oil prices forecasts, given a high weights on demand side, over the sample period 1992:01 to 2012:09.

This chapter repeats the analysis presented by Baumeister & Kilian (2015) comparing the forecasting performance of six econometric models for the real oil price, individually and in combination relative to a no-change benchmark model, over an evaluation period from 1992:01 to 2016:12. Wang et al. (2017) also proposed combinations, based on single predictor time-varying parameter models, utilising the most variables in Baumeister & Kilian (2015), and illustrated the predictability of WTI and RAC measures over an evaluation period from 1992:01 to 2015:12. Two conspicuous features of their analysis are: (1) their forecasts results do not perform significantly different with Baumeister & Kilian (2015)'s combination, and, (2) they just extended the data back to 1986:01, while Baumeister & Kilian (2015) consider the most series back to 1973:01. Hence, a multivariate real-time database vital for subsequent research on the oil market is constructed with this thesis. The database collates real-time measurements by data vintage for variables similar to those described by Baumeister & Kilian (2012), where 2017:06 represents the last observation for all data series.[1] The data provides qualitatively similar but not identical results to those reported by Baumeister & Kilian (2015) over their sample.

Two extensions to the analysis are considered: first, the robustness of the results are examined utilising the Brent, as opposed to the WTI and RAC oil price measures; and second, the extension of the evaluation period to end in 2016:12. The switch to the Brent measure results in similar predictability relative to measures preferred in Baumeister & Kilian (2015). Predictability is generally not substantially altered by consideration of the longer evaluation period.

The remainder of this chapter is structured as follows. The next section sum-

[1] National Institute of Economic and Social Research (NIESR) has offered to host the real-time data set.

marises the real-time oil market data set, followed by a description of Baumeister & Kilian's (2015) models and combinations. The subsequent section describes the results, recursively extending to a longer evaluation period, before the final section concludes the chapter.

2.2 A real-time data set for oil price forecasting

When constructing the real-time oil market data set, I broadly follow the nowcast and backcast methods described in Baumeister & Kilian (2012, 2015), with the main differences between my data set and Baumeister & Kilian's (2012) being the extended sample to 2017:06 and the inclusion of the Brent crude oil price. The data emerging from this thesis provides economically similar although not identical results to those reported by Baumeister & Kilian (2015) both over their sample and my extended sample, with their preferred measure of oil prices.

In this chapter, I utilise 18 time-series data, split into series with and without revisions and briefly summarised in Tables 2.1 and 2.2. These series are available in real time and without any further revisions including crude oil futures prices, oil price forecasts from the *Short-Term Energy Outlook* published by the U.S. Energy Information Administration (EIA), gasoline price, heating oil price, and an index of industrial raw materials. However, the rest of the series are revised after the first release due to the publication delay between the date of the time series observation and its release. Following the conventional terminology in the real-time macroeconomic forecasting literature, I define "vintage of data" as the historical time series observed by forecasters at a specific point in time (known as the "vintage date"); for example, the 2017:06 vintage includes observations only available at the end of June 2017. There are 307 vintages in total.

Table 2.1 summarises the series' unit of measurement, frequency, maturities (or forecast horizons), first observation $\underline{\tau}$, the last observation $\bar{\tau}$, whether the series have missing values, and the source key in the resources under each column. Meanwhile, the resources of each series are detailed in supporting notes below the table. The inclusion of EIA forecasts' data facilitates replication of the comparison between the quarterly forecasts from Baumeister & Kilian's (2015) combination and the EIA's quarterly forecasts of RAC, as detailed in Section 2.4.6. Baumeister & Kilian (2015) specify EIA quarterly forecasts according to the timing of the

Table 2.1: Data (without subsequent revisions) resources for observations from 1991:12 to 2017:06

Series	Unit of measure-ment	Frequency	Maturity/Forecast horizon(s)	$\underline{\tau}$	$\bar{\tau}$	Missing values	Source key
WTI futures	Dollars per Barrel	Monthly	1–24 months	Maturities 1-17 months: 1991:12 /Maturities 18 and 21 months: 1992:01 /Maturities 19, 20 and 22-23 months: 1991:12 /Maturity 24 months: 1992:04[1]	2017:06	Maturities 1-17 months: NO /Maturities 18 and 21 from 1991:12 to 1995:06: YES /Maturities 19, 22 and 24 from 1991:12 to 1995:08: YES /Maturities 20 and 23 from 1991:12 to 1995:07: YES	CL1:COM to CL24:COM
Brent futures	Dollars per Barrel	Monthly	1–24 months	Maturities 1-8 months: 1991:12 /Maturity 9 months: 1992:02 /Maturities 10-11 months: 1994:04 /Maturity 12 months: 1994:07 /Maturities 13, 16, 19, and 22 months: 1997:12 /Maturities 15, 18, and 21 months: 1998:01 /Maturities 14, 17, 20, and 23-24 months: 1998:02	2017:06	Maturities 1-8 months: NO /Maturities 9-11 from 1991:12 to 1994:03: YES /Maturities 12, 15, 18, and 21 from 1991:12 to 2004:12: YES /Maturities 13, 16, 19, and 22 from 1991:12 to 2004:11: YES /Maturities 14, 17, 20, and 23-24 from 1991:12 to 2005:01: YES	CO1:COM to CO24:COM
Wiki crude oil futures[2]	Dollars per Barrel	Monthly	1–24 months	1991:12	2017:06	NO	CHRIS-CME_CL1 to CHRIS-CME_CL24
RAC forecasts	Dollars per Barrel	Quarterly	Convention I: 1-7 quarters /Convention II: 1-8 quarters	Conventions I&II (1-4 quarters): $1991Q4$[3]	Conventions I&II (1-4 quarters): $2017Q1$	Conventions I&II (1-4 quarters): NO /Conventions I&II (5-7(8) quarters): YES	Energy Prices
WTI forecasts	Dollars per Barrel	Quarterly	Convention I: 1-7 quarters /Convention II: 1-8 quarters	Conventions I&II (1-4 quarters): $1991Q4$	Conventions I&II (1-4 quarters): $2017Q1$	Conventions I&II (1-4 quarters): NO /Conventions I&II (5-7(8) quarters): YES	Energy Prices
Brent forecasts	Dollars per Barrel	Quarterly	Convention I: 1-7 quarters /Convention II: 1-8 quarters	Conventions I&II (1-4 quarters): $1991Q4$	Conventions I&II (1-4 quarters): $2017Q1$	Conventions I&II (1-4 quarters): NO /Conventions I&II (5-7(8) quarters): YES	Energy Prices
gasoline prices	Dollars per Gallon	Monthly	−−	1973:01	2017:06	NO	eer_epmru_pf4_y35ny_dpgd
Heating oil prices	Dollars per Gallon	Monthly	−−	1973:01	2017:06	NO	eer_epd2f_pf4_y35ny_dpgd
Industrial raw materials prices (non-oil)	INDEX	Monthly	−−	1973:01	2017:06	NO	BVY00

Note: These data sets are not not revised. Here I use 307 observations commencing in 1991:12 and ending in 2017:06. The data for each of them are $\underline{\tau}, ..., \bar{\tau}$, where $\underline{\tau}$ and $\bar{\tau}$ are the first and last observations.

[1] First observations with different futures' maturities are not available at the same time. Hence, I separately report the first observations at different maturities with a forward slash.

[2] Wiki crude oil futures are employed for the backcast WTI and Brent futures. (Details of the backcast are available in Appendix A.3.2.)

[3] Here I only consider the EIA forecasts without missing values. At forecast horizons beyond 4 quarters, here are missing values for the both two timing conventions, which is not available on 1991Q4. Hence the timing of the first observation of EIA quarterly forecasts ($\underline{\tau}$) only presents 1-4 quarters in this table.

RESOURCES: WTI and Brent futures are collected from Bloomberg. Wiki crude oil futures are from Quandl. RAC, WTI, and Brent quarterly forecasts are published by the *Short-Term Energy Outlook* from the U.S. Energy Information Administration. Gasoline and heating oil prices are available in the web of the U.S. Energy Information Administration. Industrial raw material non-oil prices are stored in the Commodity Research Bureau.

Table 2.2: Real-time data (with subsequent revisions) resources for vintages dated 1991:12 to 2017:06

Series	Unit of measure-ment	$\underline{\tau}$	$\bar{\tau}$	Number of Obs.	The publication delay	Average number of Obs. revised	Mnemonic/Source key
Brent	Dollars per Barrel	1980:01	1991:11-2017:05	1991:12-2017:06: 143-449[1]	1	1	POILBREUSDM
WTI	Dollars per Barrel	1973:01	1991:11-2013:06 /2013:08-2017:06[2]	1991:12-2013:07: 227-486 /2013:08-2017:06: 488-534	1991:12-2013:07: 1 /2013:08-2017:06: 0	1991:12-2013:07: 1 /2013:08-2017:06: 0	WTISPLC
RAC	Dollars per Barrel	1974:01	1991:09-2005:04 /2005:06-2017:04[3]	1991:12-2005:07: 213-376 /2005:08-2017:06: 378-520	1991:12-2005:07: 3 /2005:08-2017:06: 2	3.89*	Crude oil price summary
World crude oil production	Thousand Barrels per Day	1975:01	1991:09-2017:03	1991:12-2017:06: 189-495	3	24.08*	International petroleum
The U.S. crude oil inventories	Million Barrels	1973:01	1991:11-2017:05	1991:12-2017:06: 227-533	1	3.01*	Petroleum stocks
The U.S. petroleum inventories	Million Barrels	1973:01	1991:11-2017:05	1991:12-2017:06: 227-533	1	3.20*	Petroleum stocks
The organisation for economic co-operation and development (OECD) petroleum inventories	Million Barrels	1973:01	1991:07-2002:05 /2002:07-2016:02 /2016:04-2017:03	1991:12-2002:10: 223-353 /2002:11-2016:06: 355-518 /2016:07-2017:06: 520-531	1991:12-2002:10: 5 /2002:10-2016:06: 4 /2016:07-2017:06: 3	13.68*	Petroleum stocks in OECD countries
The U.S. CPI for all urban consumers	INDEX	1973:01	1991:11-2017:05	1991:12-2017:06: 227-533	1	4.43	Consumer prices—All urban consumers
World real economic activity (rea) index[4]	INDEX	1973:01	1991:11-2017:05	1991:12-2017:06: 227-533	1	––[5]	BDIY:INDEX

Note: The vintage is defined as historical data observed at a specific point in time, for example, the 2017:06 vintage includes observations only available in June 2017. Here I use 307 vintages commencing in 1991:12 and concluding in 2017:06. The data for each vintage are $\underline{\tau}, ..., \bar{\tau}$, where $\underline{\tau}$ and $\bar{\tau}$ are the first and last observations available in vintages dated 1991:12 to 2017:06.

[1] The range shown after the colon is number of observations available in vintages. For example, here are 143 observations in vintage 1991:12.

[2] WTI in the Federal Reserve Economic Data (FRED) data base combines two resources: 1) the Wall Street Journal's commodity energy prices with one-month lag, discontinued on August 2013; and 2) the monthly oil spot prices reported by the U.S. Department of Energy are calculated as the average of the daily observations.

[3] Dates separated by a slash is because the publication delay between the date of the time series observation and its release has changed in EIA.

[4] The world real economic activity (rea) index includes: 1) the average monthly single voyage freight rates for the period between 1970:05 and 1985:11, collected by Drewry Shipping Consultants Ltd; and 2) the nominal shipping rate raw data (the Baltic Dry Cargo Index), available post-1985.01. Details of the construction are available in Appendix A.2.

[5] Due to the construction of the rea index, including deflation by the U.S. CPI and detrending, revision occur for every observations in each vintage.

Sign * denotes the consideration of yearly revisions, which differs from Baumeister & Kilian (2012, 2015).

RESOURCE: Bent and WTI prices are available in the FRED database. RAC, the world crude oil production, the U.S. crude oil inventories, the U.S. petroleum inventories, and OECD petroleum inventories are published by the U.S. Energy Information Administration monthly energy review. The U.S. consumers price index for all urban consumers is from the *Economic Indicators* published by Council of Economic Advisers and the real time dataset provided by the Federal reserve bank of Philadelphia. The world real economic activity index is combined from Drewry Shipping Consultants Ltd and Bloomberg as mentioned above.

Short-Term Energy Outlook issues, a monthly EIA publication. Timing convention I employs the end-of-quarter issues of the publication (i.e., March, June, September and December), while timing convention II utilises the first month of the following quarter (i.e., April, July, October and January), which is consistent with the definition in Baumeister & Kilian (2015). This study uses 307 observations commencing in 1991:12 and ending in 2017:06. Moreover, Wiki crude oil futures are employed for backcasting the missing values of WTI and Brent futures, with details of the backcast available in Appendix A.3.2.

Series with revisions are summarised in Table 2.2, which are stored in vintages. Columns include unit of measurement, first observation $\underline{\tau}$, the last observation $\bar{\tau}$, the number of observations in a vintage, the publication delay (between the date of the time series observation and its release, e.g. at time t+1, if I observe time t, then the number of lags or of publication delays equals 1), the average number of observations revised, and the source key in the resources. The revision process will be introduced in Subsection 2.2.1, which is not identical with Baumeister & Kilian (2012) and specified by an asterisk ($*$), and resources of each series are listed in the footnotes of the table.

The following subsections will summarise the revisions, and provide details for the nowcasting method. The revision and nowcasts implemented in the real-time data set influence the point forecast accuracy, which is illustrated in Section 2.4.5. Moreover, Appendix A fills the details of processes for constructing the U.S. consumer price index for all urban consumers (CPI), and a world real economic activity index based on the Bulk Dry Cargo Ocean shipping freight rates, which is not identical but qualitatively indifferent with from the index proposed in Kilian (2009). Additionally, the process of backcasts is in Appendix A.3.

2.2.1 Revisions

Each monthly issue of the *Monthly Energy Review*, published by the EIA, reports the corresponding vintage of five oil market data series considered in this chapter including RAC, the world crude oil production, the U.S. crude oil inventories, and the U.S. and the organisation for economic co-operation and development (OECD) petroleum inventories. The data differs from the preceding issue due to partially historical revisions and newly released observations. In all cases, the EIA reports monthly observations to a maximum of three years' worth of deep history. Baumeis-

ter & Kilian (2012, p. 327) report that they only used monthly revisions; however, the EIA provides a more comprehensive history by recording the revised measurements for the annual growth rates, with varying start dates. In contrast, I also utilise the EIA's annual revised historical data while assuming equal monthly revisions over the year, which effectively smooths the unobserved "true" monthly revisions over the twelve months.

Before illustrating the revisions, it is necessary to clarify the definition of the in-report sample, and the ex-post revised data, which is specified in Baumeister & Kilian (2012). Baumeister & Kilian (2015) utilised the same data set in Baumeister & Kilian (2012), and the most recent data set available at time of the paper was for 2013:03. Their ex-post observation defined as the 2013:03 vintage, discarding the last six months data as these data are still preliminary. Consequently, their in-report sample is vintages 1991:12–2012:09. In this chapter, I extend the sample to the 2017:06 vintage. Hence, the in-report sample in this exercise is vintages 1991:12–2016:12, and the post-report vintage is 2017:06. Considering this chapter is a replication, I illustrate the descriptive statistics on data revisions for Baumeister & Kilian's (2015) and the extended sample in upper and lower panels of Table 2.3 respectively.

Consistent with Baumeister & Kilian (2012), I present in- and post-report revisions. The in-report revision is calculated as the percentage change between the most recent vintage available for a value in subsequent issues of the *Monthly Energy Review* within the in-report sample, and the value released on the first vintage. For example, considering a value is first released in the 1991:12 vintage. The in-report revision for this value is the average of percentage changes between the value reported in subsequent vintages from 1992:01 to 2016:12 and its first released level. The number of first released values is identical with the number of vintages. More specifically, the number of observations (or of the in-sample vintages) in Baumeister & Kilian's (2015) sample is 250, while in the extended sample is 301. Therefore, the in-report average revisions (percentage changes) reported in the first column of Table 2.3 is the average of in-report revisions for the corresponding samples.

Table 2.3: Descriptive statistics on data revisions

Variables	Ave. Revisions (percentage change)		Std. Dev. relative to ex-post data (levels)		Subsequent revisions' magnitude			Ave. number of in-report revisions
	In-report	Post-report	In-report	Post-report	1^{st}	2^{nd}	3^{rd}	
	In-report vintage: 1991:12–2012:09 (post-report vintage: 2013:03)							
The U.S. Consumer Price Index	−0.057%	−0.015%	0.338%	0.191%	−0.027	−0.029	0.005	3.98
The U.S. refiners' acquisition cost of crude oil imports	0.062%	0.034%	2.940%	1.254%	0.073	0.055	−0.213	3.86
World crude oil production	0.138%	0.160%	0.958%	1.422%	23.015	19.431	20.659	18.49
The U.S. crude oil inventories	−0.231%	−0.284%	1.953%	2.162%	−1.559	−0.475	0.266	2.58
The U.S. petroleum inventories	−0.020%	−0.064%	4.406%	2.774%	−1.686	4.819	1.057	2.80
OECD petroleum inventories	−0.141%	0.375%	3.124%	3.892%	−1.813	−1.976	−2.092	11.98
	In-report vintage: 1991:12–2016:12 (post-report vintage: 2017:06)							
The U.S. Consumer Price Index	−0.051%	−0.033%	0.260%	0.038%	−0.047	−0.026	0.004	4.44
The U.S. refiners' acquisition cost of crude oil imports	0.032%	−0.020%	2.495%	0.040%	0.024	0.006	−0.173	3.89
World crude oil production	0.150%	0.201%	1.214%	0.513%	29.355	26.399	27.383	24.08
The U.S. crude oil inventories	−0.233%	−0.701%	2.935%	2.075%	−0.877	−0.878	−4.378	3.01
The U.S. petroleum inventories	0.004%	−0.259%	4.539%	2.855%	−0.826	4.603	−1.760	3.20
The OECD petroleum inventories	−0.071%	0.384%	2.837%	3.223%	−0.598	−0.733	−0.830	13.68

Note: The descriptions of each column is detailed in text. In- and post-report revisions' average of percentage changes and its standard deviation relative to ex-post data are lower than 5% and essentially zero for all variables considered in both Baumeister & Kilian's (2015) and extended samples. Extending the sample of the real-time data set, the average number of in-report revisions is increasing, indicating a raising frequency of revisions. The reasons of subsequent revisions are reported in corresponding part of issues of *Monthly Energy Review*.

The post-report revisions are distinguished from the in-report revisions, and measured as the average of percentage changes between a value reported in ex-post vintage and the corresponding value reported in all in-sample vintages after its first released vintage. More specifically, the post-report revision of the value first released in the 1991:12 vintage is calculated as the average of percentage changes between its value reported in ex-post vintage and in vintages from 1992:01 to 2016:12. The post-report average revisions (percentage change) reported in the second column of Table 2.3 is the average of post-report revisions for values first released in each vintage within the in-report sample. Standard deviations of the in- and post-report revisions relative to the standard deviation of the ex-post data are reported in the third and fourth columns in Table 2.3. However, the in- and post-report revisions are difference levels instead of percentage changes, which gives a direct comparison with respect to ex-post observations. I also present the average of first, second and

third revisions' magnitude at the corresponding series' unit of measurement listed in Table 2.2. The last column provide the average number of in-report revisions. Consistent with Baumeister & Kilian (2012) and as shown in Table 2.3, the main conclusion is that although the revisions are tiny relative to its observations, but this is unforeseeable. The influence of the revisions' magnitude on foresting accuracy may be limited; however, the timing of the revisions influences the forecasting accuracy. I will detail the forecasting sensitivity relative to the timing of revisions in Section 2.4.5.

2.2.2 Nowcasts

Baumeister & Kilian (2012) propose a nowcasting method that is able to dramatically improve the out-of-sample forecast accuracy. The nowcast of real-time RAC is different with the other oil market variables, because RAC can be nowcasted with the other crude oil price measure, like WTI. Hence, I will introduce method for RAC and the other oil market data, respectively.

The real-time RAC becomes available with a lag of 3 months from 1991:12 to 2005:07, and with a lag of 2 months post-2005:07. Therefore, the most recent observations of RAC is extrapolated at the rate of growth of the WTI price through:

1. Calculating the growth rate of WTI, denoted as r_t^{WTI} at time t:

$$r_t^{WTI} = \frac{(WTI_t - WTI_{t-1})}{WTI_{t-1}},$$

where $t = [\bar{\tau}-2, \bar{\tau}-1, \bar{\tau}]$ for vintages from 1991:12 to 2005:07, and $t = [\bar{\tau}-1, \bar{\tau}]$ for the remaining vintages, while $\bar{\tau}$ is the final month in each vintage; for example, $\bar{\tau}$ is December, 1911 for the 1991:12 vintage.

2. Nowcasting the delayed values $\widehat{RAC_t}$ through

$$\widehat{RAC_t} = (1 + r_t^{WTI})RAC_{t-1},$$

where $t = [\bar{\tau} - 2, \bar{\tau} - 1, \bar{\tau}]$ for vintages 1991:12 to 2005:07, and $t = [\bar{\tau} - 1, \bar{\tau}]$ for the remainder.

Moreover, the oil market data published in the *Monthly Energy Review* (denominated as X_t in vintage t) includes the world crude oil production, the U.S.

crude oil inventories, the U.S. petroleum inventories, and the OECD petroleum inventories. These data have different lag lengths between the date of the time series observation and its release, as detailed in Table 2.2 and denoted as lag. To echo Baumeister & Kilian (2012), I extrapolate them based on the average rate of change in history up to that point in time through:

$$\widehat{X_{t-j}} = X_{t-lag} + \frac{\sum_{i=\underline{\tau}}^{n-lag}(X_{i+1} - X_i)}{t - \underline{\tau} - lag} * (lag - j),$$

where $n = [227, \ldots, 527]$ in vintages t=[1991:12, ..., 2016:12] respectively, and j is from zero to the lag length $lag - 1$. $\underline{\tau}$ is the timing of the first observation in each vintage defined in Table 2.2, and X_{t-lag} is the final observation in each vintage.

Additionally, to ensure robustness and sensitivity checks, I alternatively use a rolling window for the nowcasts in Section 2.4.5, applying $\underline{\tau} = [t-13, t-25, t-61]$ for the fixed window size of 12, 24, and 60 months in the equation above, respectively.

2.3 The model space and combinations

Baumeister & Kilian (2015) use forecast combination to mitigate issues of model misspecification. They combine six specifications using an equal weight and inverse mean squared predictive error (MSPE) weights. The six specifications include, an unrestricted global oil market vector autoregression (VAR), a commodity price model, an oil price futures spread model, a gasoline spread model, a time varying parameter (TVP) product spread, and a no-change benchmark model; see details below or in Baumeister & Kilian (2015).

Baumeister & Kilian (2015) forecast the nominal crude oil price deflated by the U.S. CPI. The forecasts of $\widehat{R}^{oil}_{t+h|t}$ are based on information at time t for the period $t + h$, where h is the forecast horizon. The forecast is a "point forecast" combination from six different model specifications without any restriction:

$$\widehat{R}^{oil}_{t+h|t} = \sum_{k=1}^{6} w_{k,t}\widehat{R}^{k}_{t+h|t}, \tag{2.1}$$

where the weights, $w_{k,t}$, are assigned to model k at time t. An equal weight, $w_{k,t} = \frac{1}{6}$,

and MSPE based weights are used, where the latter are defined as:

$$w_{k,t} = \frac{m_{k,t}^{-1}}{\sum_{j=1}^{6} m_{j,t}^{-1}};$$

and, $m_{k,t}^{-1}$ is the inverse MSPE of model k calculated with respect to observed outcomes available at time t. The use of inverse MSPE weights has a long history (e.g. Diebold & Pauly, 1987; Stock & Watson, 2004). Inverse MSPE weights are calculated at time t for forecasts at horizon h and are evaluated over a sample up to $t-(6+h)$, with equal weight utilised on the initial $h+6$ months, which is consistent with Baumeister & Kilian (2015). The practice addresses the issue of robustness to model misspecification and structural change, as per Baumeister & Kilian (2015), through adopting a recursive approach and utilising rolling fixed window sizes of 36, 24 and 12 months.

Moreover, $\widehat{R}^{k}_{t+h|t}$ is the forecasts from the k^{th} specification, and the six models are defined as:

1. An unrestricted global oil market VAR:

$$\widehat{R}^{1}_{t+h|t} = exp(\widehat{r}^{\text{VAR}}_{t+h|t}), \tag{2.2}$$

   where $\widehat{r}^{VAR}_{t+h|t}$ is the forecast of log measures based on an unrestricted global oil market VAR model, with the following four variables: the percent change in the global crude oil production, a business cycle index of global real activity (rea), the log of the RAC oil price deflated by the log of the U.S. CPI, and the change in global crude oil inventories. The forecasts of the RAC measure are predicted from the endogenous variable of the VAR, while forecasts of WTI and Brent are calculated as the spread between the prices and RAC multiplied by the forecast of RAC respectively.

2. A commodity price based model:

$$\widehat{R}^{2}_{t+h|t} = R^{oil}_{t}(1 + \pi^{h,\text{raw}}_{t} - E_t(\pi^{h}_{t+h})), \tag{2.3}$$

   where $\pi^{h,\text{raw}}_{t}$ is the difference between the log price of non-oil industrial raw materials at t and $t-h$, and R^{oil}_{t} is the real oil price measure at time t. Baumeister & Kilian (2015) define $E_t(\pi^{h}_{t+h})$, the expected U.S. inflation, as

the average U.S. CPI inflation available at time t, where the averaging begins in 1986:07, hence:

$$E_t(\pi^h_{t+h}) = [1 + \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}}^{\overline{\tau}} (ln(CPI_{t+1}) - ln(CPI_t))]^h - 1, \tag{2.4}$$

where $t = [\underline{\tau}, ..., \overline{\tau}]$, $\underline{\tau}$=1986:07 and $\overline{\tau}$ is the month of the final observation in a specific vintage.

3. An oil futures spread model:

$$\widehat{R}^3_{t+h|t} = R^{oil}_t(1 + f^h_t - s_t - E_t(\pi^h_{t+h})), \tag{2.5}$$

where s_t is the corresponding log of monthly WTI spot price, and $E_t(\pi^h_{t+h})$ is the inflation expectation introduced in Equation (2.4). f^h_t is the log of oil price futures with the maturity h observed at t. Forecasts of WTI and RAC are based on WTI futures, but the forecast of the Brent measure is based on Brent futures. The monthly oil futures prices for WTI with different maturities are the average of daily futures closed prices collected from Bloomberg. Missing values constrain the inclusion of forecasts based on oil futures in Baumeister & Kilian's (2015) forecast combinations at forecast horizons between 18 and 24 months. Plausible values for missed futures are backcasted according to Wiki Futures, which is available at https://www.quandl.com/data/CHRIS-Wiki-Continuous-Futures. (Details of the backcast are explained in Appendix A.3.2.) Therefore, forecasts based on oil futures at horizons beyond 18 months are included in this study.

4. A gasoline spread based model:

$$\widehat{R}^4_{t+h|t} = R^{oil}_t exp\{\widehat{\beta}[s^{gas}_t - s_t] - E_t(\pi^h_{t+h})\}, \tag{2.6}$$

where s^{gas}_t is the log spot price of gasoline which is the production of crude oil; $E_t(\pi^h_{t+h})$ again is the inflation expectation formulated in Equation (2.4); and, $\widehat{\beta}$ is estimated from the regression $\Delta s_{t+h} = \beta[s^{gas}_t - s_t] + \varepsilon_{t+h}$, employing ordinary least squares. Δs_{t+h} is the h-period ahead log-difference of spot WTI prices$(s_{t+h} - s_t)$.

5. A TVP product spread model:

$$\widehat{R}^5_{t+h|t} = R^{oil}_t exp\{\widehat{\delta_{1t}}[s^{gas}_t - s_t] + \widehat{\delta_{2t}}[s^{heat}_t - s_t] - E_t(\pi^h_{t+h})\}, \tag{2.7}$$

where the TVP $\widehat{\delta_{1t}}$ and $\widehat{\delta_{2t}}$, are estimated from:

$$\Delta s_{t+h} = \delta_{1t}[s_t^{gas} - s_t] + \delta_{2t}[s_t^{heat} - s_t] + e_{t+h}.$$

The s_t^{gas} and s_t^{heat} are the log spot prices of gasoline and heating oil, respectively, and $e_{t+h} \sim NID(0, \sigma^2)$, σ^2 is the constant variance of the error term. The TVP model of gasoline and heating oil spreads, motivated by Reeve & Vigfusson (2011) and developed by Baumeister et al. (2013), employs independent Normal-Wishart prior and the Gibbs sampling algorithm for the forecasts. The estimation uses a Bayesian approach involving the Gibbs sampler. Moreover, $E_t(\pi_{t+h}^h)$ is the inflation expectation introduced in Equation (2.4).

6. No change or random walk model:

$$\widehat{R}_{t+h|t}^6 = R_t^{oil}. \tag{2.8}$$

The no-change forecast is not only included in the combinations, but also acts as the benchmark in Baumeister & Kilian (2015) and my exercise.

Each specification is estimated over different samples, where Baumeister & Kilian (2015) include the maximum number of observations, as opposed to identifying a common estimation period. For example, the estimation of the VAR model in Equation (2.2) used data from 1973:01, when oil market dataset is first available. However, the TVP product spread model is estimated using a sample beginning in 1986:07 from Equation (2.7), due to gasoline and heating oil data are first available in July 1986.

2.4 Empirical results

Baumeister & Kilian (2015) evaluate the forecasts of WTI and RAC oil prices over the 1992:01 to 2012:09 evaluation period using the data observed in 2013:03 as the target variable. In following subsection, I will illustrate the narrow replication based on the same period but using my real-time data generates similar findings to Baumeister & Kilian (2015). Then, I extend the sample to 2016:12, followed by the extension to the Brent measure.

2.4.1 Narrow replication of the sample period 1992:01—2012:09

In a narrow replication, my data set generates quantitatively similar results to those reported by Baumeister & Kilian (2015). The same as Baumeister & Kilian (2015), recursive MSPE and success ratios are utilised for the evaluation of point and sign forecast accuracy over the evaluation period 1992:01–2012:09. The recursive MSPE ratio is defined as a ratio of the recursive MSPE of various combinations relative to the MSPE of no-change forecast. Hence, if a combination's MSPE ratio is below 1, the forecast is more accurate than the benchmark. A rough guide of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is used for statistically significant tests.

Success ratios describe the directional accuracy of the forecast accuracy. Under the null hypothesis that there is no directional forecast accuracy, the model would be no more successful at predicting the sign of change in crude oil prices relative to tossing a fair coin. Then, a success ratio higher than 0.5 indicates that the forecast combination would improve directional accuracy relative to the benchmark. The Pesaran & Timmermann (2009) test are utilised for the null hypothesis of no directional accuracy. The results for equal, recursive MSPE and rolling MSPE weights utilising window sizes of 36, 24, and 12 months are reported in the columns of Tables 2.4 and 2.5.

As shown in the tables, WTI and RAC oil price measures with equal weights have lower MSPE ratios and higher success ratios relative to inverse-MSPE-weights forecasts at the majority of horizons. Recursive MSPE ratios show the predictive power of point forecast of RAC, as shown in the upper panel of Table 2.4. For example the equal weight dominates the benchmark at horizons within 18 months excepting at horizon 6 months, significantly at 10% level based on a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test. The similar results for the WTI oil price measure are shown in the upper panel of Table 2.5. The highest recursive MSPE ratio reduction of real crude oil price forecasts is produced by the equally weighted forecasts of WTI at the 1-month horizon and is 9.7% lower. Notably, success ratios at all horizons under equal weight are higher than 0.5, indicating an improved sign forecasts relative to the benchmark. In the lower panel of Table 2.4, I can reject the null at least at 10% level for all horizons excepting at 6 and 9 months for RAC forecasts produced by equal weight combination based on the Pesaran & Timmermann (2009) test. The highest success ratio is 66.0%,

achieved by equal-weight forecasts of RAC at the 12-month horizon. Sign forecasts of WTI are illustrated in the lower panel of Table 2.5, and shown the similar results with the RAC oil price measure.

Table 2.4: Real-time forecast accuracy of baseline forecast combination based on all six forecasting models for the real U.S. refiners' acquisition cost for oil imports (the evaluation period: 1992:01–2012:09)

			Rolling weights based on windows of length		
MH	Equal weight	Recursive weights	36	24	12
		Recursive MSPE ratios			
1	**0.928****(0.029)	**0.933****(0.040)	**0.936****(0.046)	**0.929****(0.034)	**0.926****(0.031)
3	**0.921****(0.008)	**0.924****(0.009)	**0.925****(0.010)	**0.920****(0.005)	**0.926****(0.005)
6	**0.983**(0.153)	**0.988**(0.219)	**0.989**(0.237)	**0.991**(0.274)	**0.990**(0.251)
9	**0.976***(0.087)	**0.982**(0.145)	**0.980**(0.125)	**0.984**(0.181)	**0.985**(0.215)
12	**0.937****(0.000)	**0.943****(0.001)	**0.945****(0.003)	**0.944****(0.002)	**0.942****(0.001)
15	**0.930****(0.000)	**0.940****(0.001)	**0.952****(0.015)	**0.952****(0.013)	**0.972**(0.102)
18	**0.973***(0.068)	**0.994**(0.364)	1.022(0.847)	1.029(0.903)	1.059(0.989)
21	1.006(0.649)	1.028(0.952)	1.053(0.994)	1.060(0.996)	1.106(1.000)
24	**0.988**(0.238)	**0.997**(0.436)	1.000(0.503)	1.004(0.580)	1.045(0.958)
			Success ratios		
1	**0.558***(0.064)	**0.554***(0.079)	**0.558***(0.062)	**0.554***(0.084)	**0.554***(0.086)
3	**0.583****(0.018)	**0.587****(0.012)	**0.583****(0.017)	**0.587****(0.013)	**0.575****(0.034)
6	**0.545**(0.190)	**0.533**(0.309)	**0.525**(0.425)	**0.508**(0.652)	**0.529**(0.350)
9	**0.548**(0.123)	**0.556***(0.063)	**0.560***(0.056)	**0.548**(0.112)	**0.544**(0.123)
12	**0.660****(0.000)	**0.639****(0.000)	**0.630****(0.000)	**0.634****(0.000)	**0.660****(0.000)
15	**0.621****(0.000)	**0.596****(0.001)	**0.570****(0.019)	**0.562****(0.030)	**0.574****(0.018)
18	**0.578****(0.001)	**0.556****(0.003)	**0.517**(0.135)	0.500(0.317)	**0.509**(0.226)
21	**0.576****(0.002)	**0.511***(0.062)	**0.541****(0.041)	**0.528**(0.155)	0.489(0.676)
24	**0.544***(0.083)	**0.527***(0.053)	**0.544***(0.095)	**0.540**(0.182)	**0.535**(0.232)

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p-values of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported in brackets after recursive MSPE ratios. I also report p-values for the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

Table 2.5: Real-time forecast accuracy of baseline forecast combination based on all six forecasting models for the real WTI price (the evaluation period: 1992:01–2012:09)

			Rolling weights based on windows of length		
MH	Equal weight	Recursive weights	36	24	12
		Recursive MSPE ratios			
1	**0.903****(0.007)	**0.908****(0.008)	**0.909****(0.009)	**0.909****(0.011)	**0.906****(0.011)
3	**0.922****(0.010)	**0.926****(0.011)	**0.928****(0.013)	**0.928****(0.011)	**0.932****(0.012)
6	**0.986**(0.220)	**0.991**(0.292)	**0.992**(0.310)	**0.994**(0.358)	**0.998**(0.457)
9	**0.979**(0.119)	**0.984**(0.190)	**0.980**(0.135)	**0.984**(0.194)	**0.986**(0.247)
12	**0.945****(0.002)	**0.951****(0.006)	**0.949****(0.006)	**0.946****(0.004)	**0.940****(0.002)
15	**0.940****(0.001)	**0.953****(0.009)	**0.962***(0.050)	**0.962****(0.046)	**0.945****(0.008)
18	**0.972***(0.063)	**0.998**(0.447)	1.025(0.880)	1.041(0.960)	1.071(0.991)
21	1.004(0.600)	1.034(0.972)	1.056(0.994)	1.066(0.994)	1.087(0.999)
24	**0.980**(0.150)	1.001(0.521)	1.002(0.545)	1.010(0.660)	1.057(0.976)
			Success ratios		
1	**0.558***(0.081)	**0.550**(0.150)	**0.550**(0.150)	**0.554**(0.131)	**0.550**(0.157)
3	**0.543**(0.220)	**0.547**(0.198)	**0.555**(0.124)	**0.551**(0.167)	**0.534**(0.348)
6	**0.537**(0.271)	**0.520**(0.470)	**0.520**(0.504)	**0.508**(0.635)	0.500(0.699)
9	**0.556***(0.067)	**0.552***(0.075)	**0.556***(0.061)	**0.552***(0.082)	**0.531**(0.211)
12	**0.597****(0.002)	**0.580****(0.007)	**0.584****(0.009)	**0.597****(0.002)	**0.601****(0.002)
15	**0.579****(0.009)	**0.583****(0.003)	**0.562****(0.035)	**0.570****(0.025)	**0.570****(0.016)
18	**0.560****(0.011)	**0.552****(0.005)	**0.526**(0.113)	**0.509**(0.274)	**0.504**(0.286)
21	**0.585****(0.001)	0.493(0.146)	**0.520**(0.120)	**0.528**(0.200)	**0.528**(0.225)
24	**0.531**(0.185)	0.482(0.406)	**0.522**(0.225)	**0.509**(0.513)	**0.518**(0.403)

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p-values of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported in brackets after recursive MSPE ratios. I also report p-values for the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

2.4.2 Extension: the sample period 1992:01–2016:12 for RAC and WTI measures

The forecast accuracy of measures preferred by Baumeister & Kilian (2015) for the extended 1992:01 to 2016:12 evaluation period is illustrated in Tables 2.6 and 2.7. Generally, I find increased accuracy and significance level of real RAC and WTI forecasts relative to Baumeister & Kilian (2015). Owing to the inclusion of forecasts based on oil futures prices, there are significant gains in both the RAC and WTI real price forecast combinations relative to the no-change benchmark at horizons longer than 18 months. In particular, success ratios at all horizons under the equal weight are significantly higher than 0.5, based on the Pesaran & Timmermann (2009) test for the two measures, excepting at horizon 9 months for the RAC measure. I confirm the predictability of Baumeister & Kilian (2015) combination over the extended sample.

Table 2.6: Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real U.S. refiners' acquisition cost for oil imports

			Rolling weights based on windows of length		
MH	Equal weight	Recursive weights	36	24	12
			Recursive MSPE ratios		
1	**0.931****(0.016)	**0.935****(0.023)	**0.939****(0.030)	**0.933****(0.021)	**0.929****(0.017)
3	**0.921****(0.002)	**0.924****(0.002)	**0.928****(0.004)	**0.921****(0.001)	**0.921****(0.001)
6	**0.975****(0.041)	**0.981***(0.085)	**0.982***(0.080)	**0.978****(0.048)	**0.969****(0.022)
9	**0.968****(0.012)	**0.976****(0.046)	**0.971****(0.022)	**0.970****(0.024)	**0.966****(0.020)
12	**0.935****(0.000)	**0.945****(0.000)	**0.941****(0.000)	**0.930****(0.000)	**0.922****(0.000)
15	**0.931****(0.000)	**0.944****(0.000)	**0.944****(0.001)	**0.935****(0.000)	**0.929****(0.001)
18	**0.943****(0.000)	**0.963****(0.010)	**0.975***(0.082)	**0.972***(0.081)	**0.954***(0.054)
21	**0.970****(0.018)	**0.997**(0.413)	**0.999**(0.466)	1.001(0.529)	**1.000**(0.499)
24	**0.965****(0.013)	1.008(0.721)	**0.987**(0.202)	**0.988**(0.243)	1.005(0.573)
			Success ratios		
1	**0.550***(0.064)	**0.547***(0.078)	**0.567****(0.017)	**0.560****(0.031)	**0.563****(0.024)
3	**0.604****(0.001)	**0.597****(0.001)	**0.604****(0.001)	**0.604****(0.001)	**0.601****(0.001)
6	**0.569****(0.024)	**0.556***(0.071)	**0.553***(0.087)	**0.542**(0.151)	**0.563****(0.036)
9	**0.541**(0.102)	**0.548***(0.064)	**0.572****(0.009)	**0.562****(0.022)	**0.562****(0.020)
12	**0.651****(0.000)	**0.623****(0.000)	**0.633****(0.000)	**0.637****(0.000)	**0.668****(0.000)
15	**0.601****(0.000)	**0.584****(0.002)	**0.566****(0.016)	**0.573****(0.007)	**0.573****(0.007)
18	**0.583****(0.001)	**0.562****(0.009)	**0.523**(0.187)	**0.505**(0.396)	**0.509**(0.341)
21	**0.586****(0.001)	**0.514**(0.211)	**0.550****(0.037)	**0.532**(0.135)	0.489(0.681)
24	**0.556****(0.035)	**0.516**(0.254)	**0.552****(0.049)	**0.542**(0.105)	**0.538**(0.126)

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p-values of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported in brackets after recursive MSPE ratios. I also report p-values for the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

Table 2.7: Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real WTI price

			Rolling weights based on windows of length		
MH	Equal weight	Recursive weights	36	24	12
			Recursive MSPE ratios		
1	**0.905****(0.002)	**0.908****(0.002)	**0.910****(0.002)	**0.910****(0.003)	**0.908****(0.003)
3	**0.920****(0.002)	**0.923****(0.002)	**0.926****(0.003)	**0.925****(0.002)	**0.926****(0.002)
6	**0.976***(0.060)	**0.982**(0.108)	**0.981***(0.097)	**0.981***(0.092)	**0.987**(0.182)
9	**0.969****(0.019)	**0.976***(0.056)	**0.971****(0.026)	**0.971****(0.031)	**0.975***(0.068)
12	**0.940****(0.000)	**0.948****(0.000)	**0.941****(0.000)	**0.930****(0.000)	**0.935****(0.000)
15	**0.936****(0.000)	**0.948****(0.000)	**0.944****(0.001)	**0.934****(0.000)	**0.919****(0.000)
18	**0.943****(0.000)	**0.963****(0.012)	**0.971***(0.064)	**0.972***(0.096)	**0.985**(0.280)
21	**0.969****(0.020)	**0.994**(0.347)	**0.992**(0.338)	**0.992**(0.346)	**0.992**(0.372)
24	**0.952****(0.003)	**0.991**(0.284)	**0.965****(0.030)	**0.962****(0.028)	**0.993**(0.407)
			Success ratios		
1	**0.567****(0.024)	**0.563****(0.037)	**0.547**(0.116)	**0.550***(0.099)	**0.560****(0.049)
3	**0.574****(0.017)	**0.574****(0.018)	**0.581****(0.009)	**0.577****(0.014)	**0.570****(0.027)
6	**0.556***(0.067)	**0.539**(0.196)	**0.542**(0.174)	**0.539**(0.182)	**0.522**(0.375)
9	**0.575****(0.008)	**0.572****(0.011)	**0.575****(0.008)	**0.565****(0.022)	**0.558****(0.030)
12	**0.616****(0.000)	**0.606****(0.000)	**0.606****(0.000)	**0.612****(0.000)	**0.595****(0.001)
15	**0.591****(0.001)	**0.594****(0.001)	**0.584****(0.003)	**0.598****(0.001)	**0.587****(0.002)
18	**0.562****(0.009)	**0.565****(0.005)	**0.555****(0.023)	**0.537***(0.092)	**0.534**(0.105)
21	**0.607****(0.000)	**0.521**(0.106)	**0.546****(0.036)	**0.561****(0.021)	**0.568****(0.014)
24	**0.545***(0.075)	0.491(0.558)	**0.538**(0.107)	**0.531**(0.175)	**0.534**(0.143)

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p-values of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported in brackets after recursive MSPE ratios. I also report p-values for the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

2.4.3 Extension: the Brent measure

The Brent measure represents an increasingly important benchmark for the world crude oil price (see, among others, Morana, 2001; Alberola et al., 2008; Baumeister & Kilian, 2016c). Hence, this subsection extends the combination method proposed in Baumeister & Kilian (2015) to include the Brent measure, based on the 1992:01–2016:12 evaluation period. It is not a surprise that the combination method improves the forecasting performance for the Brent measure relative to no-change forecasts using recursive MSPE and success ratios.

Echoing the WTI and RAC results on the shorter evaluation period reported by Baumeister & Kilian (2015), I find evidence of significant predictability from forecast combinations for the Brent measure. As shown in the second column of Table 2.8, the MSPE and success ratios imply equal weights are more accurate than the no-change forecast at all forecast horizons from 1 to 24 months. The Pesaran & Timmermann (2009) test suggests a significant improvement of directional forecast accuracy relative to no-change forecast at all horizons. The results using MSPE

weights are similar to those for equal weights.

Table 2.8: Real-time forecast accuracy of baseline forecast combination based on all six forecasting models extended sample 1992:01 to 2016:12 for the real Brent price

			Rolling weights based on windows of length		
MH	Equal weight	Recursive weights	36	24	12
			Recursive MSPE ratios		
1	**0.925****(0.001)	**0.927****(0.001)	**0.929****(0.001)	**0.930****(0.001)	**0.934****(0.003)
3	**0.945****(0.006)	**0.948****(0.006)	**0.951****(0.010)	**0.954****(0.013)	**0.954****(0.008)
6	**0.991**(0.231)	**0.998**(0.423)	**1.000**(0.495)	**0.997**(0.405)	**0.993**(0.276)
9	**0.980***(0.058)	**0.989**(0.181)	**0.986**(0.129)	**0.979***(0.056)	**0.975****(0.046)
12	**0.948****(0.000)	**0.959****(0.001)	**0.954****(0.001)	**0.943****(0.000)	**0.926****(0.000)
15	**0.942****(0.000)	**0.957****(0.001)	**0.958****(0.003)	**0.955****(0.002)	**0.939****(0.000)
18	**0.957****(0.001)	**0.979***(0.061)	**0.990**(0.243)	**0.995**(0.360)	**0.987**(0.274)
21	**0.980***(0.056)	1.008(0.741)	1.012(0.815)	1.013(0.814)	1.012(0.710)
24	**0.972****(0.027)	1.009(0.776)	**0.999**(0.460)	**0.998**(0.426)	1.031(0.931)
			Success ratios		
1	**0.537**(0.201)	**0.537**(0.211)	**0.540**(0.183)	**0.547**(0.125)	**0.547**(0.131)
3	**0.581****(0.011)	**0.567****(0.033)	**0.560***(0.056)	**0.564****(0.044)	**0.567****(0.034)
6	**0.529**(0.280)	**0.502**(0.634)	**0.505**(0.601)	**0.522**(0.344)	**0.525**(0.290)
9	**0.562****(0.026)	**0.551***(0.052)	**0.568****(0.017)	**0.568****(0.017)	**0.568****(0.010)
12	**0.599****(0.000)	**0.595****(0.001)	**0.606****(0.000)	**0.599****(0.000)	**0.640****(0.000)
15	**0.612****(0.000)	**0.591****(0.001)	**0.570****(0.009)	**0.577****(0.005)	**0.633****(0.000)
18	**0.601****(0.000)	**0.558****(0.009)	**0.548****(0.032)	**0.548****(0.030)	**0.523**(0.128)
21	**0.568****(0.005)	0.500(0.347)	**0.543****(0.041)	**0.536***(0.089)	**0.521**(0.218)
24	**0.542***(0.076)	0.498(0.459)	**0.531**(0.144)	**0.509**(0.374)	**0.509**(0.386)

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p-values of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported in brackets after recursive MSPE ratios. I also report p-values for the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

2.4.4 Extension: recursive forecast accuracy of equal weighted combinations

This subsection evaluates the recursive forecast accuracy of equal weighted combinations for the three measures over an extending window from 1992:01–2008:03 to 1992:01–2016:12. Generally, I found the equal-weight combination consistently performs better than no-change forecasts for at all forecasting horizons.

In Figure 2.2, I plot the recursive MSPE and success ratios of the equal weight combinations at selected horizons (1, 6, 12, 18 and 24 months), evaluated in periods recursively extended from the period commencing with the financial crisis

2008:03 to 2016:12. The recursive evaluation periods overlapping with Baumeister & Kilian's (2015) can be found in the left of the y axis in each sub plot. The Brent measure is marked as solid line, while a straight horizontal dashed slim line indicates the benchmarks (1 for recursive MSPE ratios and 0.5 for success ratios). It is clear that both recursive MSPE and success ratios consistently indicate that equal weight combinations dominate the benchmark before and after the plunge of crude oil prices in 2014. With the increase of forecasting horizon length, the gain of combination relative to the benchmark is raising. For example, downward and upward trend of MSPE and success ratios after 2014:06 are far steeper at horizon 24 months than at 6 months respectively. The equivalent plots for WTI and RAC measures (shown in dotted and dashed lines) reveal the qualitatively similar predictability relative to Brent, MSPE and success ratios of the Brent measure show a slightly lower predictability than the other oil price measures at horizon 6 months though.

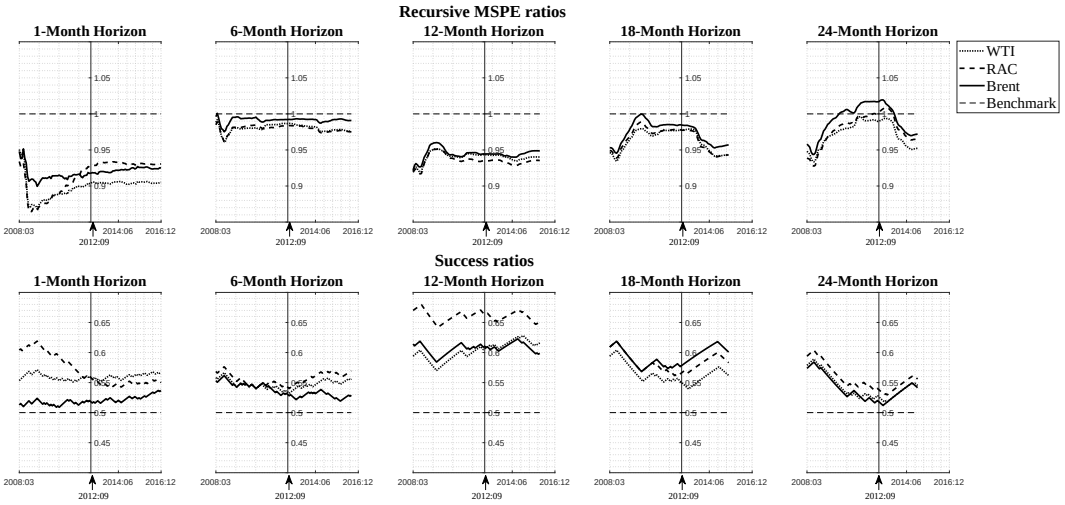

Figure 2.2: Historical recursive forecast accuracy of Baumeister & Kilian's (2015) equal weighted averages at selected horizons (recursive MSPE and success ratios)

The following subsection demonstrates the robustness and sensitivity analysis of the monthly forecasts based on the equal weight as well as the quarterly forecasts for the three measures.

2.4.5 Robustness and sensitivity

The revisions and the choice of nowcasting method would impact on the combinations' forecast accuracy. In this section I analyse the robustness and sensitivity of

the equal-weight combination with analysis of each individual model.

(I) Alternative revision assumptions

In this section, I present the performance of equal-weight combinations under different assumptions of revisions. The general observation in this subsection is that it is possible to reduce the forecasts' recursive MSPE ratios relative to the ratio of recursive revisions utilised in Baumeister & Kilian (2015) by an average of approximately 1.2% controlling the same evaluation period 1992:01–2012:09, and an average of approximately 2.6% over the period 1992:01–2016:12. The following paragraphs will illustrate different assumptions of revisions and corresponding point forecast results evaluated by recursive MSPE and success ratios.

As shown in Subsection 2.2.1, the revision is statistically unforeseeable, and distributed around zero. This is consistent with the description in Baumeister & Kilian (2012). However, different series are not revised over the same duration. For example, the OECD petroleum inventories may be continually revised for more than 50 months following the initial release, as some of the OECD countries correct or first report their petroleum inventories; while for RAC, most revisions are concluded within 5 months. Theoretically, not every revision is informative for out-of-sample forecasts, and while it is impossible to specify which revision is informative, there is space to posit three different assumptions regarding the timing of the revision: (1) All revisions are informative, and are labelled as 'recursive revisions' in Table 2.9. (2) Only the first revision is informative, which I refer to as a 'one-time revision'. More specifically, I employ the data in each vintage only revised once after the vintage of 1991:12, with any subsequent revisions after the first one disregarded. (3) The informative revisions occur once a quarter, and are referred to as 'four equal revisions (once every quarter)' in the tables, including the start quarter (January, April, July and October), mid quarter (February, May, August and November), and end quarter (March, June, September and December). This is motivated by the observation that variables, such as the U.S. CPI, are revised at quarterly frequency, and by the fact that different series are revised asynchronously, with the aim being to use the up-to-date information harmoniously; for example, I revise the data during every start quarter, with the revisions that occur in the mid-quarter vintages and end-quarter vintages ignored.

Table 2.9: Real-time forecast accuracy of equal weighted forecast combination based on all six forecasting models (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12) under different revision assumptions

					4 equal revisions, one ever quarter					
Revision assumptions	Recursive revision		One-time revision		Start quarter		Mid quarter		End quarter	
MH	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12
Real WTI price										
Recursive MSPE ratios										
1	**0.903***	**0.905***	1.014	**0.990**	1.024	**0.999**	**0.907***	**0.908***	**0.899***	**0.900***
3	**0.922***	**0.920***	**0.972**	**0.951***	**0.974**	**0.952***	**0.924***	**0.924***	**0.915***	**0.913***
6	**0.986**	**0.976***	1.011	**0.985**	1.012	**0.986**	**0.980**	**0.974***	**0.969**	**0.960***
9	**0.979**	**0.969***	1.004	**0.973***	1.007	**0.975***	**0.976**	**0.965***	**0.966***	**0.954***
12	**0.945***	**0.940***	**0.969***	**0.938***	**0.974***	**0.939***	**0.942***	**0.934***	**0.917***	**0.908***
15	**0.940***	**0.936***	**0.966***	**0.929***	**0.970***	**0.931***	**0.938***	**0.932***	**0.920***	**0.911***
18	**0.972***	**0.943***	1.007	**0.940***	1.012	**0.942***	**0.975***	**0.943***	**0.959***	**0.925***
21	1.004	**0.969***	1.044	**0.961***	1.049	**0.962***	1.004	**0.963***	**0.986**	**0.943***
24	**0.980**	**0.952***	1.020	**0.942***	1.021	**0.939***	**0.981**	**0.947***	**0.963***	**0.928***
Success ratios										
1	**0.558***	**0.567***	0.434	0.453	0.434	0.453	**0.554**	**0.563***	**0.538**	**0.553***
3	**0.543**	**0.574***	0.474	0.483	0.470	0.477	**0.547**	**0.560***	**0.547**	**0.567***
6	**0.537**	**0.556***	0.377	0.400	0.381	0.403	**0.549**	**0.563***	**0.545**	**0.556***
9	**0.556***	**0.575***	0.419	0.469	0.423	0.476	**0.531**	**0.565***	**0.585***	**0.603***
12	**0.597***	**0.616***	0.466	**0.502**	0.475	**0.512**	**0.588***	**0.609***	**0.563***	**0.588***
15	**0.579***	**0.591***	0.498	**0.528***	0.498	**0.535***	**0.587***	**0.615***	**0.570***	**0.594***
18	**0.560***	**0.562***	0.474	**0.509***	0.478	**0.512***	**0.547***	**0.558***	**0.573***	**0.572***
21	**0.585***	**0.607***	0.489	**0.521***	0.485	**0.511***	**0.572***	**0.596***	**0.603***	**0.621***
24	**0.531**	**0.545***	0.438	0.495	0.447	**0.509**	**0.540**	**0.563***	**0.535**	**0.567***
Real U.S. refiners' acquisition cost for oil imports										
Recursive MSPE ratios										
1	**0.928***	**0.931***	**0.992**	**0.972**	**0.997**	**0.977**	**0.930***	**0.933***	**0.923***	**0.924***
3	**0.921***	**0.921***	**0.962**	**0.941***	**0.967**	**0.944***	**0.922***	**0.925***	**0.913***	**0.912***
6	**0.983**	**0.975***	1.004	**0.978**	1.005	**0.978**	**0.973***	**0.969***	**0.965***	**0.958***
9	**0.976***	**0.968***	1.005	**0.968***	1.006	**0.968***	**0.973***	**0.964***	**0.965***	**0.952***
12	**0.937***	**0.935***	**0.970***	**0.931***	**0.972***	**0.929***	**0.936***	**0.928***	**0.916***	**0.905***
15	**0.930***	**0.931***	**0.960***	**0.920***	**0.964***	**0.920***	**0.934***	**0.928***	**0.918***	**0.908***
18	**0.973***	**0.943***	1.012	**0.935***	1.017	**0.935***	**0.979**	**0.939***	**0.968***	**0.924***
21	1.006	**0.970***	1.050	**0.958***	1.056	**0.956***	1.013	**0.964***	**0.998**	**0.945***
24	**0.988**	**0.965***	1.029	**0.948***	1.036	**0.947***	**0.993**	**0.958***	**0.976**	**0.936***
Success ratios										
1	**0.558***	**0.550***	**0.510***	**0.513***	**0.506***	**0.510***	**0.554***	**0.557***	**0.554***	**0.547***
3	**0.583***	**0.604***	0.498	**0.507***	0.486	0.500	**0.607***	**0.617***	**0.591***	**0.607***
6	**0.545**	**0.569***	0.410	0.444	0.410	0.451	**0.557***	**0.566***	**0.545**	**0.559***
9	**0.548**	**0.541**	0.398	0.469	0.402	0.473	**0.535**	**0.538**	**0.564***	**0.575***
12	**0.660***	**0.651***	**0.538***	**0.585***	**0.538***	**0.585***	**0.664***	**0.671***	**0.630***	**0.644***
15	**0.621***	**0.601***	**0.511***	**0.552***	**0.511***	**0.556***	**0.617***	**0.626***	**0.609***	**0.619***
18	**0.578***	**0.583***	0.483	**0.537***	0.496	**0.548***	**0.573***	**0.601***	**0.578***	**0.608***
21	**0.576***	**0.586***	0.485	**0.532***	0.476	**0.529***	**0.568***	**0.593***	**0.603***	**0.629***
24	**0.544***	**0.556***	0.456	**0.513**	0.460	**0.520**	**0.531**	**0.556***	**0.544***	**0.578***
Real Brent price										
Recursive MSPE ratios										
1	**0.918***	**0.925***	1.141	1.107	1.157	1.125	**0.966**	**0.946***	**0.959**	**0.942***
3	**0.937***	**0.945***	1.015	**0.983**	1.019	**0.987**	**0.949***	**0.941***	**0.941***	**0.932***
6	**0.992**	**0.991**	1.032	**0.994**	1.033	**0.994**	**0.992**	**0.976***	**0.987**	**0.969**
9	**0.985**	**0.980***	1.017	**0.969***	1.020	**0.970***	**0.980**	**0.957***	**0.970**	**0.945***
12	**0.946***	**0.948***	**0.986**	**0.932***	**0.989**	**0.932***	**0.949***	**0.924***	**0.928***	**0.903***
15	**0.939***	**0.942***	**0.981**	**0.920***	**0.986**	**0.920***	**0.947***	**0.919***	**0.932***	**0.902***
18	**0.983**	**0.957***	1.029	**0.934***	1.034	**0.934***	**0.992**	**0.934***	**0.979**	**0.918***
21	1.015	**0.980***	1.066	**0.949***	1.070	**0.947***	1.025	**0.951***	1.010	**0.933***
24	1.006	**0.972***	1.058	**0.938***	1.063	**0.937***	1.022	**0.943***	1.006	**0.923***
Success ratios										
1	**0.518**	**0.537**	0.486	0.483	0.478	0.477	**0.506**	**0.513**	**0.518**	**0.520***
3	**0.575***	**0.581***	0.409	0.446	0.417	0.453	**0.518**	**0.537***	**0.506**	**0.530***
6	**0.529**	**0.529**	0.377	0.434	0.369	0.427	**0.504**	**0.539***	**0.520**	**0.549***
9	**0.544**	**0.562***	0.386	0.445	0.390	0.445	0.477	**0.534***	**0.519***	**0.558***
12	**0.605***	**0.599***	0.496	**0.554***	0.483	**0.543***	**0.584***	**0.606***	**0.597***	**0.616***
15	**0.609***	**0.612***	**0.511***	**0.573***	0.494	**0.563***	**0.587***	**0.605***	**0.591***	**0.615***
18	**0.586***	**0.601***	0.448	**0.537***	0.444	**0.534***	**0.534***	**0.587***	**0.534***	**0.583***
21	**0.563***	**0.568***	0.419	0.500	0.428	**0.514***	**0.515***	**0.561***	**0.528***	**0.575***
24	**0.527**	**0.542***	0.425	0.498	0.407	0.487	0.491	**0.552***	0.487	**0.552***

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported for significance test. I also report the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy. $*$ denotes significance at least at the 10% level.

Accordingly, I summarise the forecast combinations' recursive MSPE and success ratios in periods 1992:01–2012:09 and 1992:01–2016:12 under different revision assumptions in Table 2.9 for RAC, WTI and Brent measures, respectively. I evaluate these forecasts based on the ex-post observations in vintages 2013:03 and 2017:06, which mirrors that employed in the paper, and find smaller recursive MSPE ratios and larger success ratios of the end-quarter and mid-quarter revisions than of the recursive revisions at the majority of forecast horizons; for instance, the greatest reduction of recursive MSPE relative to no-change is 11% at the 1-month horizon, revised for every end quarter for WTI and evaluated within the sample 1992:01–2012:09. While, the highest success ratio is 67.1% at the 12-month horizon, revised every mid quarter for RAC based on the sample 1990:01-2016:12. Hence, the quarterly uniform revisions of different real-time series promote increased forecasting accuracy.

(II) Alternative nowcast assumptions

A nowcast method for delayed observations under each vintage is proposed in Baumeister & Kilian (2012), which is also utilised in Baumeister & Kilian (2015). However, it is of interest to consider the forecast performance under alternative assumptions on nowcasts. In this subsection, I generally find: (1) nowcasts will dramatically improve the forecast accuracy relative to using real-time data set without nowcasts; and (2) if the nowcasts are only based on the most recent 12 observations, the forecast accuracy, based on recursive MSPE and success ratios, improved by around 3% on average.

In Table 2.10, I provide recursive MSPE and success ratios, relative to no-change forecasts, of equal-weight forecast combinations for WTI, RAC and Brent measures based on different nowcasts for the 1992:01–2012:09 and 1992:01–2016:12 evaluation periods respectively. An oil futures spread model is not included in Baumeister & Kilian (2015) at horizons beyond 18 months as there are missing values of the WTI futures data with maturities 18 to 24 months, denoted as 'BK (2015)' in Table 2.10. Moreover, futures prices for the Brent oil measure have missing values during the evaluation periods considered in this chapter with maturities from 9 to 24 months. Hence, I strictly follow Baumeister & Kilian (2015), and exclude the forecasts from oil futures spreads in the equal-weight combination at these horizons. Columns under 'No-nowcast' in Table 2.10, combines forecasts of each specification without nowcasts, and excludes the oil futures spread model at

horizons 18-24 months for WTI and RAC (9-24 months for the Brent measure). For example, oil market VAR includes the variables RAC, the world crude oil production, OECD crude oil stocks, and the world real economic activity index proposed by Kilian & Park (2009), which become available with different lags, see Table 2.2. I iteratively nowcast missing values, then forecast out-of-sample for horizons 1 to 24 months. In this chapter, I proposed a backcast method for these missing values based on Wiki crude oil futures continuous contracts from CL1 to Cl24 (available at https://www.quandl.com/data/CHRIS-Wiki-Continuous-Futures), see Appendix A.3.2. Thereafter, I am able to add the oil futures spread model at all forecast horizons 1 to 24 months for 'No-nowcast' and 'BK (2015)', presented under columns 'No-nowcast + Futures' and 'BK (2015) + Futures' respectively. Moreover, the nowcasts proposed in Baumeister & Kilian (2012) are based on all historical observations available at the real time point, for instance, they extrapolate the world oil production data based on the average rate of change in world oil production up to that point in time. However, the oil market data set presents highly frequent regime changes. Therefore, I fixed the rolling nowcasts' window size as 12, 24 and 60 months, which are listed under the last six columns. As an example, I nowcast the world oil production data based on the average rate of change in the last 12, 24 and 60 observations.

As shown in Table 2.10, I confirm that the nowcast method have positive contributions on real oil prices forecasts as mentioned in Baumeister & Kilian (2012), and especially at short forecast horizons comparing No-nowcast and BK (2015). Moreover, I found additional positive contributions at horizons 18 to 24 months for WTI, RAC, and at 9-24 months for Brent by adding forecasts of the oil futures spread model at these horizons. For example, 'BK (2015) +Futures' forecast combinations, under evaluation period 1992:09-2016:12, are significantly improved no-change forecasts at all forecast horizons from 1 to 24 months for all three measures. Moreover, although fixing the rolling-nowcast window size does not add significant improvement relative to nowcasts proposed in Baumeister & Kilian (2012), I found around an average of 3% more deductions on recursive MSPE and improvements on success ratios with window length 12 months, comparing columns under the '12' rolling nowcast window size, relative to 'BK (2015) +Futures'. More specifically, the highest deductions of recursive MSPE ratios are observed at the 1-month horizon for WTI, and the highest success ratio at the horizon 12 months for RAC, under the '12' rolling nowcast window size.

Table 2.10: Real-time recursive MSPE and success ratios of forecast combinations with equal weights under different nowcast assumptions (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12)

Nowcast assumptions	No-Nowcast		BK (2015)		No-Nowcast + Futures		BK (2015)+ Futures		Rolling nowcasts based on windows of length: 12		24		60	
MH	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12
Real WTI price														
Recursive MSPE ratios														
1	3.001	2.927	**0.903***	**0.905***	3.001	2.927	**0.903***	**0.905***	**0.897***	**0.900***	**0.901***	**0.903***	**0.902***	**0.904***
3	1.429	1.399	**0.922***	**0.920***	1.429	1.399	**0.922***	**0.920***	**0.918***	**0.917***	**0.920***	**0.919***	**0.921***	**0.919***
6	1.073	1.086	**0.986**	**0.976***	1.073	1.086	**0.986**	**0.976***	**0.982**	**0.972***	**0.984**	**0.974***	**0.985**	**0.975***
9	**0.987**	**0.997**	**0.979**	**0.969***	**0.987**	**0.997**	**0.979**	**0.969***	**0.977**	**0.968***	**0.979**	**0.970***	**0.979**	**0.969***
12	**0.949**	**0.976**	**0.945***	**0.940***	**0.949**	**0.976**	**0.945***	**0.940***	**0.943***	**0.940***	**0.945***	**0.941***	**0.945***	**0.940***
15	**0.907***	**0.944***	**0.940***	**0.936***	**0.907***	**0.944***	**0.940***	**0.936***	**0.938***	**0.935***	**0.940***	**0.936***	**0.940***	**0.936***
18	**0.928***	**0.932***	**0.972***	**0.943***	**0.928***	**0.932***	**0.972***	**0.943***	**0.970***	**0.943***	**0.972***	**0.944***	**0.972***	**0.944***
21	1.006	1.019	1.039	1.029	**0.987**	**0.972**	1.004	**0.969***	1.003	**0.969***	1.004	**0.970***	1.005	**0.969***
24	**0.994**	1.006	1.002	1.012	**0.985**	**0.958**	**0.980**	**0.952***	**0.980**	**0.953***	**0.980**	**0.953***	**0.980**	**0.953***
Success ratios														
1	0.486	0.477	**0.558***	**0.567***	0.486	0.477	**0.558***	**0.567***	**0.550**	**0.563***	**0.558***	**0.570***	**0.558***	**0.570***
3	0.478	0.493	**0.543**	**0.574***	0.478	0.493	**0.543**	**0.574***	**0.547**	**0.577***	**0.547**	**0.577***	**0.543**	**0.574***
6	0.492	**0.512**	**0.537**	**0.556***	0.492	**0.512**	**0.537**	**0.556***	**0.533**	**0.553***	**0.537**	**0.556***	**0.537**	**0.556***
9	**0.510**	**0.527**	**0.556***	**0.575***	**0.510**	**0.527**	**0.556***	**0.575***	**0.564***	**0.582***	**0.560***	**0.579***	**0.556***	**0.575***
12	**0.521**	**0.526**	**0.597***	**0.616***	**0.521**	**0.526**	**0.597***	**0.616***	**0.601***	**0.619***	**0.592***	**0.612***	**0.597***	**0.616***
15	**0.570***	**0.580***	**0.579***	**0.591***	**0.570***	**0.580***	**0.579***	**0.591***	**0.579***	**0.594***	**0.579***	**0.591***	**0.583***	**0.594***
18	**0.565***	**0.562***	**0.560***	**0.562***	**0.565***	**0.562***	**0.560***	**0.562***	**0.578***	**0.580***	**0.560***	**0.562***	**0.565***	**0.565***
21	**0.581***	**0.564***	**0.511**	**0.518**	**0.563***	**0.568***	**0.585***	**0.607***	**0.581***	**0.604***	**0.585***	**0.607***	**0.585***	**0.607***
24	**0.535**	**0.531**	**0.540**	**0.513**	**0.540**	**0.560***	**0.531**	**0.545***	**0.535**	**0.549***	**0.535**	**0.549***	**0.531**	**0.545***
Real U.S. refiners' acquisition cost for oil imports														
Recursive MSPE ratios														
1	4.007	4.034	**0.928***	**0.931***	4.007	4.034	**0.928***	**0.931***	**0.926***	**0.930***	**0.928***	**0.931***	**0.927***	**0.930***
3	1.731	1.710	**0.921***	**0.921***	1.731	1.710	**0.921***	**0.921***	**0.919***	**0.919***	**0.920***	**0.920***	**0.920***	**0.920***
6	1.136	1.171	**0.983**	**0.975***	1.136	1.171	**0.983**	**0.975***	**0.979**	**0.972***	**0.981**	**0.973***	**0.982**	**0.974***
9	1.042	1.053	**0.976***	**0.968***	1.042	1.053	**0.976***	**0.968***	**0.975***	**0.967***	**0.976***	**0.968***	**0.976***	**0.968***
12	**0.991**	1.031	**0.937***	**0.935***	**0.991**	1.031	**0.937***	**0.935***	**0.936***	**0.936***	**0.938***	**0.936***	**0.937***	**0.935***
15	**0.915***	**0.973**	**0.930***	**0.931***	**0.915***	**0.973**	**0.930***	**0.931***	**0.928***	**0.931***	**0.929***	**0.931***	**0.930***	**0.931***
18	**0.958**	**0.966**	**0.973***	**0.943***	**0.958**	**0.966**	**0.973***	**0.943***	**0.972***	**0.942***	**0.973***	**0.943***	**0.973***	**0.943***
21	1.055	1.060	1.044	1.033	1.038	1.015	1.006	**0.970***	1.005	**0.970***	1.006	**0.970***	1.007	**0.970***
24	1.068	1.058	1.015	1.030	1.059	1.008	**0.988**	**0.965***	**0.987**	**0.966***	**0.988**	**0.966***	**0.988**	**0.966***
Success ratios														
1	0.442	0.447	**0.558***	**0.550***	0.442	0.447	**0.558***	**0.550***	**0.566***	**0.560***	**0.550**	**0.543**	**0.558***	**0.550***
3	0.470	0.497	**0.583***	**0.604***	0.470	0.497	**0.583***	**0.604***	**0.579***	**0.604***	**0.587***	**0.607***	**0.583***	**0.607***
6	0.500	**0.525**	**0.545**	**0.569***	0.500	**0.525**	**0.545**	**0.569***	**0.541**	**0.563***	**0.545**	**0.566***	**0.545**	**0.569***
9	0.498	**0.531**	**0.548**	**0.541**	0.498	**0.531**	**0.548**	**0.541**	**0.548**	**0.541**	**0.548**	**0.541**	**0.544**	**0.538**
12	**0.534**	**0.540***	**0.660***	**0.651***	**0.534**	**0.540***	**0.660***	**0.651***	**0.660***	**0.651***	**0.660***	**0.651***	**0.655***	**0.647***
15	**0.545***	**0.563***	**0.621***	**0.601***	**0.545***	**0.563***	**0.621***	**0.601***	**0.621***	**0.605***	**0.613***	**0.598***	**0.617***	**0.601***
18	**0.569***	**0.569***	**0.578***	**0.583***	**0.569***	**0.569***	**0.578***	**0.583***	**0.591***	**0.594***	**0.582***	**0.587***	**0.582***	**0.587***
21	**0.541**	**0.546***	**0.507**	0.489	**0.520**	**0.543***	**0.576***	**0.586***	**0.568***	**0.579***	**0.568***	**0.579***	**0.581***	**0.589***
24	**0.553**	**0.549***	**0.549**	**0.520**	**0.549***	**0.563***	**0.544***	**0.556***	**0.549***	**0.560***	**0.553***	**0.563***	**0.549***	**0.560***
Real Brent price														
Recursive MSPE ratios														
1	3.026	2.938	**0.918***	**0.925***	3.026	2.938	**0.918***	**0.925***	**0.912***	**0.921***	**0.915***	**0.923***	**0.916***	**0.924***
3	1.448	1.442	**0.937***	**0.945***	1.448	1.442	**0.937***	**0.945***	**0.934***	**0.942***	**0.935***	**0.944***	**0.936***	**0.944***
6	1.110	1.131	**0.992**	**0.991**	1.110	1.131	**0.992**	**0.991**	**0.988**	**0.987**	**0.989**	**0.989**	**0.991**	**0.990**
9	1.003	1.022	**0.985**	**0.980***	1.003	1.022	**0.985**	**0.980***	**0.982**	**0.979***	**0.984**	**0.981***	**0.984**	**0.980***
12	**0.976**	1.021	**0.968***	**0.973***	**0.971**	1.010	**0.946***	**0.948***	**0.945***	**0.948***	**0.946***	**0.949***	**0.946***	**0.949***
15	**0.954**	**0.999**	**0.969***	**0.979**	**0.940***	**0.975**	**0.939***	**0.942***	**0.938***	**0.942***	**0.939***	**0.942***	**0.939***	**0.942***
18	**0.988**	1.001	1.018	1.002	**0.967**	**0.968**	**0.983**	**0.957***	**0.981**	**0.956***	**0.982**	**0.957***	**0.983**	**0.957***
21	1.040	1.047	1.051	1.036	1.017	1.002	1.015	**0.980***	1.014	**0.980***	1.015	**0.980***	1.015	**0.980***
24	1.041	1.050	1.034	1.033	1.024	**0.999**	1.006	**0.972***	1.006	**0.972***	1.006	**0.972***	1.007	**0.972***
Success ratios														
1	**0.514**	**0.510**	**0.518**	**0.537**	**0.514**	**0.510**	**0.518**	**0.537**	**0.526**	**0.543**	**0.526**	**0.543**	**0.522**	**0.540**
3	0.470	0.483	**0.575***	**0.581***	0.470	0.483	**0.575***	**0.581***	**0.571***	**0.581***	**0.575***	**0.581***	**0.575***	**0.581***
6	0.475	0.485	**0.529**	**0.529**	0.475	0.485	**0.529**	**0.529**	**0.541**	**0.539**	**0.545**	**0.542**	**0.529**	**0.529**
9	**0.502**	**0.531**	**0.544**	**0.562***	**0.502**	**0.531**	**0.544**	**0.562***	**0.556***	**0.572***	**0.548**	**0.565***	**0.544**	**0.562***
12	0.487	0.491	**0.567**	**0.564***	**0.504**	**0.516**	**0.605***	**0.599***	**0.622***	**0.612***	**0.609***	**0.599***	**0.609***	**0.602***
15	**0.532**	**0.517**	**0.591***	**0.566***	**0.511**	**0.521**	**0.609***	**0.612***	**0.609***	**0.612***	**0.609***	**0.612***	**0.613***	**0.615***
18	**0.552***	**0.544***	**0.547**	**0.544***	**0.543***	**0.558***	**0.586***	**0.601***	**0.591***	**0.604***	**0.591***	**0.601***	**0.591***	**0.604***
21	**0.572***	**0.554***	**0.528**	**0.518**	**0.555***	**0.554***	**0.563***	**0.568***	**0.568***	**0.571***	**0.563***	**0.568***	**0.563***	**0.568***
24	**0.558**	**0.542**	**0.522**	**0.505**	**0.535***	**0.534**	**0.527**	**0.542***	**0.531**	**0.545***	**0.527**	**0.542***	**0.522**	**0.538***

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported for significance test. I also report the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy. $*$ denotes significance at least at the 10% level.

(III) Historical sensitivity analysis

In practice, it is also of interest to check whether all six models are systematically necessary in Baumeister & Kilian's (2015) combination when extending the evaluation period. Baumeister & Kilian (2015) suggested to use a small set of models excluding 'gasoline spread' and 'no-change' specifications. From this subsection, I find that 'VAR', 'commodity prices', and the 'oil futures spread' models have become increasingly important within Baumeister & Kilian's (2015) six-models combination during the recursive extensions of the evaluation period from 1992:01–2012:09 to 1992:01–2016:12, which echoes their conclusion. However, 'TVP product spread' has tended to be less accurate than was the case prior to 2014. From the view of MSPE ratios, 'gasoline spread', 'TVP product spread' and 'no-change' specifications do not significantly contribute to a combination positively over a longer evaluation period, while, Baumeister & Kilian's (2015) small sample is still appropriate from the success ratios.

Table 2.11 presents the changes in real-time recursive MSPE and success ratios of six "leave-one-out" forecast equal-weight combinations at selected monthly horizons, for the 1992:01–2012:09 and 1992:01–2016:12 evaluation periods. Moreover, Figures 2.3 and 2.4 illustrate the changes of the leave-one-out forecast equal-weight combinations at the 1-, 12- and 24-month horizons as examples for the 1992:01–2008:03 period, recursively updated to the period 1992:01 to 2016:12.

Leave-one-out forecasts are combinations that eliminate one of Baumeister & Kilian's (2015) six equal-weight combination models, and hence the equal weight is $\frac{1}{5}$. For example, the changes in recursive MSPE and success ratios of a combination excluding the no-change forecast under the 'No change' column in Table 2.11 are calculated as the recursive MSPE and the success ratios of the leave-one-out forecast (omitting the no-change forecast), minus the corresponding ratios of Baumeister & Kilian's (2015) six-model combination at the selected forecast horizons within the recursively extended evaluation periods, respectively. This is consistent with Baumeister & Kilian's (2015) sensitivity analysis; however, in my study I consider both the changes in recursive MSPE and in the success ratios.

Since lower recursive MSPE and higher success ratios imply a more accurate forecast, the positive changes of MSPE and negative changes of success ratios (shown as boldface in Table 2.11) indicate that excluding the corresponding model

from Baumeister & Kilian's (2015) combination reduces accuracy results than their six-model combination. Therefore, the positive contribution via the changes due to the omitted model are positive for the recursive MSPE ratios and negative for the success ratios. A rough guide of a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) and Pesaran & Timmermann (2009) tests are calculated for recursive MSPE and the success ratios of Baumeister & Kilian's (2015) six-model forecast combination relative to the leave-one-out forecasts respectively, where an asterisk (*) in Table 2.11 denotes a minimum statistical significance at the 10% level.

I present the changes of recursive MSPE and success ratios for the three real crude oil price measures under the 1992:01–2012:09 and 1992:01–2016:12 evaluation periods, respectively (Table 2.11). Comparing the two evaluation periods, I generally confirm the consistently positive contributions of VAR, commodity prices and the oil futures spread model, at least at some forecast horizons. For example, excluding VAR, the changes of recursive MSPE ratios become positive at all forecast horizons for the WTI and RAC measures, and an increasing number of significant observations extend Baumeister & Kilian's (2015) evaluation period. Moreover, the magnitude of all the positive contributions of leaving-oil-futures-spread-out forecasts increases further under the extended evaluation period compared to the 1992:01 to 2012:09 evaluation period for all three measures. However, gasoline spread and no change forecast lose their power to contribute positively. More specifically, there is no positive change of recursive MSPE ratios committing gasoline spread or no change for WTI and RAC at all forecast horizons when extending Baumeister & Kilian's (2015) evaluation period to 1992:01–2016:12. Although significant negative changes of success ratios exist here for all three measures at some horizons, the magnitude is negligible. Additionally, TVP product spread shows decreasing predictability, despite the fact that it can still have a positive contribution within the combination at some horizons. For example, there are significantly positive changes of recursive MSPE ratios for WTI at the 6- to 24-month horizons under the 1992:01 to 2012:09 evaluation period, while under the 1992:01 to 2016:12 evaluation period only the change of recursive MSPE ratios at the 9-month horizon is significantly positive, albeit at a far smaller magnitude of change than under Baumeister & Kilian's (2015) evaluation period (0.009 relative to 0.028).

To present recursive dynamics of the information from Table 2.11 after 1992:01–2016:12 evaluation period, I integrate selected forecast horizons and re-

cursively extended evaluation periods from 1992:01–2008:03 to 1992:01–2016:12 in Figures 2.3 and 2.4. Changes in the real-time recursive MSPE (Figure 2.3) and success (Figure 2.4) ratios of the six leave-one-out forecast equal-weight combinations (in the columns) at horizons 1, 12, and 24 months.[2] As per Figure 2.2, the dotted, dashed, and solid lines represent the WTI, RAC and Brent measures, respectively, but here the benchmark is zero, again drawn as a slim dashed line.

A general observation comparing the ex-post relative to Baumeister & Kilian's (2015) evaluation period (the right relative to the left of the y axis in each sub plot) is that the volatility of changes in real-time recursive MSPE and success ratios of the leave-one-out forecast combinations with equal weights attenuates distinctly, as shown in Figures 2.3 and 2.4. For example, the recursive MSPE ratio changes of the leave-commodity-prices-out combination at the 12-month horizon fluctuates considerably from at 0.05 to at -0.03 during Baumeister & Kilian's (2015) evaluation period, while fluctuating slightly around zero within the interval between -0.01 and 0.01 after the 1992:01–2012:09 evaluation period.

In the first column of Figure 2.3, where the combination excludes VAR, positive changes of the recursive MSPE ratios for all three oil price measures are consistent at the 1-month horizon from 2012:09. Moreover, I can observe a trend in the changes of recursive MSPE ratios, which becomes positive with the extension of the evaluation periods at the 12- and 24-month horizons (Figure 2.3), despite the changes being negative within the sample (1992:01–2012:09). Although there are clear gains when including the VAR model, the changes in directional forecast accuracy among the three measures in Figure 2.4 exhibit divergence among the three oil price measures. The change in the success ratios for the Brent measure, excluding VAR, are positive at the 1-, 12- and 24-month horizons, but I can observe a downward trend of the changes with the extension of the evaluation periods. Hence, the VAR model does not contribute to increasing the directional forecast accuracy of the combination systemically for Brent measure. For WTI and RAC, in contrast, negative changes of success ratios at the 1- and 12-month horizons indicate the necessity of including VAR within Baumeister & Kilian's (2015) combination. Despite the changes of WTI's and RAC's success ratios being higher than zero at the horizon 24 months, clear downward pressure with the extension of the evaluation period also promotes the inclusion of VAR.

[2] As per Figure 2.2, the right of the y axis in each sub plot presents extended evaluation periods relative to Baumeister & Kilian (2015).

Including the commodity price model (column 2, Figures 2.3 and 2.4) improves the forecast accuracy at short forecast horizons. This is confirmed by both the recursive MSPE and success ratios, with consistent positive changes in recursive MSPE ratios at the 1-month horizon for all three measures; however, the inclusion of commodity prices consistently introduces losses, particularly at longer horizons. For example, the changes of the recursive MSPE and success ratios are negative and positive, respectively, but they converge to zero following the decrease of oil prices in 2014. These observations support Baumeister & Kilian's (2015) conclusion.

The contribution of the oil futures spread model within Baumeister & Kilian's (2015) forecast combination (column 3, Figures 2.3 and 2.4) is mixed at the 1-month horizon, but increases with time and is considerable at the 12- and 24-month horizons for all three measures. Since the fall of oil prices in 2014, the oil futures spread model has become increasingly important in Baumeister & Kilian's (2015) combination for the WTI and RAC measures at the 12- and 24-month horizons, while the importance for the Brent measure does not increase with time and even decreases.

The TVP product spread (column 5, Figures 2.3 and 2.4) gradually loses its predictive power in Baumeister & Kilian's (2015) forecast combination, and with a divergence among the three measures across horizons, when extending the evaluations in the post-2014 period. Excluding the TVP product spread for the Brent measure causes larger reductions at the 1-month horizon, but increases at 12 months in terms of the recursive MSPE ratios compared to the other two price measures. However, the gains of including the TVP model in the recursive MSPE ratios at the 12-month horizon are not reflected in the success ratios for Brent. The success ratios of Brent consistently increase at 12 months when excluding the TVP model. For WTI and RAC at the 1-month horizon, I also find consistent additional forecast accuracy through inclusion of the TVP product spread, although the magnitudes are more modest than for Brent. It should also be mentioned that gains contributed by the TVP product spread for the forecast combination at the 12- and 24-month horizons show signs of weakness for the 2014:09 to 2016:12 period.

Table 2.11: Historical changes in real-time recursive MSPE and success ratios of leave-one-out forecast combinations with equal weights (evaluation periods: 1992:01–2012:09 and 1992:01–2016:12)

Omitted model	VAR		Commodity prices		Oil futures spread		gasoline spread		TVP product spread		No change	
MH	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12	2012:09	2016:12
Real WTI price												
Changes in recursive MSPE ratios												
1	**0.028***	**0.034***	**0.040***	**0.036***	−0.007	−0.005	−0.015	−0.016	−0.015	−0.018	−0.014	−0.014
3	**0.017**	**0.027***	**0.039***	**0.035***	−0.006	−0.002	−0.011	−0.015	−0.011	−0.017	−0.012	−0.013
6	**0.004**	**0.019***	−0.016	−0.010	**0.010***	**0.014***	**0.004**	−0.003	**0.014***	−0.001	−0.001	−0.003
9	−0.010	**0.006**	−0.025	−0.017	**0.019***	**0.025***	**0.005***	−0.003	**0.028***	**0.009**	−0.002	−0.004
12	−0.010	**0.007**	−0.019	−0.010	**0.028***	**0.035***	**0.003**	−0.007	**0.024***	**0.003**	−0.008	−0.010
15	−0.007	**0.018***	−0.024	−0.015	**0.037***	**0.048***	**0.003**	−0.010	**0.023***	−0.007	−0.009	−0.010
18	−0.008	**0.025***	−0.027	−0.021	**0.037***	**0.053***	**0.003**	−0.013	**0.022***	−0.011	−0.003	−0.009
21	−0.003	**0.043***	−0.043	−0.027	**0.035***	**0.060***	**0.006***	−0.013	**0.031***	−0.028	**0.004**	−0.003
24	−0.009	**0.050***	−0.054	−0.021	**0.022***	**0.059***	**0.011***	−0.021	**0.065***	−0.022	−0.001	−0.006
Changes in success ratios												
1	**-0.008***	**-0.010***	**-0.052***	**-0.050***	**-0.024***	**-0.027***	**-0.032***	**-0.033***	**-0.012***	**-0.017***	0.000	0.000
3	0.032	0.013	**-0.004**	**-0.013***	0.032	0.020	0.016	0.007	0.016	**-0.003***	0.000	0.000
6	0.012	**-0.010***	**-0.033**	**-0.041***	0.000	**-0.010***	**-0.025**	**-0.014***	**-0.037**	**-0.027***	0.000	0.000
9	**-0.017***	**-0.024***	**-0.029***	**-0.041***	**-0.008***	**-0.003***	0.012	0.010	**-0.037***	**-0.021***	0.000	0.000
12	**-0.013***	**-0.031***	**-0.038***	**-0.045***	**-0.017***	**-0.028***	0.004	**-0.003***	**-0.008***	**-0.017***	0.000	0.000
15	**-0.030***	**-0.045***	0.013	0.014	**-0.004***	**-0.028***	0.004	0.003	0.000	**-0.003***	0.000	0.000
18	0.009	0.000	**-0.013***	0.018	**-0.022***	**-0.035***	0.022	0.028	0.000	0.014	0.000	0.000
21	0.004	**-0.021***	**-0.013***	**-0.021***	**-0.074***	**-0.089***	0.004	**-0.018***	**-0.061***	**-0.061***	0.000	0.000
24	0.058	0.014	0.066	0.043	0.009	**-0.032***	0.004	0.022	**-0.053**	**-0.004***	0.000	0.000
Real U.S. refiners' acquisition cost for oil imports												
Changes in recursive MSPE ratios												
1	**0.023***	**0.028***	**0.027***	**0.024***	−0.008	−0.007	−0.009	−0.010	−0.009	−0.012	−0.011	−0.010
3	**0.011**	**0.024***	**0.038***	**0.035***	−0.003	−0.000	−0.010	−0.015	−0.009	−0.016	−0.013	−0.013
6	−0.004	**0.015**	−0.006	−0.000	**0.012***	**0.015***	**0.003**	−0.005	**0.012***	−0.006	−0.001	−0.003
9	−0.020	**0.002**	−0.016	−0.007	**0.021***	**0.027***	**0.005***	−0.006	**0.028***	**0.004**	−0.002	−0.004
12	−0.022	**0.004**	−0.008	**0.004**	**0.029***	**0.037***	**0.003**	−0.012	**0.026***	−0.003	−0.010	−0.011
15	−0.024	**0.014**	−0.006	**0.004**	**0.036***	**0.048***	**0.002**	−0.017	**0.024***	−0.015	−0.011	−0.011
18	−0.026	**0.020***	−0.013	−0.005	**0.039***	**0.055***	**0.003**	−0.019	**0.024***	−0.019	−0.003	−0.009
21	−0.025	**0.037***	−0.035	−0.012	**0.038***	**0.063***	**0.007***	−0.019	**0.039***	−0.036	**0.004**	−0.003
24	−0.032	**0.047***	−0.049	−0.009	**0.027***	**0.065***	**0.013***	−0.026	**0.069***	−0.037	**0.000**	−0.003
Changes in success ratios												
1	**-0.028***	0.000	0.000	0.013	0.000	0.003	**-0.008***	**-0.003***	0.008	0.010	0.000	0.000
3	0.024	0.007	**-0.065***	**-0.057***	**-0.012***	**-0.020***	**-0.008***	**-0.013***	**-0.008***	**-0.017***	0.000	0.000
6	0.029	**-0.003***	**-0.016**	**-0.020***	**-0.004**	**-0.020***	**-0.025**	**-0.020***	**-0.049**	**-0.041***	0.000	0.000
9	**-0.012**	**-0.017**	**-0.037**	**-0.034**	0.000	0.007	**-0.004**	0.010	**-0.017**	0.031	0.000	0.000
12	**-0.042***	**-0.069***	**-0.021***	**-0.038***	**-0.042***	**-0.052***	**-0.004***	0.021	**-0.025***	0.007	0.000	0.000
15	**-0.017***	**-0.038***	**-0.017***	**-0.010***	**-0.026***	**-0.049***	0.000	0.035	**-0.043***	0.003	0.000	0.000
18	0.000	**-0.035***	**-0.009***	**-0.011***	**-0.030***	**-0.053***	0.017	0.032	**-0.013***	0.018	0.000	0.000
21	0.022	**-0.025***	**-0.013***	**-0.032***	**-0.070***	**-0.096***	0.000	0.011	**-0.057***	**-0.018***	0.000	0.000
24	0.040	**-0.007***	0.053	0.025	0.004	**-0.036***	0.013	0.036	**-0.093***	**-0.029***	0.000	0.000
Real Brent price												
Changes in recursive MSPE ratios												
1	**0.031***	**0.038***	**0.040***	**0.038***	−0.014	−0.014	−0.012	−0.014	−0.012	−0.017	−0.012	−0.011
3	**0.018**	**0.030***	**0.039***	**0.036***	−0.014	−0.013	−0.009	−0.013	−0.008	−0.015	−0.010	−0.008
6	−0.000	**0.018***	−0.009	−0.002	**0.004**	**0.004**	**0.005***	−0.002	**0.017***	−0.000	**0.000**	−0.000
9	−0.017	**0.004**	−0.015	−0.006	**0.016***	**0.018***	**0.005***	−0.004	**0.028***	**0.007**	−0.001	−0.002
12	−0.022	**0.006**	−0.001	**0.009**	**0.022***	**0.025***	**0.003**	−0.010	**0.025***	−0.001	−0.009	−0.008
15	−0.026	**0.013**	−0.001	**0.007**	**0.029***	**0.037***	**0.004**	−0.014	**0.024***	−0.010	−0.010	−0.010
18	−0.027	**0.018***	−0.012	−0.004	**0.036***	**0.045***	**0.004**	−0.017	**0.022***	−0.014	−0.001	−0.007
21	−0.025	**0.034***	−0.032	−0.011	**0.036***	**0.057***	**0.008***	−0.019	**0.032***	−0.030	**0.005***	−0.002
24	−0.029	**0.042***	−0.050	−0.012	**0.028***	**0.061***	**0.014***	−0.029	**0.058***	−0.026	**0.003**	−0.003
Changes in success ratios												
1	0.064	0.043	0.000	0.003	0.008	0.003	0.016	0.013	0.008	0.007	0.000	0.000
3	0.016	**-0.020***	**-0.028***	**-0.030***	**-0.004***	0.000	**-0.020***	**-0.020***	**-0.036***	**-0.030***	0.000	0.000
6	0.029	0.003	**-0.012**	**-0.010**	**-0.029**	**-0.034**	0.012	0.027	**-0.029**	**-0.003**	0.000	0.000
9	**-0.017**	**-0.045***	**-0.037**	**-0.041***	0.000	0.000	0.004	0.007	**-0.008**	0.010	0.000	0.000
12	0.004	**-0.035***	**-0.013***	**-0.017***	**-0.038***	**-0.035***	0.017	0.031	0.017	0.042	0.000	0.000
15	**-0.017***	**-0.070***	**-0.026***	**-0.045***	**-0.017***	**-0.045***	0.013	0.021	**-0.017***	0.000	0.000	0.000
18	0.017	**-0.035***	**-0.039***	**-0.032***	**-0.039***	**-0.057***	0.022	0.032	**-0.030***	0.000	0.000	0.000
21	0.026	**-0.021***	**-0.022***	**-0.014***	**-0.035***	**-0.050***	0.000	0.021	**-0.035***	0.007	0.000	0.000
24	0.066	0.011	0.022	0.000	**-0.004**	**-0.036***	0.000	0.029	**-0.080**	**-0.029***	0.000	0.000

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is reported for significance test. I also report the Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy. $*$ denotes significance at least at the 10% level.

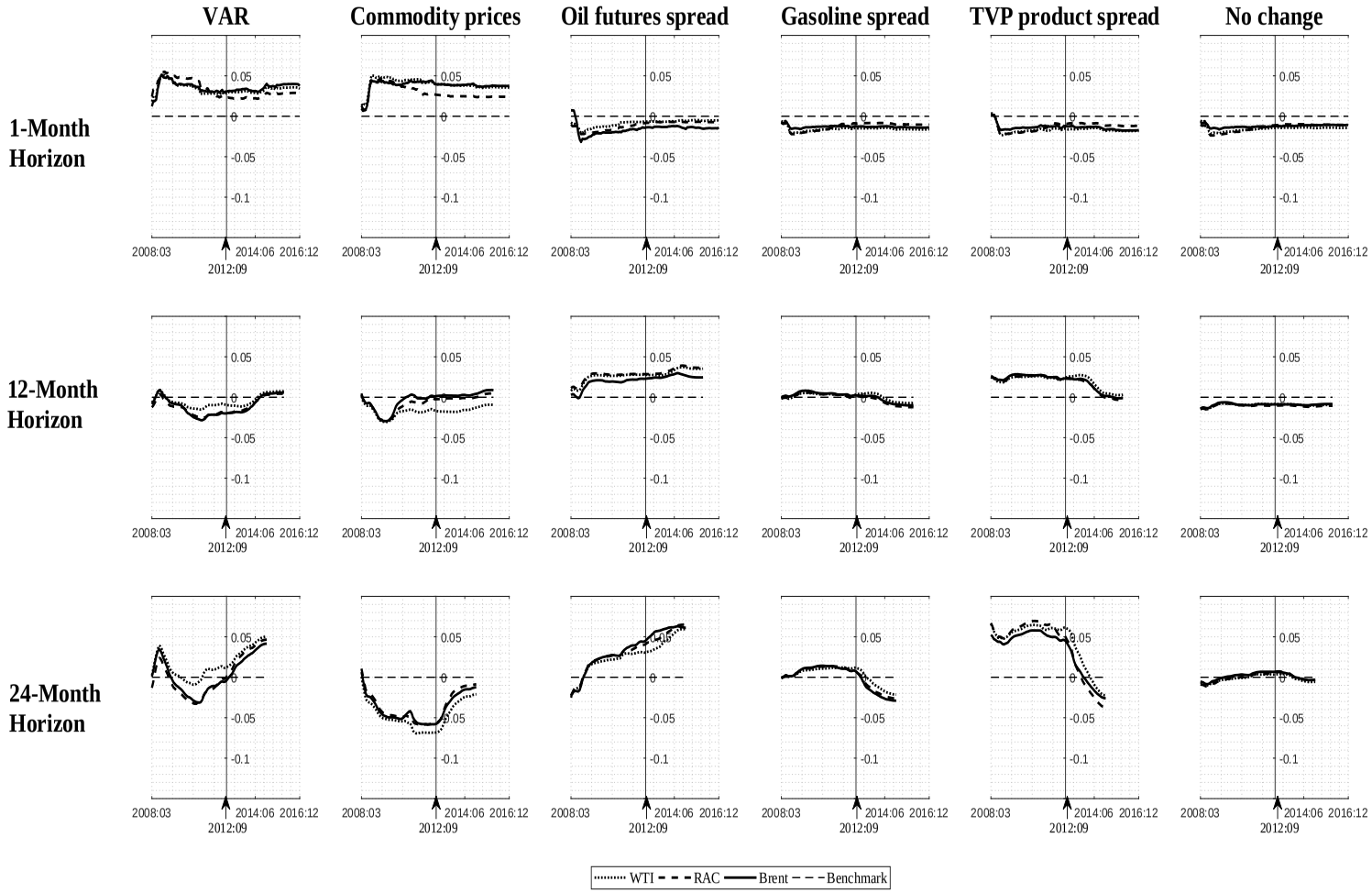

Figure 2.3: Historical changes in real-time recursive MSPE ratios of leave-one-out forecast combinations with equal weights

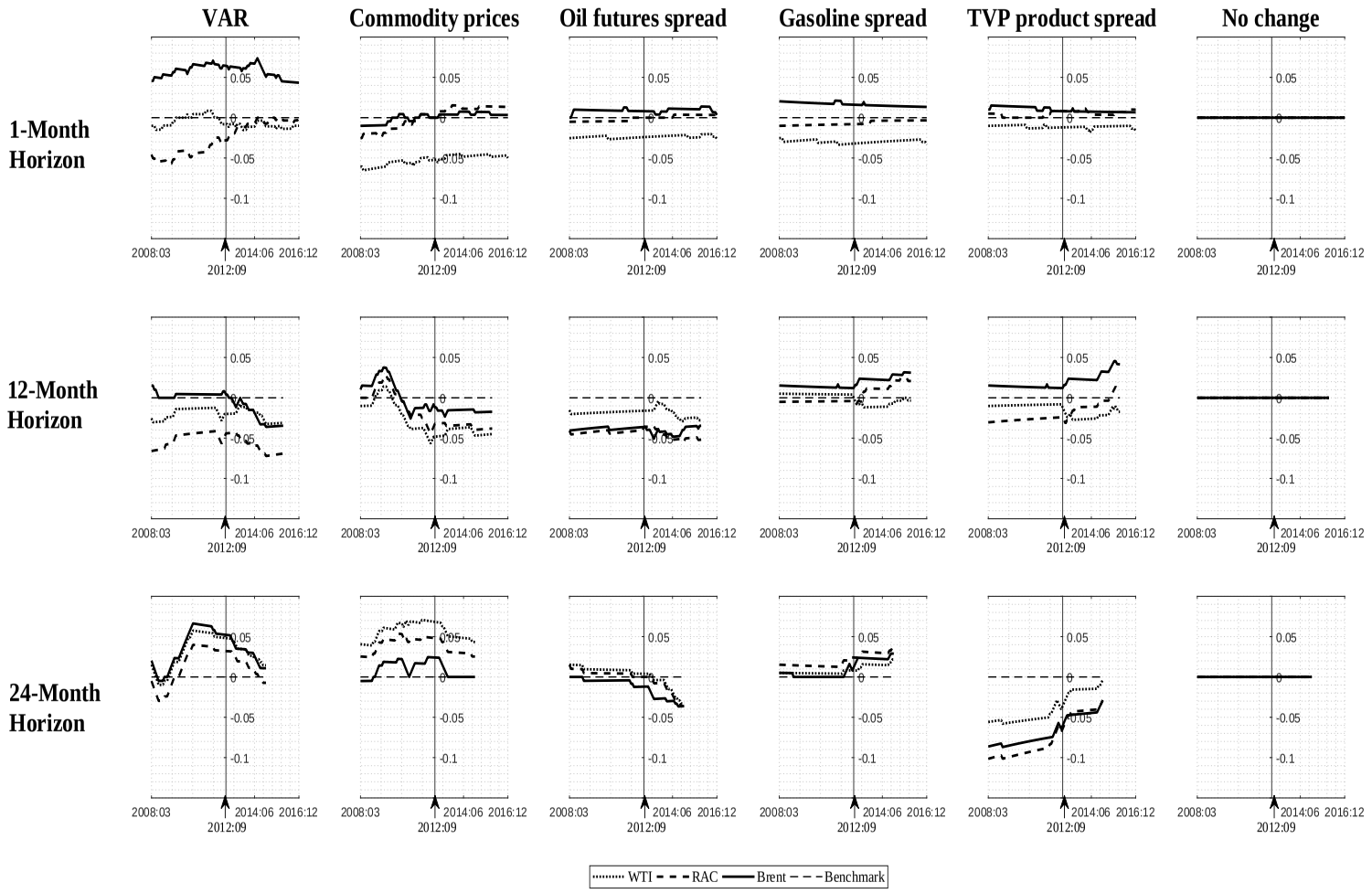

Figure 2.4: Historical changes in real-time success ratios of leave-one-out forecast combinations with equal weights

To summarise, Baumeister & Kilian (2015) exclude 'gasoline spread' (column 4) and 'no-change' (column 6) forecasts based on a sensitivity analysis, a decision confirmed by this exercise. Figures 2.3 and 2.4 show that the size of the differences in the recursive MSPE ratios is negligible across horizons, relative to the changes caused by excluding other specifications. These two models lose their ability to reduce the recursive MSPE ratios in Baumeister & Kilian's (2015) combination from the end of 2014. No-change forecasts do not contribute to forecast directional accuracy, while gasoline spread erodes the combination's directional forecast accuracy, excluding limited contributions at the horizon 12 months for WTI. Additionally, over the longer evaluation period 1992:01 to 2016:12, 'TVP product spread' gradually losses its predictive power for all three real crude oil price measures, especially from the view of recursive MSPE ratios. As a justification of the choice of models beyond Baumeister & Kilian's (2015) small set of combination, I would suggest excluding the 'TVP product spread' model, especially for forecasters who care more about MSPE rather than directional forecasts.

2.4.6 Quarterly forecasts of the sample period 1992Q1—2012Q3

In this subsection, I aggregate the forecasts of RAC, WTI and Brent measures from monthly to quarterly frequency. The method for proceeding is consistent with that introduced in Baumeister & Kilian (2015, p. 345). First, I construct equal-weight forecast combinations of the forecasts from specifications predicted in each vintage at the monthly horizons of 1 to 24 months (in similarity to the results generated in previous sections). Then, these are aggregated to quarterly forecasts; for example, the forecast at the horizon 1 quarter is the average of the monthly forecasts at 1, 2 and 3 months, which is estimated in the vintages of March, June, September, and December.

The data source for the EIA's oil price forecasts is their *Short-Term Economic Outlook*.[3] The publication provides quarterly forecasts of the U.S. refiners' acquisition cost for imports at horizons up to 7 quarters, while publications since 2000:10 and 2012:07 also include the quarterly forecasts of WTI and Brent crude oil prices, respectively. Given the irregular pattern of the reports, however, consistent time series of quarterly forecasts dating back to 1991 can only be obtained for hori-

[3] Available at https://www.eia.gov/outlooks/steo/outlook.php

zons from 1 to 4 quarters.[4]

For the 1991 to 1996 period, the *Short-Term Energy Outlook* is issued every quarter, while from 1997 onwards the publication is released on a monthly basis, but only reports quarterly forecasts.[5] The monthly publications are typically issued within the first 2 weeks of each month. The EIA updates its quarterly forecasts in each monthly report, incorporating new information as it becomes available.

To make the forecast comparison meaningful it is important to match the information set of the EIA as closely as possible with the information set in this study. Following Baumeister & Kilian (2015), I also distinguish two timing conventions for EIA forecasts. Timing convention I utilise the end-of-quarter issues of the *Short-Term Energy Outlook* (i.e., March, June, September and December). The oil price reported for the current quarter is taken as the nowcast and the oil prices reported for the subsequent quarters represent the forecasts. The corresponding real oil price forecasts are obtained by adjusting the nominal EIA oil price forecasts for expected inflation. The process is embodied in the following forecasting model of quarterly EIA forecasts $\hat{R}^{oil,Q}_{t+h|t}$ at horizon h on quarter t:

$$\hat{R}^{oil,Q}_{t+h|t} = R^{oil,Q}_{t}(1 + P^{h,EIA}_{t} - s_t - E_t(\pi^{h}_{t+h})),$$

where $R^{oil,Q}_{t}$ is the real quarterly observation of crude oil measures (deflated by the U.S. CPI), $P^{h,EIA}_{t}$ is the log of the current EIA forecasts at horizon h quarter(s), s_t is the corresponding log spot oil price, and $E_t(\pi^{h}_{t+h})$ is the expected quarterly U.S. inflation, as the average U.S. CPI inflation available at time t, where the averaging begins in 1986Q3, hence:

$$E_t(\pi^{h}_{t+h}) = [1 + \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}}^{\overline{\tau}} (ln(CPI_{t+1}) - ln(CPI_t))]^h - 1,$$

where $t = [\underline{\tau}, ..., \overline{\tau}]$, $\underline{\tau}$=1986Q3 and $\overline{\tau}$ is the quarter of the final observation in a specific vintage. This method is consistent with the forecasts from the oil futures spread model, introduced in Baumeister & Kilian (2015, p. 348)

[4]This quarterly time series allows us to gauge more directly the accuracy of the EIA's judgemental forecasts and to compare their accuracy with that of alternative forecasts. Establishing this is also pertinent because of the possibility that the EIA may have early access to oil market data, allowing it to generate more accurate real-time forecasts than econometricians.

[5]Monthly forecasts have been available in the EIA's *Short-Term Economic Outlook* since August 2004 for RAC and WTI, and since July 2012 for Brent.

Timing convention II instead relies on the issue of the *Short-Term Energy Outlook* that appears in the first month of the following quarter (i.e., April, July, October and January). Under this convention, the price reported for the previous quarter (e.g., the Q1 price in the April issue) is considered as the nowcast, and the price quoted for the current quarter (e.g., Q2 in the April issue) is the one-quarter-ahead forecast.

Baumeister & Kilian (2015) compared quarterly forecast combinations' performance relative to forecasts of RAC measure from *Short-Term Energy Outlook*. I also summarise the quarterly forecast accuracy with a similar evaluation period 1992Q1–2012Q3 in Table 2.12. Baumeister & Kilian's (2015) six-model combination (BK6), four-model combination (BK4), and the EIA's forecast for quarterly real RAC measure are listed in the first three columns. BK4 exclude no-change and gasoline spread forecasts, guided by sensitivity analysis in Section 2.4.5 and consistent with Baumeister & Kilian (2015). Equal weights are utilised here. As illustrated in Baumeister & Kilian (2015), I also found that the forecasts of EIA timing convention II are far more accurate than of timing convention I. Hence, I just report the timing convention II forecasts, which I call it 'EIA forecasts', allowing the EIA to offer an informational advantage of up to one month.

Quarterly point forecasts of RAC from Baumeister & Kilian's (2015) combination (especially BK4) dominate EIA forecasts, and are significant at 10% level at forecast horizons 4 and 5 quarters based on a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test. EIA forecasts can improve the sign forecasts at the 1-quarter horizon to a significant level of 5%, based on the Pesaran & Timmermann (2009) test and have a higher success ratio relative to Baumeister & Kilian's (2015) combinations. However, it is clear that Baumeister & Kilian's (2015) combinations can provide more accurate directional forecasts relative to EIA forecasts at horizons beyond 2 quarters, and at the most horizons are significant at 5% level. Moreover, BK4 for RAC is generally more accurate than BK6. These observations are consistent with the results in Baumeister & Kilian (2015). Additionally, I provide recursive MSPE and success ratios of BK6 and BK4 for WTI and Brent oil price measures in the rest columns of Table 2.12, respectively. And the results are qualitatively comparable with the forecasts of the RAC measure.

Table 2.12: Real-time forecast accuracy of equal-weighted forecast combinations at quarterly horizons

	Real U.S. refiners' acquisition cost for oil imports			Real WTI price		Real Brent price	
QH	Six models	Four models	EIA forecasts	Six models	Four models	Six models	Four models
			Recursive MSPE ratios				
1	**0.923**	**0.895**	**0.998**	**0.915**	**0.881**	**0.957**	**0.957**
2	**0.973**	**0.968**	1.134	**0.976**	**0.972**	**0.991**	1.004
3	**0.974**	**0.975**	1.060	**0.980**	**0.982**	**0.984**	**0.994**
4	**0.935****	**0.920***	**0.927**	**0.943***	**0.932***	**0.946***	**0.938**
5	**0.925****	**0.909****	$--$	**0.939****	**0.930***	**0.939****	**0.932***
6	**0.960**	**0.955**	$--$	**0.968**	**0.966**	**0.974**	**0.976**
7	1.009	1.020	$--$	1.013	1.025	1.014	1.025
8	1.012	1.031	$--$	1.017	1.037	1.017	1.036
			Success ratios				
1	**0.566**	**0.578**	**0.639****	**0.542**	**0.554**	**0.663****	**0.627****
2	**0.549**	**0.537**	**0.598**	**0.524**	**0.512**	0.500	**0.524**
3	**0.642****	**0.617***	**0.593**	**0.617***	**0.630****	**0.531**	**0.519**
4	**0.688****	**0.688****	**0.600**	**0.650****	**0.662****	**0.625****	**0.625****
5	**0.684****	**0.684****	$--$	**0.658****	**0.658****	**0.608****	**0.646****
6	**0.628****	**0.654****	$--$	**0.615****	**0.641****	**0.603****	**0.615****
7	**0.597****	**0.584****	$--$	**0.558***	**0.597****	**0.558****	**0.571****
8	**0.579**	**0.618****	$--$	**0.553**	**0.605****	0.500	0.487

Note: QH represents quarterly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is used for the significance test of recursive MSPE ratios. The Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level.

Moreover, I illustrate real-time recursive MSPE and success ratios relative to the no-change forecasts for the RAC measure in Figure 2.5, where the dashed, dotted, and solid lines represent the MSPE and success ratios for the forecasts from the EIA, BK4 and BK6, respectively, while the slim dashed line is the benchmarks 1 and 0.5 for the MSPE and success ratios, respectively.[6] As per Figure 2.2, an MSPE ratio below 1 and a success ratio above 0.5 indicate an improvement relative to the no-change forecast. Meanwhile, the plots present the evolution of the MSPE and success ratios over time for the evaluation period since 2012Q3.

A general observation from Figure 2.5 is that the MSPE and success ratios are stable over the period since 2012Q3. The sign forecasts of BK4 and BK6 evaluated by the success ratios are lower than the forecasts from the EIA at the 1- and 2-quarter horizons. However, the point forecasts from BK4 and BK6 dominate the forecasts from the EIA, evaluated through recursive MSPE ratios. The EIA

[6] I found the quarterly forecasting results for WTI and Brent measures show the same pattern with RAC. Hence, similar figures as Figure 2.5 for WTI and Brent measures are not reported in this thesis.

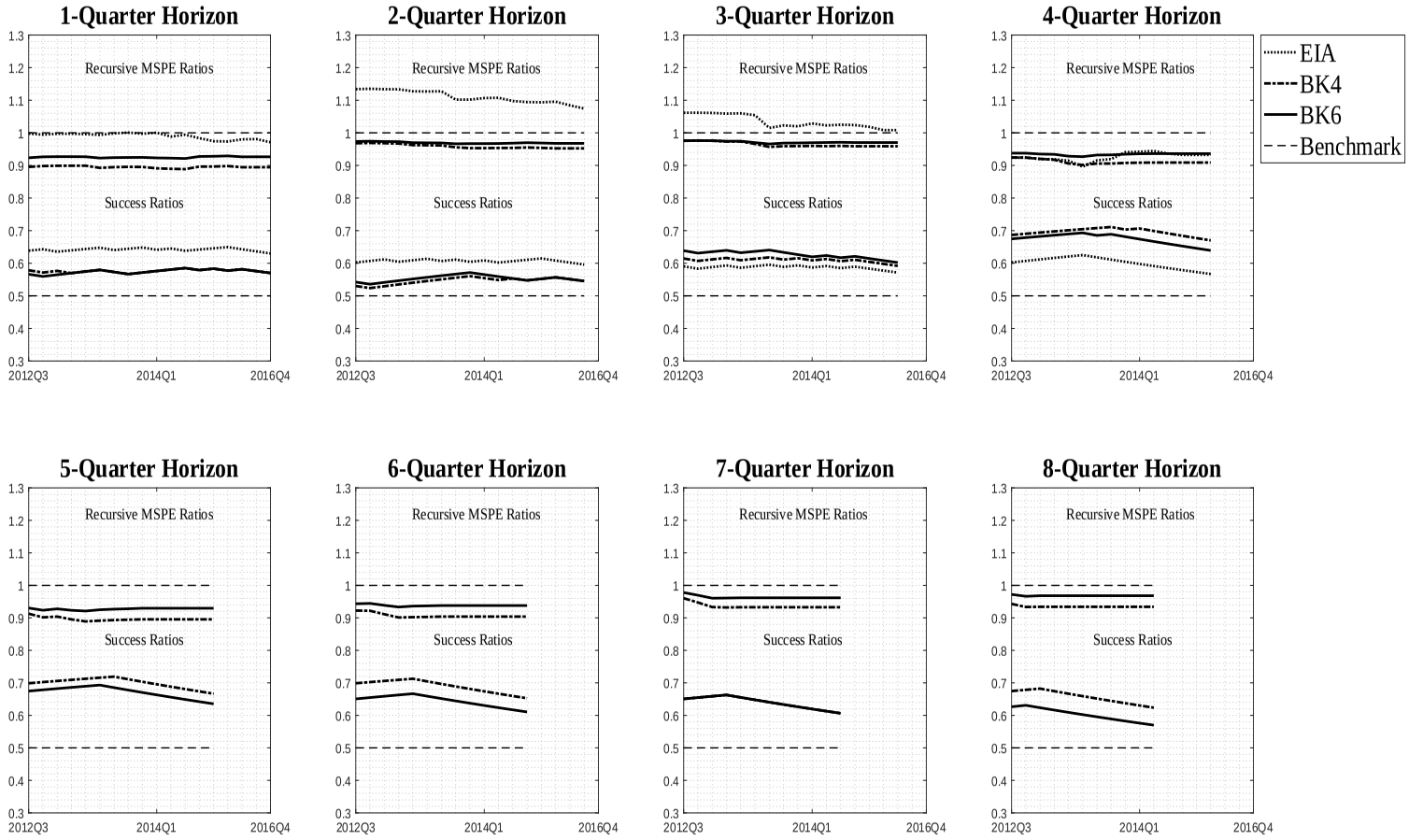

Figure 2.5: Quarterly historical recursive forecast accuracy relative to no-change forecast for refiner's acquisition cost for oil imports: EIA forecast, Equal weighted combination of all six forecasting models excepting a gasoline spread based model and no-change forecast, and Equal weighted combination of all six forecasting models

forecasts at the 2-quarter horizon present with lower accuracy than the no-change forecasts, as the recursive MSPE ratios are consistently above 1. At the 3- and 4-quarter horizons, the sign forecasts of the EIA become less accurate than BK4 and BK6. Moreover, for forecasting beyond the 4-quarter horizon, BK4 and BK6 become increasingly accurate point forecasts, but with a slightly decreasing sign-forecast accuracy following the plunge of oil prices in 2014. The forecasts from BK4 and BK6 at the 7- and 8-quarter horizons show a more accurate point forecast relative to the no-change forecast, whose recursive MSPE ratios are lower than 1, when extending the evaluation period to 2016Q4.

2.5 Conclusions

I extended Baumeister & Kilian's (2015) evaluation period to the end of 2016 using the Brent oil price measure in this chapter. The forecast combination is robust, and the predictive power is qualitatively comparable with the WTI and RAC oil price measures. Utilising alternative revision and nowcast assumptions could improve Baumeister & Kilian's (2015) equal-weight combination's forecast accuracy by around 2% on average across different forecast horizons over the period 1992:01–2016:12, respectively. From a historical sensitive analysis, a small sample set of combination including 'VAR', 'commodity prices', and the 'oil futures spread' was recommended. The following chapters consider the density forecasts but with a new model space, as well as a structural analysis of the real crude oil price.

Chapter 3

Density Forecasts of Real Crude Oil Prices Using Macro Founded Dynamic Models and Extended Evaluations

Abstract

This chapter uses a Bayesian vector autoregression (BVAR) considering time-varying parameters (TVP) and stochastic volatility (SV) modelling time variation in forecasting real crude oil prices. Two features are specific to this study. First, I minimise the one-step-ahead prediction Kullback–Leibler 'distance' in the sample and then eliminate the out-of-sample extreme predictions, allowing the highly parametric model to limit shrinkage, particularly in the long-term prediction range. Second, I extend the predictive assessment from standard statistical measures of the point and density forecasts to those that have a profitable opportunity in the futures market, as well as those that best predict the probability of real crude oil prices being extreme high or low. Over the 1992:01–2016:12 period, I find strong evidence supporting models using SV for real crude oil price density forecasts as opposed to using a conventional VAR. A constant parameter VAR with SV can provide well calibrated density forecasts and the highest probabilities for positive excess returns using crude oil futures. TVP may contribute to density forecasts, but only when incorporating SV and the majority of TVP are restricted to be time-invariant via a stochastic model specification selection prior and a linear opinion pool combining the 1- to 12-month VAR lag length choices.

3.1 Introduction

Crude oil prices such as the West Texas Intermediate (WTI) boomed to approximately $133 per barrel (p/b) in June 2008, and busted to approximately $39 per barrel (p/b) in just six months later. This extreme event has had profound implications for both oil importing and exporting countries (Alquist et al., 2013). A reliable forecast of real oil prices is being increasingly explored in the literature to support economic participants, central banks and international organisations in their monitoring of the change in oil prices. In the forecasting crude oil price literature, researchers concentrate on point, volatility and density forecasts in standard statistical terms (e.g. Ye et al., 2005; Knetsch, 2007; Yu et al., 2008; Kang et al., 2009; Mohammadi & Su, 2010; Alquist & Kilian, 2010; Baumeister & Kilian, 2012; Alquist et al., 2013; Baumeister & Kilian, 2015; Wang et al., 2015; Wang et al., 2017; and Baumeister et al., 2018).

In Chapter 3 I allow for time-varying parameters and stochastic volatility, to account for smooth structural change when considering crude oil prices, implemented via Bayesian methods (applied to vector autoregressions (VARs)). This is motivated by the limited model space often utilised in the oil price forecasting literature. It is common for constant parameter multivariate models to be utilised, for example see the oil market VAR in Baumeister & Kilian (2012, 2015), while others employ single-explanatory-variable time-varying parameters models (see Wang et al., 2015; 2017). Both approaches show the predictability for real crude oil prices. However, the former neglects the time variation of parameters in real crude oil prices forecasting, while the latter does not allow for interactions among variables.[1]

More specifically, price is based on demand and supply, which explains why the model space commences with a VAR typically including the following variables: the world crude oil production, the world real economic activity index, the real crude oil price, and crude oil inventories, as for example in the work of the structure VAR process, which is consistent with Baumeister & Kilian (2012, 2014, 2015). Bayesian VARs are widely utilised for modelling and forecasting in the macroeconomic literature, taking account of both the stochastic and parameter uncertainty associated

[1]This is also the reason why I concentrate on the oil market VAR incorporating time-varying parameters and stochastic volatility in this exercise, rather than the model space used in the previous chapter. The equal-weight combination utilised in Chapter 2 instead used to generate density forecasts as one of the benchmark models in this exercise.

with the model, and highlighting its priority for forecasting densities (for a more detailed discussion see Harrison & Stevens, 1976; Koop et al., 2007; and Koop et al., 2010). Nevertheless, constant parameters are unable to detect sources of frequent structural change in the crude oil price modelling, including non-linearities or time variation in the lag structure of the crude oil market model. On the other hand, stochastic volatility provides strong evidence for the out-performance in the energy price and commodity price analysis literature (see Duffie et al., 1999; Trolle & Schwartz, 2009; Vo, 2009; Larsson & Nossman, 2011; Jebabli et al., 2014; Wang et al., 2015; and Wang et al., 2017). However, these models limit the frequency of parameters' change, or ignore the economic foundations. To track time variation in multivariate linear structures, time-varying parameters and stochastic volatility are generally considered in the macroeconomic literature. The utilisation of multivariate stochastic volatility is intended to account for the potential heteroscedasticity of the structural shocks and non-linearities of the simultaneous relations among the variables of interests.

The inclusion of time-varying parameters and stochastic volatility is vulnerable to the over-parametrisation problem. Researchers propose priors achieving shrinkage in a time-varying parameter model framework, including the least absolute shrinkage and selection operator (LASSO), stochastic model specification search (SMSS), and Bayesian variable stochastic selection priors in order to restrict the in-sample estimates on shrinkage (see Frühwirth-Schnatter & Wagner, 2010; Korobilis, 2013; Belmonte et al., 2014; Belmonte et al., 2014; and Eisenstat et al., 2016). Although the shrinkage method can restrict in-sample estimations, the random-walk innovation of time-varying coefficients in out-of-sample periods would be too volatile to make accurate forecasts. Nonetheless, to my knowledge, there is no method in the forecasting literature to restrict out-of-sample forecasts on shrinkage.

When incorporating time-varying parameters and stochastic volatility, allowing parameters varying in the out-of-sample forecasting horizons is limited, which potentially circumvents extreme volatility in forecasts (see a discussion in Korobilis, 2013). However, allowing for the innovation of parameters in the out-of-sample is of interest, because the parameters at the end of sample, estimated using time-varying parameters and information at time T, are temporary and driven by specific timing factors. For example, Korobilis (2013) proposes an alternative method that ignores the uncertainty of parameters' time variation in the out-of-sample forecast period, but two problems arise when applying the method to oil price forecasting: (1) the

forecasts are unable to measure the structural uncertainty during the forecast period; and (2) the coefficients at the end of a sample are not reliable, such as the sample end with the financial crisis in 2008 where the crude oil price surged and crashed within 6 months.

Therefore, in this chapter I propose an econometric method applied to out-of-sample forecasts that minimises the one-step-ahead in-sample forecasts' Kullback-Leibler 'distance', and then, eliminates extreme forecasts, defined as a value that is more than three scaled median absolute deviations away from the median, in the out-of-sample density forecasts. (I name the method as MKLD.) This exercise utilises MKLD and allows for the out-of-sample innovation of time-varying parameters and stochastic volatility.

The second contribution of Chapter 3 extends the evaluation of model comparisons from a standard statistical evaluation of point and density forecasts to one that selects specifications that are more valuable in terms of (1) forecasting excess returns when trading in the crude oil futures market, and (2) forecasting the likelihood of an extreme high or low real crude oil price. Evidence in the stock return forecasting literature suggests that good statistical criteria does not necessarily contribute to a profitable portfolio allocation (Leitch & Tanner, 1991; Pettenuzzo & Ravazzolo, 2016). This broadens the set of evaluation methods deployed in the existing oil price forecasting literature. It is of interest to evaluate the point and density forecasts using higher frequency financial data, as crude oil futures are traded in large volumes in commodity markets. This chapter designs a realistic *long–short* trading strategy utilising daily crude oil futures and real-time data, the aim of which is assessing whether forecasts are able to generate high probabilities for positive excess returns. Following Garratt et al. (2016), I utilise the hit rate, Kuipers Score, symmetric (asymmetric) fair bet, as well as Brier Score and its decomposition for checking the models' forecasting accuracy of extreme high and low real crude oil prices.

The sample used in this exercise is the same as in the previous chapter, and the out-of-sample real crude oil price forecasting are evaluated in the 1992:01–2016:12 period. I find that the SMSS proposed in Eisenstat et al. (2016) and a linear opinion pool (LOP) combining time-varying-parameter VARs with stochastic volatility for 1–12 VAR lag lengths can efficiently reduce the time variation of time-varying parameters in in-sample estimation. As for the point forecast, the new

generation using the time-varying MKLD model is dominated by equal-weight combination using Baumeister & Kilian's (2015) model space (BK). However, the use of MKLD, a constant parameter VAR with stochastic volatility, and a model using both SMSS and LOP, can provide well calibrated density forecasts and dominate a density combination using BK. (Note uses density forecasts from BK's combinations, whereas the previous Chapter (and BK) considered point forecasts.) The density forecasts from BK's combination use simple Monte Carlo methods, at all forecasting horizons from 1 to 24 months.

Moreover, a constant parameter VAR with stochastic volatility using MKLD can provide higher probabilities of positive excess returns in oil futures market over the 1992:01–2016:12 period when taking a realistic *long–short* strategy, at all risk-aversion levels, than all other specifications considered in oil price forecasting literature including BK and models with time varying parameters. While, the narrow densities predicted by the traditional VAR using ordinary least squares (OLS) are unresponsive to different levels of risk aversion assumed in the evaluation.

Although, the density forecasts, based on a model shrinking the time-varying parameters via both SMSS and LOP, are slightly better for the extreme low price, the constant parameter VAR with stochastic volatility using MKLD provides one of the highest two hit rates, Kuipers Scores, and fair bet rates for forecasting excessively high real crude prices, as well as one of the lowest two Brier Scores at all the forecast horizons from 1 to 24 months. Hence, it seems that the evidence points to the gain from stochastic volatility rather than the time-varying parameters, which is also observed in Chan & Eisenstat (2018) for forecasting the U.S. macroeconomic data.

The layout of the remainder of this chapter is as follows. Section 3.2 will introduce the methodology, including a description of the estimation, model specifications, and forecasting process. Section 3.3 will detail the sources of the data used in the application. In Section 3.4, in-sample estimated statistics will be presented, followed by first a standard statistical evaluation of the point and density forecasts, and second by the extended evaluation scheme, which considers excess returns and extreme event analysis. Finally, Section 3.6 will conclude the chapter.

3.2 Models

The following subsections explore the Bayesian estimation, the specifications illustrated in this chapter and the process of MKLD.

3.2.1 Bayesian estimation of SMSS for VAR with time-varying parameters and stochastic volatility

Starting with the general formation of a time-varying parameter oil market VAR with stochastic volatility, this subsection will detail the reparameterisation of Eisenstat et al. (2016). The focus of this subsection is Eisenstat et al.'s (2016) decomposition for a hybrid model, which allows some of parameters to be time-varying while some to be constant using SMSS priors.

A time varying parameter oil market VAR with stochastic volatility is formulated as:

$$B_{0,t}y_t = X_t\beta_t + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma_t); \tag{3.1}$$

$$\beta_t = \beta_{t-1} + \eta_t,\ \eta_t \sim N(0, \widetilde{\Omega}); \tag{3.2}$$

where $t = [1, \ldots, T]$, and T is the last observations of a sample. For the oil market, $y_t = [\Delta prod_t, rea_t, r_t^{oil}, \Delta inv_t]^{'}$, which is a 4×1 vector following Kilian & Murphy (2014); $prod_t$ is the percent change in the global crude oil production; rea_t is a business cycle index of world real economic activity; r_t^{oil} is the log of the U.S. refiners' acquisition cost for crude oil imports (RAC) deflated by the log of the U.S. Consumer Price Index (CPI); Δinv_t is the change in global crude oil inventories; X_t is defined as:

$$X_t^{'} = I_n \otimes [1, y_{t-1}^{'}, ..., y_{t-k}^{'}],$$

where k is the choice of VAR lag length; the first k observations of each sample (or vintage) are dropped for estimation (in other words, $t = 1$ means the $(k+1)^{th}$ sample observations are y_t); and the symbol $\otimes$ denotes the Kronecker product. $B_{0,t}$ is a lower uni-triangular matrix—i.e., a lower triangular matrix with unit diagonal elements. Σ_t is diagonal with diagonal elements $\sigma_{i,t}^2 = exp(h_{i,t})$, where in defining $h_t = (h_{1,t}, ..., h_{n,t})^{'}$ the log-volatilities follow a random walk:

$$h_t = h_{t-1} + \eta_t^h, \quad \eta_t^h \sim N(0, R). \tag{3.3}$$

I assume the random walk is initialised with $h_0 \sim N(0, V_0^h)$ and the prior on the transition covariance is $R \sim IW(\upsilon_0, R_0)$, where IW inverse-Wishart distribution.

Following Eisenstat et al. (2016), I reparameterise the general model above as:

$$y_t = Z_t\alpha + W_t\gamma_t + \varepsilon_t, \ \varepsilon_t \sim N(0, \Sigma_t); \tag{3.4}$$

$$\gamma_t = \gamma_{t-1} + \widetilde{\eta}_t, \ \widetilde{\eta}_t \sim N(0, I_m); \tag{3.5}$$

where $\widetilde{\eta}_t$ is independent of each other for all leads and lags. Z_t includes the contemporaneous values y_t. For example, Z_t is defined below for a four-variable unrestricted time-varying-parameter VAR, with Cholesky decomposition, as:

$$Z_t = \begin{bmatrix} [1, y_{lags}] & 0 & 0 & 0 \\ 0 & [-y_{1,t}, 1, y_{lags}] & 0 & 0 \\ 0 & 0 & [-y_{1,t}, -y_{2,t}, 1, y_{lags}] & 0 \\ 0 & 0 & 0 & [-y_{1,t}, -y_{2,t}, -y_{3,t}, 1, y_{lags}] \end{bmatrix},$$

where $y_{lags} = [y_{t-1}^{'}, ..., y_{t-k}^{'}]$. Therefore, Z_t is reconstructed as the X_t defined above, and m is the size of time varying parameters, $m = (k \times n + 1) \times n + \frac{n \times (n-1)}{2}$. This type of decomposition was used by Chen & Dunson (2003), Kinney & Dunson (2007), and Eisenstat et al. (2016) for the Cholesky decomposition to a random-effects setting.

Moreover, $W_t = Z_t\Omega^{\frac{1}{2}}\Phi$, where $\Omega^{\frac{1}{2}} = diag(\omega_1, ..., \omega_m)$ is the diagonal covariance matrix for the state transition in Equation (3.1). The Φ is lower uni-triangular and contains the correlations of the time-varying parameters. It is necessary to note that this exercise restricts the non-diagonal elements in Φ to zero in order to avoid a very high-dimensional variance-covariance matrix, which may dramatically increase the computational burden. (See a discussion in Eisenstat et al. 2016, p. 1648). The Tobit prior proposed in Eisenstat et al. (2016), referred to as SMSS, incorporates the essential elements in Frühwirth-Schnatter & Wagner (2010) and Belmonte et al. (2014). The SMSS is utilised for automatically restricting $\omega_j \geq 0$, while still allow-

ing for a straightforward implementation of hierarchical shrinkage, namely LASSO, through the hyper-parameters of ω_j^* sampled from a truncated normal distribution, as:

$$\omega_j = \begin{cases} 0 & \text{if } \omega_j^* \leq 0 \\ \omega_j^* & \text{if } \omega_j^* > 0. \end{cases} \tag{3.6}$$

Appendix B.1 provides detailed prior and the posterior computation for ω_j^* (for a more detailed discussion see Eisenstat et al., 2016). It is also necessary to note that the error covariance matrix $\widetilde{\Omega}$ for β_t in Equation (3.2) is a full matrix below

$$\widetilde{\Omega} = \Omega^{\frac{1}{2}} \Phi \Phi' \Omega^{\frac{1}{2}}.$$

Further, the α follows a standard independent prior, $\alpha \sim N(a_0, A_0^{-1})$, and the variance A is shrunk with the standard LASSO used in Markov chain Monte Carlo (MCMC), see Eisenstat et al. (2016) for further details. More precisely, the unrestricted error for the state transition Equation (3.1) is from $N(0, Q_t^{-1})$, where $Q_t = B_{0,t}' \Sigma_t^{-1} B_{0,t}$. It is clear that $\alpha + \gamma_t$ includes the free elements in $B_{0,t}$ and β_t in Equation (3.1).

To sum, the distributional assumptions as regards to (ε_t, $\widetilde{\eta}_t$, η_t^h) are stated below. The elements of the vector γ_t are modelled as random walks. The standard deviations are assumed to evolve as geometric random walks, defined as stochastic volatility. This constitutes an alternative to the autoregressive conditional heteroscedasticity (ARCH) model, with the difference being that the variances generated by Equation (3.3) are unobservable components. On the other hand, assuming Equations (3.3) and (3.5) hold for a finite period of time, the effect should be innocuous. Moreover, the random walk assumption presents the advantages of focusing on permanent shifts and reducing the number of parameters in the estimation procedure. Notice that, in principle, the model can be easily extended to consider more general autoregressive processes. Particularly for the parameters of the variance–covariance matrix, this exemplifies this model's advantage over the so-called local scale models (see for example Shephard (1996) and the multivariate generalisation of Uhlig (1997)). Furthermore, all innovations in the model are assumed to be jointly normally distributed as the variance–covariance matrix:

$$V = \begin{bmatrix} I_m & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & V_0^h \end{bmatrix}, \tag{3.7}$$

where I_m is an m-dimensional identity matrix. R and V_0^h are positive definite matrices, known as hyperparameters. It is worth noting that none of the restrictions on the structure of V are essential. All the zero blocks could be replaced by non-zero blocks, with only small modifications of the estimation procedure. The following hyper-parameters on the priors are standard as in Eisenstat et al. (2016):

$$\begin{aligned} a_0 &= 0, & A_0 &= I_m, \\ h_0 &= 0, & V_0^h &= In, \\ \nu_0 &= n + 11, & R_0 &= 0.01^2(\nu_0 - n - 1)In, \\ \lambda_{01} &= 0.1, & \lambda_{02} &= 0.1. \end{aligned}$$

The following subsection will detail the other specifications considered in this Chapter.

3.2.2 Model specifications

In the real crude oil price forecasting literature, an oil market based VAR typically uses a lag length of 12 months, without sign restrictions (Baumeister & Kilian, 2012, 2014, 2015). As one of the benchmarks, I will report the VAR estimated by OLS, which is unrestricted, constant parameter with no stochastic volatility, denoted as VAR. As mentioned in the introduction, this chapter aims to incorporating the time varying parameters and stochastic volatility into the real crude oil prices modelling. Mindful of the over-parametrization problem, I will report a Bayesian VAR with time varying parameters restricted by SMSS and stochastic volatility, which will be denoted as TVPSVsmss introduced in Subsection 3.2.1.

It is also interesting to compare the out-of-sample forecast performance among the following models: a Bayesian VAR with constant parameters using LASSO and inverse gamma prior on volatility (BVAR), a Bayesian VAR with constant parameters using LASSO and stochastic volatility (SV), a Bayesian VAR with time varying parameters based on LASSO prior and inverse gamma prior on volatility (TVP), a Bayesian VAR with time varying parameters restricted by SMSS and inverse gamma prior on volatility (TVPsmss), and a Bayesian VAR with time varying parameters restricted by LASSO prior and stochastic volatility (TVPSV). The technical details are in Appendix B.2.

Moreover, I use a linear opinion pool (LOP) combining TVPSVsmss according to the choice of VAR lag lengths, k from 1 to 12 months (TVPSVsmss(k), where $k = [1, \ldots, 12]$), denoted as CombineTVPSVsmss. (See a more detailed discussion of the LOP in Stone et al., 1961 and Wallis, 2005.) The advantages of using the LOP include: (1) the fact that forecasting accuracy can be improved through the combination has been shown in forecasting literature, e.g. Geweke & Amisano (2011), Pettenuzzo & Ravazzolo (2016), Geweke (2010), Gneiting & Ranjan (2011), Gneiting et al. (2013), Garratt, Henckel, & Vahey (2019), and among others; (2) LOP provides a pool that will not converge to one specific model, excepting in the case that one specification dominates others in out-of-sample Kullback-Leibler divergence; and (3) the LOP combination of TVPSVsmss(k) is a dynamic method providing additional shrinkage on coefficients of the predictors at long lag length, motivated by a Minnesota type prior. On the other hand, due to the fact that LOP contains the specifications' disagreements, the density combination always has a different shape with respect to the specifications' marginal density forecasts, which will possibly influence the combination's out-of-sample forecasting performance as well as modelling asymmetries in risk (Garratt, Henckel, & Vahey, 2019). Therefore, this chapter focuses on the marginal forecasting performance of economic founded models (VAR based) with time-varying parameter and stochastic volatility, where I use CombineTVPSVsmss only for a model comparison purpose.[2]

Formally, the aggregated real-oil-price density of CombineTVPSVsmss from the 12 specifications based on density scores using LOP is:

$$g(r^{oil}_{T+h|CombineTVPSVsmss}) = \sum_{k=1}^{12} \omega^{CRPS}_{k,h,t} g(r^{oil}_{T+h|TVPSVsmss(k)}), \tag{3.8}$$

where $g(r^{oil}_{T+h|TVPSVsmss(k)})$ denotes the density forecasts of TVPSVsmss with k VAR lag length at forecasting horizon h; $T = \underline{t}, ..., \bar{t}$, depending on the length of evaluation period; here, for example, the out-of-sample is from 1992:01 to 2016:12. There are two alternative weighting schemes — logarithmic score-based weights and continuous ranked probability score (CRPS)-based weights — in the literature for

[2]The use of LOP in this Chapter is motivated by the purpose for a well calibrated density forecast. Given the range of measures used to evaluate the forecasts, it would be of interest to also using these measures to construct weights (e.g. models which have had high financial market returns recently get more weight). For example, Leitch & Tanner (1991) and Pettenuzzo & Ravazzolo (2016) argue for that forecasts preferred by statistical criteria, is not necessarily useful for locating to a profitable portfolio. I leave this analysis for the future research, because this Chapter mainly contributes to quantify the use of time varying parameters and stochastic volatility but not for the discussion of various optimal weighting schemes.

a linear mixture of experts' framework, but they give standard conclusions. In this chapter, I will present the results based on CRPS weights, and this idea is come from Ravazzolo & Vahey (2014). As described in Baumeister & Kilian (2015), a reliable oil market data becomes available with a 6-month delay. Hence, the construction of weights at time T is based on the fit of a model in time $T-h-6$. Consequently, the gain from a combination is limited. I will only show the results of the combined density, for CombineTVPSVsmss.

The CRPS weights are defined as

$$\omega_{k,h,T}^{CRPS} = \frac{\sum_{\underline{t}+h+6}^{T-h-6} \Gamma(g(r_{T+h|TVPSVsmss(k)}^{oil}))}{\sum_{k=1}^{12}[\sum_{\underline{t}+h+6}^{T-h-6} \Gamma(g(r_{T+h|TVPSVsmss(k)}^{oil}))]}, \quad T = \underline{t}, ..., \bar{t},$$

where Γ is the inverse of the CRPS, which is defined as $CRPS(g(.), r_{T+h}^{oil}) = E_g\{|R - r_{T+h}^{oil}|\} - \frac{1}{2}E_g\{|R - R^*|\}$; r_{T+h}^{oil} is the observation of log real RAC in the month $T+h$; R and R^* are independent draws of a linear random variable with forecast distribution $g(.)$; and $E_g\{.\}$ denotes the expectation operator for the predictive $g(r_{T+h|TVPSVsmss(k)}^{oil})$. The initial $h+6$ months utilise the equal weights, due to the data constraint (see the discussion in Baumeister & Kilian, 2015). 'Sharpness', the centre of a forecast density, and 'distance', the distance between the centre of the forecast density and the out-turn, could be considered and concentrated when evaluating the forecast density (Gneiting & Raftery, 2007). The CRPS metric prefers densities with high sharpness and small distance. Hence, employing the CRPS enables us to generalise the absolute error, and to provide a direct path for comparing various deterministic and probabilistic forecasts according to a single metric. This is why the CRPS incentives predict densities with a high probability around the ex-post released observations (see Hersbach, 2000; Gel et al., 2004; Gneiting & Raftery, 2007; Panagiotelis & Smith, 2008; Groen et al., 2013; and Ravazzolo & Vahey, 2014). Two different methods for computing the CRPS are as follows: method 1 is based on closed form expectation using iterates, while Panagiotelis & Smith (2008) introduced the computational steps required for method 2, which utilises Monte Carlo draws from the component forecast density. Because I did not find significant difference between the two methods, I will only present the forecasts based on the method 2.

Lastly, I extend the specifications introduced in Baumeister & Kilian (2015), and detailed in Section 2.3, to density forecasts. Then, I use equal weights to combine the densities, denominated as BK (for the extension see Appendix B.3). All the

Table 3.1: A list of specifications used for this application, including their label, estimator, prior and uncertainty

Label	Specification	Estimator	Prior	Uncertainty
VAR	Vector Autoregression	OLS	−−	Stochastic uncertainty
BVAR	Bayesian Vector Autoregression	Gibbs sampler	LASSO for coefficients; Inverse gamma for variance	Stochastic and parameter uncertainty
SV	BVAR with stochastic volatility	Gibbs sampler	LASSO for coefficients	Stochastic and parameter uncertainty
TVP	BVAR with time varying parameter	Gibbs sampler	LASSO for coefficients; Inverse gamma for variance	Stochastic, parameter, and time varying uncertainty
TVPsmss	BVAR with time varying parameter	Gibbs sampler	LASSO for constant coefficients; SMSS for time varying parameters; Inverse gamma for variance	Stochastic, parameter, and time varying uncertainty
TVPSV	BVAR with time varying parameter and stochastic volatility	Gibbs sampler	LASSO for coefficients	Stochastic, parameter, and time varying uncertainty
TVPSVsmss	BVAR with time varying parameter and stochastic volatility	Gibbs sampler	LASSO for constant coefficients; SMSS for time varying parameters	Stochastic, parameter, and time varying uncertainty
Combine TVPSVsmss	BVAR with time varying parameter and stochastic volatility	Gibbs sampler; LOP	LASSO for constant coefficients; SMSS for time varying parameters	Stochastic, parameter, and time varying uncertainty
BK	Equal-weight combination	Multi methods	−−	Stochastic, parameter, and time varying uncertainty

specifications considered in this chapter are listed in Table 3.1, including the label and the specification itself, the estimator, the prior, and the uncertainty considered by the model.

Further, the size of the burn-in is 10,000 draws, with 1,000 draws of time-varying (or time invariant) posterior estimates stored to produce point and density forecasts, for all specifications. I choose the number of posterior draws to be 1000 for the following reasons. First, the results did not change with larger numbers of draws, for example I tried 10,000 draws for a VAR with time varying parameters, stochastic volatility, and SMSS, and found that the results are not statistically different from the results implied by 1,000 posterior draws. The diagnostics of convergence seem satisfactory given the large dimensionality of the specifications. In addition, in Appendix B.4 I illustrate the convergence through, Geweke's (1992) inefficiency factor (IF), the $Z-$score and Gelman & Rubin's (1992) potential scale reduction factor (PSRF), which are widely used in the MCMC convergence diagnosing literature. (See reviews in Cowles & Carlin (1996) and Roy (2019).) Second, this exercise undertakes recursive estimation through 1991:12–2016:12, using 300 real-time vintages, where the statistical evaluation for the in-sample estimates are based on $300 \times 1,000 = 300,000$ posterior draws for each specification, which is computationally prohibitive. (Appendix B.4 provides evidence for the specifications estimation consistency across vintages through the PSRF test.)

3.2.3 Out-of-sample forecast implementation

Here, two methods can achieve the out-of-sample forecasts at horizon h. The first one, proposed in Korobilis (2013), ignores the uncertainty of structural change in the forecast period at 1- to 24-month horizons. I reconstruct TVPSVsmss as the companion form of the standard VAR model for forecasts:

$$E(y_{T+h}) = \Sigma_{i=0}^{h-1}(B)^i(c) + (B)^h y_{lags|T}, \tag{3.9}$$

$$c = B_{0,T}^{-1} c_T,$$

$$B = \begin{bmatrix} B_{0,T}^{-1}B_{1,T} \cdots B_{0,T}^{-1}B_{k-1,T} & B_{0,T}^{-1}B_{k,T} \\ I_{n\times(k-1)} & 0_{n(k-1)\times n} \end{bmatrix},$$

where $y_{lags|T} = [y'_{T-1}, ..., y'_{T-k}]$; $B_{0,T}$, c_T, and B_T are component estimated draws from $\alpha + \Omega^{\frac{1}{2}}\Phi\gamma_T$ in Equations (3.4) and (3.5), plugging in the last known values of the coefficients at T in the sample for the nonlinear model (see discussions around the forecast method in D'Agostino et al., 2013; Korobilis, 2013; and Cogley & Sbordone, 2008). Within the 1- to 4-month horizons, the method described above is consistent over the evaluation period; however, when extending the out-of-sample forecast horizons, such as 12 months or longer, large forecast errors are inevitably introduced into the system.

As mentioned in the introduction of this chapter, the allowance of parameter variation in an out-of-sample is interesting. Hence, through the second method, MKLD, a model with drifting coefficients, predictive simulation can be implemented to forecast breaks in the coefficients' out-of-sample. This implies random walk evolution of the mean coefficients in a time-varying parameter VAR and simulation of future path using Monte Carlo iterations (see Bauwens et al. (2011) for the discussion).

In this subsection I introduce MKLD process by an example of TVPSVsmss. For a posterior draw ι, the forecast at horizon 1 month, $y_{T+1|\iota}$ is calculated through:

1. Simulate the time-varying parameters:

$$\gamma_{T+1|\iota} = \gamma_{T|\iota} + \widetilde{\eta}_{T+1},$$

   where $\widetilde{\eta}_{T+1} \sim N(0, I_m)$.

2. Innovate stochastic volatility:

$$h_{T+1|\iota} = h_{T|\iota} + \eta^h_{T+1}, \quad \eta^h_{T+1|\iota} \sim N(0, R|_\iota),$$

   where $R|_\iota$ is the posterior draw ι's transition covariance for the log-volatilities.

3. Construct $B_{0,T+1|\iota}$:

$$B_{0,T+1|\iota} = \alpha^{B_0}_\iota + \Omega^{\frac{1}{2},B_0}_\iota \gamma^{B_0}_{T+1|\iota},$$

   where $\alpha^{B_0}_\iota$, $\Omega^{\frac{1}{2},B_0}_\iota$ and $\gamma^{B_0}_{T+1|\iota}$ are the coefficients stored in α_ι, $\Omega^{\frac{1}{2}}_\iota$ and $\gamma_{T+1|\iota}$ for contemporary observations in the matrix Z_T respectively. $B_{0,T+1|\iota}$ is formed as a lower uni-triangular matrix.

4. Forecast $y_{T+1|\iota}$ with a specific posterior draw ι:

$$y_{T+1|\iota} = B_{0,T+1|\iota}^{-1}(\widetilde{Z}_T\alpha_\iota^{\setminus B_0} + \widetilde{Z}_T\Omega_\iota^{\frac{1}{2},\setminus B_0}\gamma_{T+1|\iota}^{\setminus B_0} + \varepsilon_{T+1|\iota}),$$

where $\varepsilon_{T+1|\iota} \sim N(0, \Sigma_{T+1|\iota})$, and $\Sigma_{T+1|\iota}$ is diagonal matrix with diagnal elements $exp(h_{T+1|\iota})$. The $\alpha_\iota^{\setminus B_0}$, $\Omega_\iota^{\frac{1}{2},\setminus B_0}$ and $\gamma_{T+1|\iota}^{\setminus B_0}$ are the coefficients stored in α_ι, $\Omega_\iota^{\frac{1}{2}}$ and $\gamma_{T+1|\iota}$ for lagged observations in matrix Z_T respectively.

The $\widetilde{Z}_T$ is Z_T without contemporary observations, where

$$\widetilde{Z}_T = \begin{bmatrix} [1, y_{lags|T}] & 0 & 0 & 0 \\ 0 & [1, y_{lags|T}] & 0 & 0 \\ 0 & 0 & [1, y_{lags|T}] & 0 \\ 0 & 0 & 0 & [1, y_{lags|T}] \end{bmatrix}.$$

Again, $y_{lags|T} = [y_{T-1}^{'}, ..., y_{T-k}^{'}]$.

The forecasts at a longer horizon h from 2 to 24 months for this application, $y_{T+h|\iota}$, incorporate (or conditional on) the forecasts at previous horizons from $y_{T+1|\iota}$ to $y_{T+h-1|\iota}$, and recursively go through the steps 1 to 4 mentioned above. Hence, the forecasts of logarithmic real RAC measure at horizon h with draw ι, denoted as $r_{T+h|\iota}^{oil}$, are included in $y_{T+h|\iota}$.

In the forecasting literature, one model's marginal performance (likelihood) is closely related to its out-of-sample forecasting performance, for example, see Koop & Korobilis (2013). Researchers utilise combination methods to minimise the logarithmic score or the CRPS to reduce the out-of-sample error (see for example Garratt et al., 2014; Ravazzolo & Vahey, 2014; Garratt & Mise, 2014). Here, I use average one-step ahead in-sample Gaussian analytic CRPS, $AvCRPS$, which is calculated as:

$$AvCRPS = \frac{1}{T}\sum_{t=1}^{T}(CRPS(\mathcal{N}(X_t\beta_t, e^{h_t}), B_{0,t}y_t)), \tag{3.10}$$

where $CRPS(\mathcal{N}(\mu, \sigma^2), x) = \sigma[2\phi(\frac{x-\mu}{\sigma}) - \frac{1}{\sqrt{\pi}} + \frac{x-\mu}{\sigma} \times (2\Phi(\frac{x-\mu}{\sigma}) - 1)]$ is defined in Grimit et al. (2006); β_t and $B_{0,t}$ can be uniformly calculated via α, γ_t, and $\Omega^{\frac{1}{2}}$.[3]

[3]The use of analytic Gaussian CRPS is motivated by both the efficiency and that Equation 3.1

Then, I combine 1,000 draws of α, $\Omega^{\frac{1}{2}}$, γ_t, h_t, and R for the out-of-sample forecasts by minimising $AvCRPS$ through a LOP:

$$p(r^{oil}_{T+h}) = \sum_{\iota=1}^{1000} \omega_{\iota}^{AvCRPS} g(r^{oil}_{T+h|\iota}), \tag{3.11}$$

$$\omega_{\iota}^{AvCRPS} = \frac{AvCRPS_{\iota}^{-1}}{\Sigma_{\iota=1}^{1000} AvCRPS_{\iota}^{-1}}, \tag{3.12}$$

where there are 1,000 independent draws of a linear random variable with forecast distribution $g(.)$ from Bayesian iterations. r^{oil}_{T+h} is the log-level forecast of the real RAC measure at horizon h from a specification. The $AvCRPS_{\iota}$ is calculated for all endogenous variables considered in the specification for each random draw ι of estimated coefficients.

It is necessary to note the reason why I do not use dynamic model selection (DMS) nor Bayesian model averaging (BMA). There are three concerns. First, DMS and BMA will finally converge to a true model (in the limit), for example, please see a review on BMA in Hoeting et al. (1999). However, Diebold (1991) pointed that all models could be misspecified given a long out-of-sample forecasting horizons, referred to as the model imcompleteness in Geweke (2010). Second, comparing DMS and BMA, LOP considers the parameters' uncertainty and will no longer converge to unique solution, except in the case where one specification dominates in terms of Kullback–Leibler divergence (Pettenuzzo & Ravazzolo, 2016). Third, this exercise uses the Gibbs sampler for estimation, and the variation of posterior draws reflects parameter uncertainty. This is different to Koop & Korobilis (2013), as their DMS is applied on hyper-parameters, which determine the prior shrinkage of parameters.

Further, the probability of extreme forecasts is raised with the increase of the number of coefficients to be estimated. This is because the probability of explosive out-of-sample random walk evolutions of the parameter estimates is higher for a large model than for a small one. For example, assuming the probability of explosive out-of-sample random walk evolutions for a one-parameter model is i.e. $1\%(= 1 - 99\%)$, then which for a 100-parameter model it is approximately i.e. $63\%(= 1 - 99\%^{100})$. Therefore, I eliminate the draws in the posterior, when produces extreme forecasts for any one of the out-of-sample forecast 1–24 month horizons. An

is locally Gaussian at t. I also did try logarithmic score, and did not find any different performance relative to CRPS. In this exercise I only present the results based on CRPS.

extreme forecast is defined as a value that is more than three scaled median absolute deviations away from the median. Appendix B.5 justifies that MKLD, accompanied with extreme-forecasts-elimination process, is equivalent to placing zero prior weight on a specific region of the parameter space. Under the zero-weight region, out-of-sample time-varying parameters' AR(1) coefficients have standard–error estimates that are around three times the size of the accepted estimates. I also illustrate that the process is closely follows the estimates of the specifications, therefore rules out the argument that 'the resulting forecasts are not from the model being estimated'.

The forecasts of WTI and Brent oil price measures are extrapolated from the forecast of RAC multiplied by the most recent spread between them and RAC, respectively, from each specification considered in this chapter, through:

$$R_{T+h|M}^{\text{WTI/Brent}} = exp(r_{T+h|M}^{oil}) \times spread_T^{\text{WTI/Brent, RAC}},$$

$$spread_T^{\text{WTI/Brent, RAC}} = \frac{p_T^{\text{WTI/Brent}}}{p_T^{RAC}},$$

where, $p_T^{\text{WTI/Brent}}$ and p_T^{RAC} are the nominal observations for WTI, Brent and RAC, respectively, in time T.

Here are two motivations to do so. First, the characteristics of the three measures over period 1992:01–2016:12 are illustrated in Figure 3.1. The co-movement of three measures is consistent over the evaluation period. Although, a large divergence among the three measures happened between 2011:01 and 2014:08, the spread among three crude oil prices are generally consistent over the divergent period. And the volatility of $spread_T^{\text{WTI/Brent, RAC}}$, is negligible relative to the volatility of crude oil prices over the evaluation period. Second, extrapolating the forecast of real WTI measure from the forecast of real RAC price is common in the real crude oil price forecast literature, see Baumeister & Kilian (2012) and Baumeister & Kilian (2015).

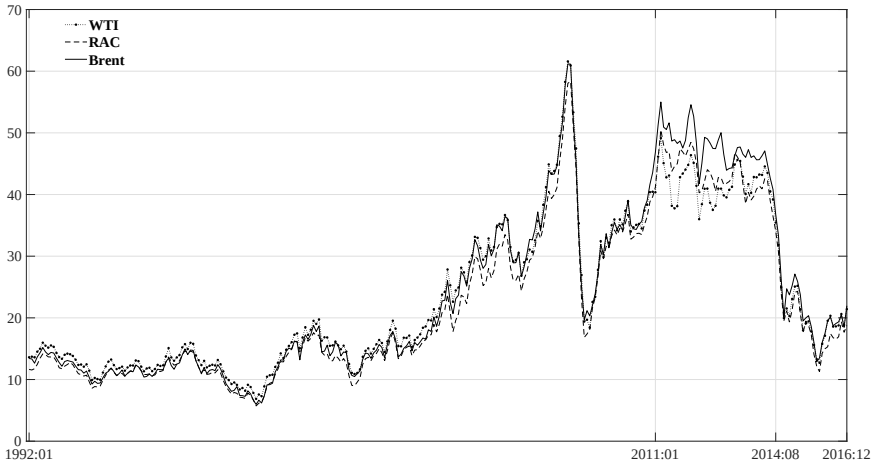

Figure 3.1: Monthly real crude oil prices (nominal prices divided by the U.S. CPI)

It is necessary to note that the MKLD method is utilised for the BVAR, SV, TVP, TVPsmss, TVPSV, and TVPSVsmss specifications, while the sub-models included in CombineTVPSVsmss also employ the method. The following section will detail the data, then I evaluate the specifications for the period 1992:01 to 2016:12.

3.3 The data

The data employed in this chapter include two different date frequencies for the forecasting purpose and the calculation of futures returns respectively. First, the data used for forecasting purpose in this chapter is the real time data set for crude oil market developed in Chapter 2. The real time monthly time series include the U.S. refiners' acquisition cost for crude oil imports, WTI spot prices and the New York Mercantile Exchange (NYMEX) WTI oil futures with maturities from 1 to 24 months, the price of Brent crude from the North Sea and its futures with maturities from 1 to 24 months, the global crude oil production, a business cycle index of global real activity proposed in Kilian (2009), the U.S. CPI, the U.S. crude oil inventories, the U.S. and the OECD countries petroleum inventories. The sources of the data include the U.S. Energy Information Administration, Bloomberg, Drewry Shipping Consultants Ltd., the Federal Reserve Economic Data (FRED) and the Federal Reserve Archival System for Economic Research (FRASER) database of the Federal

Reserve Bank of St. Louis. The details of the data are described in the previous chapter, and the advantages of a real-time data set are discussed in Baumeister & Kilian (2012), Kilian & Murphy (2014), and Baumeister & Kilian (2015).

Second, additional to the data employed for forecasting, Subsection 3.4.4 uses daily data for a *long–short* investment strategy to measure the economic performance using density forecasts. Daily closing spot prices of WTI and Brent oil during the 1991:12:01–2016:12:31 period are available respectively at https://www.eia.gov/dnav/pet/hist/RWTCD.htm and https://www.eia.gov/dnav/pet/hist/rbrteD.htm from the U.S. Energy Information Administration (EIA). WTI futures are traded on the NYMEX, and Brent futures on the Intercontinental Exchange (ICE) Futures Europe. Each crude oil futures contract represents 1,000 barrels, and the price is denominated as USD per barrel. The daily closing futures prices for WTI and Brent have maturities ranging from 1 to 24 months for the period 1991:12:01–2016:12:31 and are obtained from Bloomberg using the source keys CL1:COM to CL24:COM, and CO1:COM to CO24:COM respectively.[4]

Further, I collect the last trading date (LTD) of futures contracts from 1991:11 to 2019:01 using the tickers CLZ91 to CLF9 for WTI (source key: CL_EXS) and COZ91 to COF9 for Brent (source key: CO_EXS) from Bloomberg. The publishing dates for each *Monthly Energy Review* are also collected, and are available at https://www.eia.gov/totalenergy/data/monthly/previous.php#1973-92. This information has been taken into the consideration for a *long–short* investment strategy. For example, a 1-month-maturity futures contract's LTD is 2016:12:31, while the EIA just published its *Monthly Energy Review* at 2016:12:28. The density forecast at the 1-month horizon, based on this *Monthly Energy Review*, is not appropriate to guide a *long–short* investment strategy for the 1-month-maturity contract, which is going to mature in few days. Instead, this density forecast is used for deciding the *long–short* positions of a 2-month-maturity contract. Please see details in Appendix B.7.

[4]The daily data used in this exercise is only available before the 31^{st} December 2016, and the calculation of excess returns requires the last trading day price of a contract. Hence, the last contract ends at 2016:12:31. The calculation of excess return is described in Appendix B.7.

3.4 Forecast evaluation

In this subsection, I evaluate the specifications mentioned above based on point and density forecasts. However, there is considerable ambiguity trying to decide which density forecast is better among SV, TVP, TVPSV, TVPsmss, TVPSVsmss, and CombineTVPSVsmsss based on standard statistics. Further, standard statistical measures of point and density forecasts limit the ability to communicate with decision makers and market participants. Therefore, this chapter designs a realistic *long–short* trading strategy using daily crude oil futures and real-time data, aiming to answer if forecasts are able to provide a greater probability for abnormal returns. Finally, I extend the analysis to evaluate the probability forecasts of extreme high and low real crude oil prices.

3.4.1 In-sample diagnostic statistic

The world crude oil prices present highly frequent time variation, which provided the motivation for introducing time-varying parameters and stochastic volatility to measure time varying uncertainty. However, the inclusion of the TVP and SV introduces a non-informative structural uncertainty in the out-of-sample forecast period, which is why I use LASSO, SMSS, LOP and MKLD to restrict on the in-sample and out-of-sample shrinkage. Essentially, the majority of estimates of parameters are restricted to be time-invariant through SMSS, the inclusion of stochastic volatility and LOP. Further, MKLD eliminates approximately 20% of the draws from posteriors on average in out-of-sample forecasts.

Figure 3.2 plots the weights utilised for constructing CombineTVPSVsmss at the 1-, 3-, 6-, 12-, 18- and 24-month forecast horizons. I also highlight the highest and lowest weights at the end of the evaluation period per sub-figures. It is clear that the weights are very consistent over the evaluation period. At the 3-, 6- and 12-month horizons, the highest weight is assigned to the model with a lag order of one month but is slightly higher than the other VAR lag choices. Figure 3.3 shows the estimated time-invariance probabilities, which are calculated as the number of time invariant estimates divided by the size of the posterior, over 300 vintages' real-time estimations for the TVPsmss, TVPSVsmss, and CombineTVPSVsmss specifications, respectively. The sub-figures present the time invariant probabilities of all estimates. The left panel is for the estimates in $B_{0,t}$ that are lower unitriangular con-

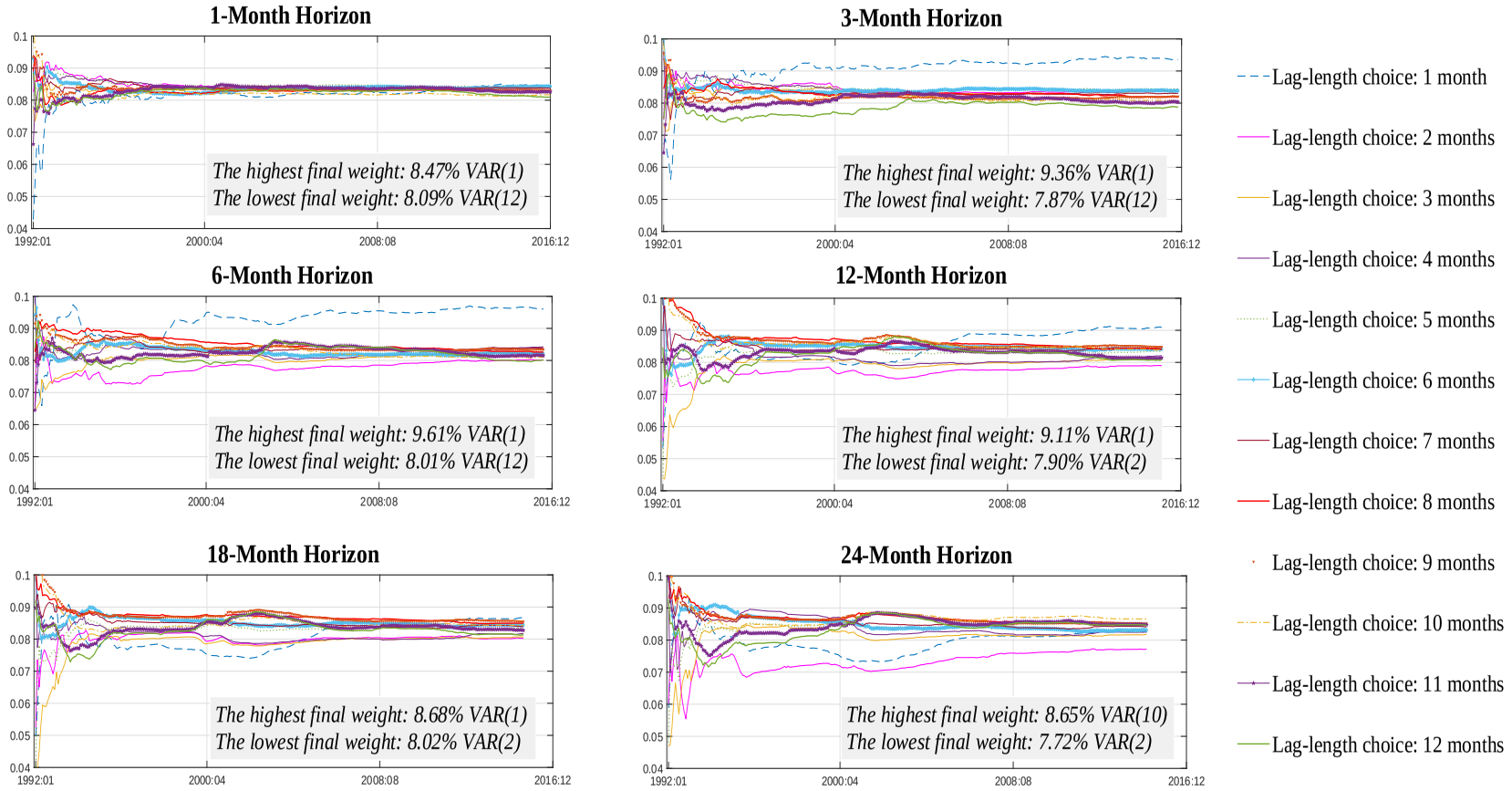

Figure 3.2: Estimated weights for different VAR lag-length choices at forecast horizons 1, 3, 6, 12, 18 and 24 months

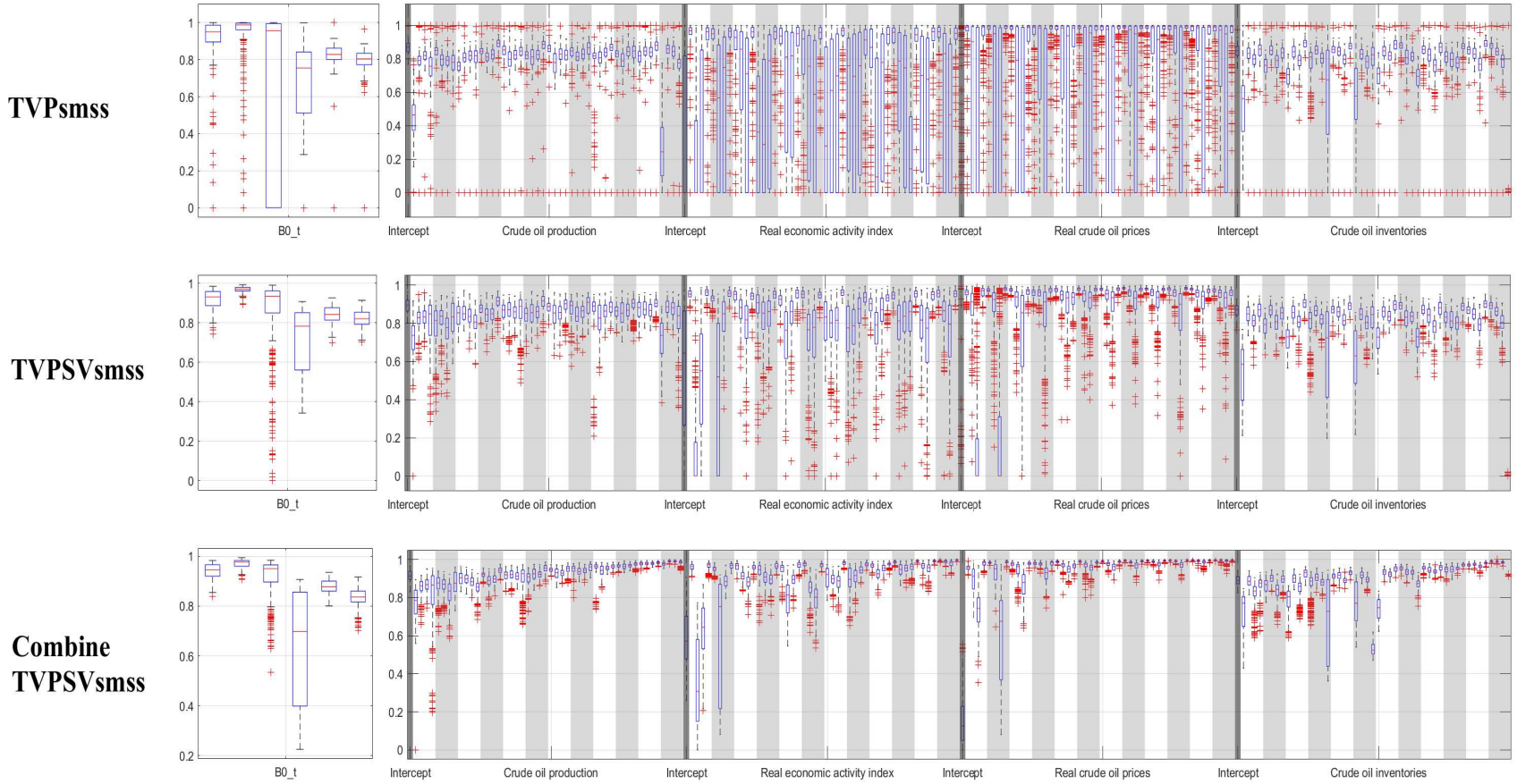

Figure 3.3: Estimated time-invariance probabilities for TVPsmss, TVPSVsmss, and CombineTVPSVsmss

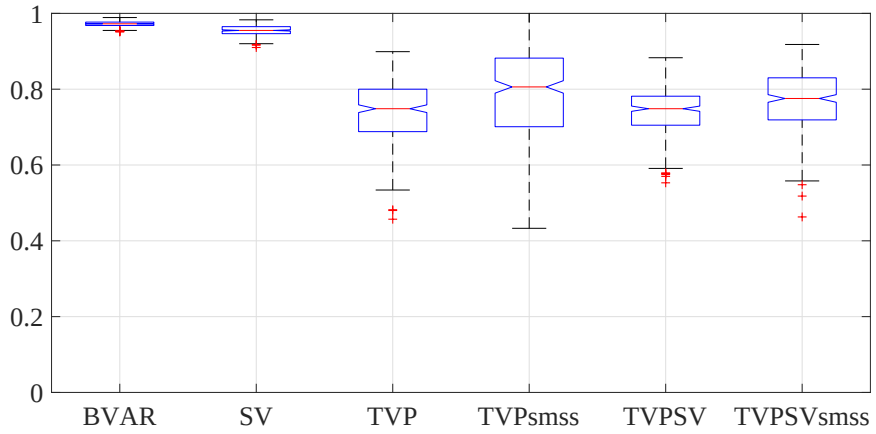

Figure 3.4: The percentage of permissible draws from posterior for BAVR, SV, TVP, TVPsmss, TVPSV and CombineTVPSVsmss over the evaluation period 1992:01–2016:12

stituting an extension of the Cholesky decomposition, as defined in Equation (3.1). Then, the coefficients assigned on the regressions for the crude oil production, the real economic index, crude oil prices and crude oil inventories are presented consecutively. Intercepts are highlighted, with white and black boxes indicating different VAR lag lengths; for example, the white box after the intercept in the sub-figures for all regressions is lag one, $y_{t-1}^{'}$, stocked in the Z_t matrix in Equation (3.4).

Generally, it is clear that the majority of coefficients are well approximated as being time invariant, as they have a low probability, whilst some coefficients still show a high probability of being time variant. Comparing the TVPsmss and TVPSVsmss models, the inclusion of stochactic volatility reduces the time varying probabilities, especially for the real economic activity and the real oil price. Comparing CombineTVPSVsmss relative to TVPSVsmss, the probability of time invariance becomes higher. Therefore, the LOP also restricts the coefficients to be time invariant. This is understandable, since the LOP combines TVPSVsmsss model with the VAR lag length choices from 1 to 12 months. The time-invariance probability of coefficients on TVPSVsmsss with VAR lag 1 is 100% for $y_{t-i}^{'}$, where i=[2, ..., 12].

Moreover, Figure 3.4 presents the percentage of the draws kept for sampling the density forecasts after MKLD, while the probability is based on 300 real-time vintage estimates for each specification including BVAR, SV, TVP, TVPsmss, TVPSV, and TVPSVsmss. It is clear that the TVP introduces a higher probability

of inactive draws than the specifications without it or restricted through SMSS; for example, the inclusion of SMSS, when comparing TVPsmss and TVP, will increase the probability of active draws in the posterior from approximately 70% to 80% on average. Surprisingly, the inclusion of SV, when comparing TVPSV and TVP, will not significantly reduce the probability of efficient draws.

3.4.2 Point forecast evaluation

The point forecasts of all specifications are median estimates of the density forecasts, as it is easier to explain the relation between point and density forecasts. For consistency, the evaluation of point forecasts will also use the recursive inverse mean squared predictive error (MSPE) and success ratios relative to the no-change forecasts introduced in Chapter 2, while the corresponding significance tests are the rough guide of Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test and the Pesaran & Timmermann (2009) test. If a recursive MSPE ratio is lower than 1, the forecast is superior to a no-change forecast. Meanwhile, if the success ratio of a forecast is higher than 0.5, then the forecast is informative in directional forecasts. Hence, those forecasts with lower recursive MSPE and higher success ratios are preferred.

Essentially, the point forecasts indicate that a BK equal combination dominates all other specifications in both point and directional forecasts at all horizons from 1 to 24 months. Table 3.2 presents the real-time forecast accuracy for WTI, RAC, and Brent measures at the 1-, 3-, 6-, 12-, 18- and 24-month horizons during the 1992:01–2016:12 evaluation period, where the forecasts of BVAR, SV, TVP, TVPsmss, TVPSV, TVPSVsmss, and CombineTVPSVsmss are highlighted within rectangular boxes as they utilise the MKLD method for out-of-sample density forecasts. The forecasts from BVAR, TVP, TVPsmss and TVPSV reveal that the recursive MSPE ratios are higher than 1 across all horizons for all three measures considered. Although SV and CombineTVPSVsmss are able to provide lower than 1 recursive MSPE ratios at 1-month horizons for all three measures, none of the forecasts within the box are able to provide statistically significant improvements relative to no-change forecasts. Additionally, it is interesting to note that the inclusion of the SV and SMSS processes in the estimation leads to a reduction of the recursive MSPE ratios from the comparison between TVP and TVPsmss, and

Table 3.2: Real-time forecast accuracy of model specifications at selected forecast monthly horizons

Specifications/MH	Forecast Horizons 1-Month	3-Month	6-Month	12-Month	18-Month	24-Month
			Real WTI price			
			Recursive MSPE ratios			
VAR	**0.985**	1.008	1.124	1.207	1.205	1.157
BVAR	**0.971**	1.022	1.162	1.293	1.268	1.205
SV	**0.936**	**0.997**	1.148	1.333	1.383	1.338
TVP	1.354	1.412	1.451	1.649	1.907	2.553
TVPsmss	1.147	1.180	1.277	1.465	1.814	3.021
TVPSV	1.115	1.329	1.545	2.174	3.383	7.413
TVPSVsmss	1.006	1.190	1.287	1.369	1.436	1.687
CombineTVPSVsmss	**0.928**	1.034	1.163	1.350	1.483	1.760
BK	**0.906****	**0.921****	**0.978***	**0.944****	**0.946****	**0.955****
			Success ratios			
VAR	**0.527**	**0.570****	**0.539***	0.491	**0.519**	**0.520**
BVAR	**0.540***	**0.567****	**0.529**	**0.529**	**0.534**	**0.534**
SV	**0.530**	**0.564****	**0.529**	0.495	**0.516**	**0.516**
TVP	0.497	**0.550****	0.461	0.443	0.428	0.365
TVPsmss	**0.557****	**0.557****	0.471	0.426	0.445	0.433
TVPSV	**0.553****	**0.584****	0.468	0.457	0.428	0.408
TVPSVsmss	**0.517**	**0.510**	0.444	0.478	0.470	0.480
CombineTVPSVsmss	**0.517**	**0.527**	**0.502**	**0.512**	0.470	0.487
BK	**0.553***	**0.581****	**0.539**	**0.606****	**0.576****	**0.552****
			Real U.S. refiners' acquisition cost for oil imports			
			Recursive MSPE ratios			
VAR	**0.987**	1.014	1.155	1.201	1.195	1.116
BVAR	**0.954**	1.020	1.184	1.281	1.239	1.148
SV	**0.933**	**0.999**	1.160	1.302	1.313	1.217
TVP	1.279	1.390	1.421	1.597	1.799	2.314
TVPsmss	1.146	1.173	1.252	1.405	1.694	2.658
TVPSV	1.105	1.309	1.506	2.085	3.064	6.159
TVPSVsmss	**0.991**	1.159	1.280	1.371	1.393	1.544
CombineTVPSVsmss	**0.924**	1.019	1.162	1.337	1.423	1.592
BK	**0.931****	**0.922****	**0.977****	**0.940****	**0.946****	**0.969****
			Success ratios			
VAR	**0.570****	**0.567****	**0.536**	**0.540***	**0.537**	**0.523**
BVAR	**0.597****	**0.564****	**0.532**	**0.578****	**0.558****	**0.545***
SV	**0.580****	**0.574****	**0.525**	**0.550***	**0.541**	**0.527**
TVP	0.500	**0.520**	0.478	0.478	0.396	0.368
TVPsmss	**0.567****	**0.540***	0.468	0.481	0.428	0.437
TVPSV	**0.550****	**0.540***	0.471	**0.519**	0.452	0.419
TVPSVsmss	**0.547****	**0.507**	0.468	**0.526**	0.495	0.491
CombineTVPSVsmss	**0.560****	**0.550****	**0.519**	**0.561****	0.495	0.498
BK	**0.553***	**0.597****	**0.569****	**0.633****	**0.576****	**0.563****
			Real Brent price			
			Recursive MSPE ratios			
VAR	1.003	1.037	1.153	1.206	1.211	1.120
BVAR	1.004	1.055	1.201	1.293	1.266	1.161
SV	**0.968**	1.029	1.180	1.314	1.335	1.233
TVP	1.327	1.378	1.443	1.608	1.793	2.207
TVPsmss	1.146	1.157	1.274	1.424	1.688	2.491
TVPSV	1.143	1.329	1.541	2.108	3.055	6.000
TVPSVsmss	1.041	1.215	1.311	1.386	1.409	1.522
CombineTVPSVsmss	**0.959**	1.062	1.192	1.358	1.446	1.584
BK	**0.940****	**0.934****	**0.978***	**0.932****	**0.939****	**0.948****
			Success ratios			
VAR	0.497	**0.550****	**0.546***	**0.505**	**0.541***	**0.523**
BVAR	**0.510**	**0.547****	**0.536**	**0.543***	**0.562****	**0.531**
SV	0.500	**0.557****	**0.549***	**0.516**	**0.544**	**0.520**
TVP	0.493	**0.530**	0.481	0.443	0.399	0.383
TVPsmss	**0.547****	**0.564****	0.492	0.439	0.438	0.451
TVPSV	**0.550****	**0.564****	0.481	0.478	0.449	0.419
TVPSVsmss	0.500	**0.517**	0.471	0.491	0.498	0.484
CombineTVPSVsmss	0.493	**0.547****	**0.522**	**0.526**	0.484	0.484
BK	**0.520**	**0.560****	**0.525**	**0.592****	**0.572****	**0.542****

Note: MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, a Harvey et al. (1997) small-sample adjustment of the Diebold & Mariano (1995) test is used for the significance test of recursive MSPE ratios. The Pesaran & Timmermann (2009) test for the null hypothesis of no directional accuracy success ratios. $*$ denotes significance at the 10% level and $**$ at the 5% level. The Evaluation period is from 1992:01 to 2016:12.

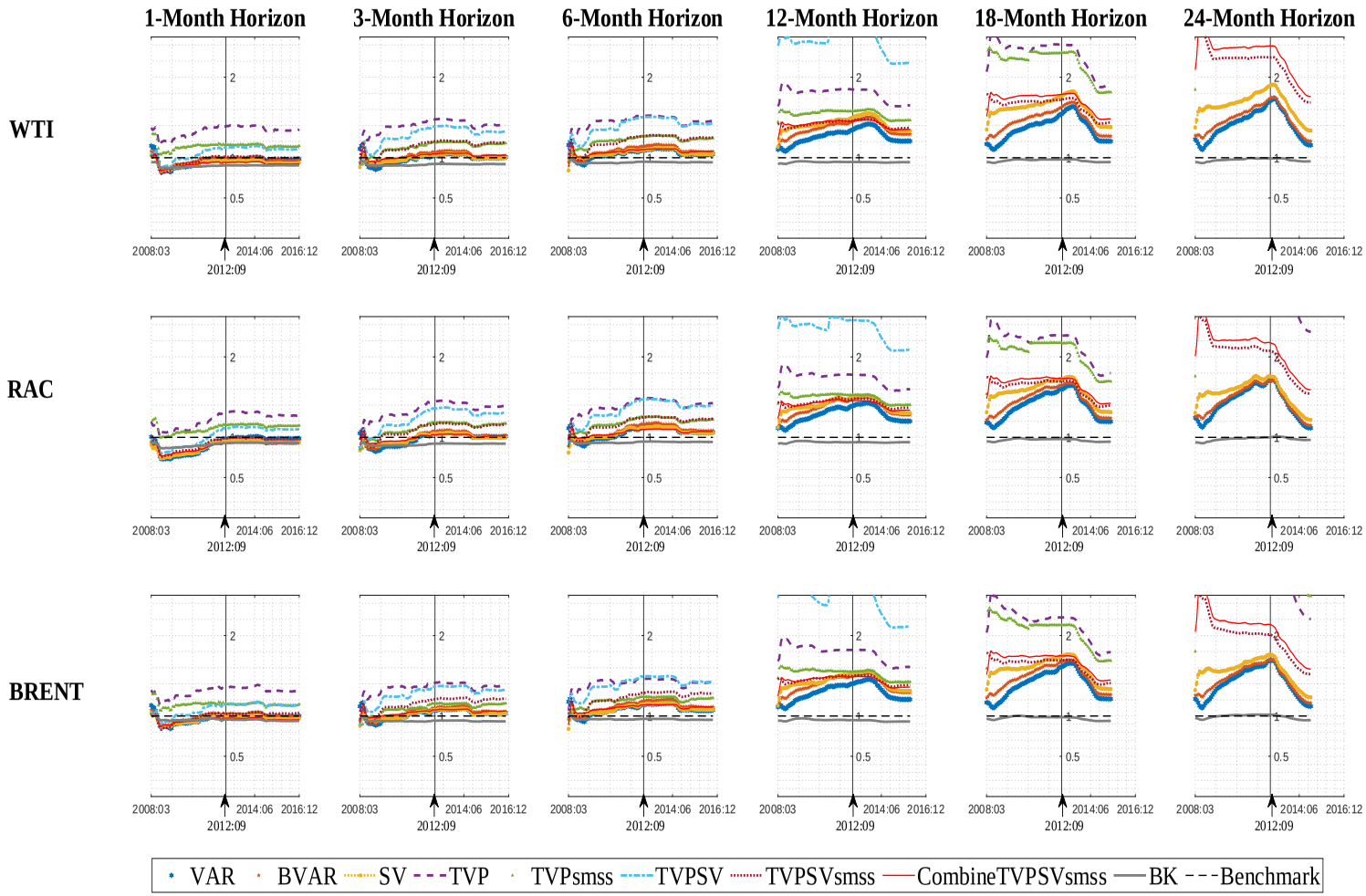

Figure 3.5: Historical recursive MSPE ratios at selected forecast horizons for the evaluation period 2008:03-2016:12

Figure 3.6: Historical Success ratios at selected forecast horizons for the evaluation period 2008:03-2016:12

between TVPsmss and TVPSVsmsss, for all three measures.

Figures 3.5 and 3.6 present the historical MSPE and success ratios over the period after the financial crisis in 2008 to verify the stability of the statistics for the RAC, WTI and Brent measures respectively.[5] Sub-figures illustrate the results generated all the models respectively. The conclusion reflects Table 3.2, whereby only BK can consistently provide preferred recursive MSPE and success ratios. The recursive MSPE ratios of TVP, TVPSV, and TVPsmss are considerably more volatile than others, especially at the horizons larger than 6 months.

From the perspective of the point forecasts, the proposed specifications in this chapter are less accurate than the BK combination. However, the point forecasts are unable to address more economically motivated questions that might be asked by financial market participants. Further, it is quite difficult to distinguish the relative performance among SV, TPVSVsmss, and CombineTVPSVsmss using the point forecasts statistics. The following subsections will turn to the evaluations based on density forecasts.

3.4.3 Density forecast evaluation

Rosenblatt (1952), Dawid (1984) and Diebold et al. (1999) propose a methodology utilising the probability integral transforms (*pits*) of the realised real oil price regarding the forecasts of densities. The *pits* is gauged as:

$$Z_{T+h} = \int_{-\infty}^{r^{oil}_{T+h|M}} p(u)du,$$

where $p(.)$ is the probability generated from the density forecasts of $g(r^{oil}_{T+h|M})$ for the log real RAC measure based on specification M, utilised as one of the endogenous variables in Equation (3.1). As mentioned in Section 3.2.3, forecasts of WTI and Brent measures are extrapolated from the forecast of RAC multiplied by the most recent spread between them and RAC, respectively, from each specification, excepting BK, considered in this chapter. The performance of density forecasts for WTI and Brent measures are closely related to the performance of forecasts for the RAC measure. To illustrate results, I show the forecast performance of the log-level

[5]The historical recursive MSPE and success ratios for WTI and Brent measures are qualitatively the same as the ratios for the RAC measure.

RAC measure, which is directly forecasted from the vector autoregressions.

The *pits*, Z_{T+h} will be uniformly, independently and identically distributed (i.i.d.) if the density forecast is well calibrated (see Diebold et al., 1999). Hereafter, the *pits* provides a means of evaluating the forecast densities relative to the real but unknown densities of the RAC measure.

In fact, the *pits*-related evaluation requires goodness-of-fit tests for both uniformity and independence at the conclusion of the evaluation period. The Anderson–Darling (AD) and Pearson chi-squared (χ^2) tests are employed for uniformity, where the AD test is adjusted through the Kolmogorov–Smirnov test to place higher weight on the tails, as encouraged by Noceti et al. (2003), while the Pearson chi-squared (χ^2) test will separate the range of the Z_{T+h} into eight equally distributed probability classes, and thus uniformity in the histogram can be tested. For independence, the Likelihood Ratio (LR) and Ljung–Box (LB) tests are considered, where the former, proposed by Berkowitz (2001), utilises a three degrees of freedom variant, and the alternative assumes that the Z_{T+h} follows an AR(1) process, while the LB test is designed up to the 4^{th} order of autocorrelation.

Table 3.3 reports all of the four *pits*-based density forecast evaluations of the log of real RAC measure in the first four columns at the 1-, 3-, 6-, 12-, 18- and 24-month horizons, with the specifications listed in rows. Boldface indicates that the test is significant at least at the 5% level. Echoing Table 3.2, I highlight those specifications that utilised MKLD within boxes. Moreover, I also provide the average CRPS, denoted as $\overline{CRPS}(g(r^{oil}_{T+h|M}), r^{oil}_{T+h})$, over the period for each specification, calculated as:

$$\overline{CRPS}(g(r^{oil}_{T+h|M}), r^{oil}_{T+h}) = \frac{1}{\bar{t}-\underline{t}} \sum_{T=\underline{t}}^{\bar{t}} [E_g\{|R - r^{oil}_{T+h}|\} - \frac{1}{2} E_g\{|R - R^*|\}],$$

where r^{oil}_{T+h} is the observation of log real RAC at $T+h$; R and R^* are independent draws of a linear random variable with forecast distribution $g(r^{oil}_{T+h|M})$ at horizon h for vintages $\underline{t}$=1992:01, to $\bar{t}$=2016:12; $E_g\{.\}$ denotes the expectation operator for the predictive $g(r^{oil}_{T+h|M})$; and $\bar{t}-\underline{t}$ is the number of in-report vintages, which is 300 in this exercise. Again, the calculation of the CRPS is based on method 2 as described above — or in Panagiotelis & Smith (2008). The sign of Giacomini & White (2006) test, using the CRPS as the objective function, is shown in the final column.

Boldface indicates an improvement of predictive ability relative to the BK density forecasts based on the Giacomini & White (2006) test, while $*$ denotes significance at least at the 5% level.

In Table 3.3, the *pits* of density forecasts at the 1-month horizon for SV and CombineTVPSVsmss are significantly uniform and independent based on χ^2 and LB tests, and their average CRPSs are smaller than the BK's. Moreover, TVPSV and TVPSVsmss also show significantly lower average CRPS than the BK's, although they are not significantly independent. The SV and CombineTVPSVsmss show qualitatively lower average CRPS and perform better with respect to BK's at all horizons based on the Giacomini & White (2006) test, as well as at least one of the diagnostic tests are significant.

Moreover, Figure 3.7 illustrates histograms of *pits* over the evaluation period for all specifications. It is clear that the density predicted by VAR and BK are too narrow at all forecast horizons, as there far more observations of *pits* near 0 and 1 than in the middle, while this is not observed in those models using the time-varying MKLD model excepting BVAR. It is also necessary to mention that density forecasts generated by TVP are diffuse relative to other densities, especially at horizons larger than 6 months.

The calibrations of *pits* reflect the relative performance of density forecasts, while it is also interesting to consider the shape of the density forecasts. Figure 3.8 presents the density forecasts (98% posterior credibility sets) of log real RAC for TVP, SV, and BK over the 1992:01 to 2016:12 evaluation period, and the corresponding observations.[6] The reason that only plots the three densities is for avoiding to getting bogged down with too many details. SV stands for well calibrated density forecasts, while TVP densities are relatively diffuse. BK is the benchmark.

Generally, SV broadens BK's density forecasts and narrows down the TVP's. The density forecasts of BK are too narrow to track the observations at horizons beyond 3 months. While, the densities forecasted by TVP are too diffuse to reflect the information in terms of the crude oil prices, especially at horizons 18 and 24 months. Different from TPV and BK, SV can track and react sensitively to the

[6]The last observation at forecast horizon h is 2016:12-h+1, as the post-report vintage 2017:06 does not cover the observation after 2017:06.

Table 3.3: Real-time density forecast accuracy of model specifications at selected forecast monthly horizons for log(RAC)

	pits related diagnostics					
	Uniformity		Independence			
Specifications	**AD test**	**χ^2 test**	**LR test**	**LB test**	**Average CRPS**	**GW**
			1-Month Horizon			
VAR	0.000	0.000	0.000	**0.155**	**0.054**	**(-)***
BVAR	0.000	0.000	0.000	**0.353**	0.056	(+)
SV	0.000	**0.285**	0.000	**0.244**	**0.050**	**(-)**
TVP	0.000	0.024	0.000	0.042	0.058	(+)
TVPsmss	0.000	0.000	0.000	0.021	0.056	(+)
TVPSV	0.000	**0.329**	0.000	0.003	**0.055**	**(-)***
TVPSVsmss	0.000	0.001	0.000	0.005	**0.054**	**(-)***
CombineTVPSVsmss	0.000	**0.071**	0.000	**0.086**	**0.051**	**(-)**
BK	0.000	0.000	0.000	0.000	0.055	−−
			3-Month Horizon			
VAR	0.000	0.000	*NaN*	0.003	0.118	(+)
BVAR	0.000	0.000	0.000	0.002	0.109	(+)
SV	0.000	**0.156**	0.000	0.010	**0.101**	**(-)**
TVP	0.000	0.005	0.002	0.000	0.115	(+)
TVPsmss	0.000	**0.439**	0.000	0.000	0.112	(+)
TVPSV	0.000	**0.134**	**0.222**	0.000	0.112	(+)
TVPSVsmss	0.000	0.047	0.000	0.000	0.111	(+)
CombineTVPSVsmss	0.000	**0.168**	0.000	0.001	**0.101**	**(-)**
BK	0.000	0.000	*NaN*	0.002	0.104	−−
			6-Month Horizon			
VAR	0.000	0.000	*NaN*	**0.502**	0.185	(+)
BVAR	0.000	0.000	0.000	**0.334**	0.159	(+)
SV	0.000	**0.290**	0.000	**0.495**	**0.148**	**(-)**
TVP	0.000	0.042	**0.194**	0.019	0.164	(+)
TVPsmss	0.000	**0.404**	0.000	0.020	0.159	(+)
TVPSV	0.000	**0.386**	**0.370**	0.037	0.167	(+)
TVPSVsmss	0.000	0.000	0.000	0.003	0.165	(+)
CombineTVPSVsmss	0.000	0.007	0.000	0.041	**0.147**	**(-)**
BK	0.000	0.000	*NaN*	**0.689**	0.152	−−
			12-Month Horizon			
VAR	0.000	0.000	*NaN*	**0.978**	0.268	(+)
BVAR	0.000	0.000	0.000	**0.957**	0.223	(+)
SV	0.000	0.018	0.012	**0.596**	**0.210**	**(-)**
TVP	0.000	0.005	**0.985**	**0.918**	0.227	(+)
TVPsmss	0.000	0.000	0.000	**0.961**	0.226	(+)
TVPSV	0.000	0.037	**0.382**	**0.752**	0.236	(+)
TVPSVsmss	0.000	0.000	0.000	**0.666**	0.224	(+)
CombineTVPSVsmss	0.000	0.000	0.044	**0.973**	**0.207**	**(-)**
BK	0.000	0.000	0.000	**0.964**	0.214	−−
			18-Month Horizon			
VAR	0.000	0.000	*NaN*	**0.498**	0.337	(+)
BVAR	0.000	0.000	0.000	**0.685**	0.268	(+)
SV	0.000	0.000	**0.243**	0.036	**0.254**	**(-)**
TVP	0.000	0.005	**0.122**	**0.968**	0.280	(+)
TVPsmss	0.000	0.013	0.000	**0.519**	0.278	(+)
TVPSV	0.000	0.017	**0.667**	**0.989**	0.295	(+)
TVPSVsmss	0.000	0.000	0.000	**0.998**	0.258	(+)
CombineTVPSVsmss	0.000	0.000	0.002	**0.402**	**0.246**	**(-)**
BK	0.000	0.000	0.000	**0.599**	0.257	−−
			24-Month Horizon			
VAR	0.000	0.000	*NaN*	**0.353**	0.364	(+)
BVAR	0.000	0.000	0.000	**0.746**	**0.284**	**(-)**
SV	0.000	0.002	**0.093**	**0.108**	**0.274**	**(-)**
TVP	0.000	0.000	0.000	**0.761**	0.332	(+)
TVPsmss	0.000	**0.088**	**0.754**	**0.286**	0.305	(+)
TVPSV	0.000	0.008	0.013	**0.705**	0.337	(+)
TVPSVsmss	0.000	0.000	0.000	**0.756**	**0.279**	**(-)**
CombineTVPSVsmss	0.000	0.002	**0.440**	0.044	**0.269**	**(-)**
BK	0.000	0.000	0.000	**0.114**	0.284	−−

Note: **AD test** is the small-sample (simulated) p-value from the Noceti et al. (2003) test for uniformity of the *pits*, assuming independence of the *pits*. **χ^2 test** is the p-value for the Pearson chi-squared test of uniformity of the *pits* histogram in eight equiprobable classes. **LR test** is the p-value for the likelihood ratio test of zero mean, unit variance and zero first order autocorrelation of the inverse normal cumulative distribution function transformed *pits*, with a maintained assumption of normality for the transformed *pits*. LR test is proposed by Berkowitz (2001). **LB test** is the p-value from a Ljung–Box test for independence of the *pits* based on autocorrelation coefficients up to four. Boldface indicates a statistical significance at least at 5%. According to the Bonferroni correction, the four tests based on pits are jointly controlled by size, reflecting a stricter p-value. For instance, 5% significant level signals a 5%/4=1.25% level. GW is the sign of the test-statistic for the Giacomini & White (2006) test of predictive ability relative to the BK density forecasts, and $*$ denotes a significant level of 5% based on the test.

Figure 3.7: Histogram of *pits* for all specifications at selected horizons

Figure 3.8: Density forecasts (1% to 99% quantiles) for log(RAC) over the period 1992:01–2016:12 based on VAR, SV and BK at selected horizons

real crude oil price at all forecasting horizons during the whole evaluation period excepting the financial crisis in 2008 at the 3- and 6-month horizons.

Nevertheless, the existing "tool box" of the density forecast evaluation can not distinguish the performances of SV, TVPSV, TVPSVsmss, and CombineTVPSV-smss. Moreover, a well-calibrated density from standard statistics is limited to concluding the use of the forecasts with market participants. In the following exercise, I will evaluate the density forecasts from the point of view for market participants, who consider higher frequency (daily) financial data. Finally, I turn to event probability forecasts.

3.4.4 Forecast performance measured using financial market pay-off

As mentioned in the introduction of this chapter, it is important to evaluate the point and density forecasts using higher frequency financial data, since the crude oil is traded in high volumes in daily commodity markets. In asset price text books, an investor can profit from a rise in the crude oil price by taking up a *long* position in the crude oil futures market, if he or she is bullish on crude oil. Meanwhile, if the investor is bearish on crude oil, it is profitable to adopt a *short* position in the crude oil futures market. Accordingly, it is possible to extend the evaluation scheme for models' forecasts to a financial market pay-off. This subsection designs a realistic *long–short* trading strategy using daily crude oil futures and real-time forecasts, which provides the probability of that forecasts are able to generate positive excess returns.

The calculation of excess returns using the "realistic" *long–short* trading strategy requires as inputs a density forecast of the oil price at investment horizon h, daily closing prices of crude oil futures (WTI and Brent) with the h-month maturity, and a threshold $\pi^c \subseteq [0.5, 1]$ reflecting the risk-aversion level. The signal to take either a *long* or *short* position is denoted as $\zeta^i_{d,h,M,\pi^c}|\cdot$ on day d, where $i \subseteq$ [WTI, Brent], and M indicates the density forecast of a model specification. The $|\cdot$ stands that the calculation of a signal is conditional on the futures' LTD and the EIA reporting date, which have multiple forms of arrangements and compositions detailed Appendix B.6. The trading rule is as follows. (1) If the implied real crude oil price using a daily futures with maturity h, which is calculated via a futures-spread

forecast of real crude oil prices widely used in oil price forecasting literature (see Baumeister & Kilian, 2015, Baumeister et al., 2018, and Garratt, Vahey, & Zhang, 2019), is smaller than the $(1-\pi^c)$-quantile of a density forecast at the horizon h, we go *long*, and (2) if the implied real oil price is larger than the π^c-quantile of the density forecast at h, we go *short*. The technical details to construct the realistic strategy are described in Appendix B.6.

I propose the probability of positive excess returns, defined in Appendix B.7, using the data set dated at time T in the financial market as $P(PR_T) = (LP_T + SP_T)/(LN_T + LP_T + SN_T + SP_T)$. The '$LP_T$' represents the total number of positive excess returns, at all horizons $h = 1$ to 24 months and on all the days between months T and $T+1$, which is conditional on that $\zeta^i_{d,h,M,\pi^c}|\cdot$ signals a *long* position; SP_T is the number of positive excess returns conditional on all the *short* positions; LN_T and SN_T indicate the numbers of negative excess returns with the *long* and *short* positions, respectively. The higher the probability of excess returns the more we favour the use of that particular model specification using this criteria.

Table 3.4 presents the average $P(PR_T)$ of vintages from 1992:01 to 2016:12. Columns are risk-aversion levels, including $\pi^c = 50\%$, $\pi^c = 68\%$, $\pi^c = 75\%$, $\pi^c = 85\%$, $\pi^c = 90\%$, and $\pi^c = 95\%$. The upper panel is for the WTI measure, and the lower is for the Brent. The probability in parenthesis after the oil price measures' names are the observed probability of positive excess returns when always taking the *long* positions at all horizons (ex-post). Boldface highlights the highest two average probabilities. The most important signal is that SV can generate at least the second highest average $P(PR_T)$ under the sets of all risk-aversion levels for both WTI and Brent measures. Moreover, CombineTVPSVsmss can generate at least the second highest average probabilities using all π^c above 50% for the Brent measure. Furthermore, VAR's average probabilities are not sensitive to the increase of risk-aversion levels, while the other models' average probabilities of positive returns rise with the increase of π^c. In Appendices B.8 and B.9 I also compare the cumulative excess returns and the probabilities of negative excess returns at horizons 1 to 24 months respectively, which broadly confirm the robustness of observations from $P(PR_T)$.

Table 3.4: Hit rate of financial market excess returns for the WTI and Brent measures, 1992:01–2016:12

	risk-aversion level					
Specifications	$\pi^c = 50\%$	$\pi^c = 68\%$	$\pi^c = 75\%$	$\pi^c = 85\%$	$\pi^c = 90\%$	$\pi^c = 95\%$
			The WTI measure (49%)			
VAR	**0.53**	0.53	0.53	0.54	0.53	0.53
BVAR	**0.53**	0.53	0.54	0.57	0.59	0.61
SV	**0.52**	**0.56**	**0.59**	**0.62**	**0.65**	**0.67**
TVP	0.46	0.51	0.55	**0.63**	0.64	0.64
TVPsmss	0.49	**0.54**	0.55	0.61	0.63	0.65
TVPSV	0.46	0.52	**0.58**	**0.62**	**0.66**	0.62
TVPSVsmss	0.49	0.51	0.56	0.61	0.60	0.65
CombineTVPSVsmss	0.49	0.52	0.57	**0.63**	0.62	**0.67**
BK	0.51	0.53	0.54	0.59	0.62	**0.66**
			The Brent measure (51%)			
VAR	**0.53**	0.54	0.54	0.55	0.55	0.56
BVAR	**0.52**	0.53	0.56	0.60	0.61	0.61
SV	**0.53**	**0.57**	**0.63**	**0.66**	**0.69**	**0.71**
TVP	0.49	0.55	0.56	0.61	0.62	0.63
TVPsmss	0.51	**0.57**	**0.60**	0.59	0.63	0.67
TVPSV	0.48	0.54	0.59	0.62	0.64	0.69
TVPSVsmss	0.48	0.54	0.57	0.63	0.61	0.64
CombineTVPSVsmss	0.49	**0.56**	**0.63**	**0.65**	**0.65**	**0.72**
BK	**0.53**	0.54	0.55	0.61	0.65	0.66

Note: MH represents the monthly forecast horizons. The π^c denotes the choice of probability to do *long* or *short* stratagem, representing the financial markets participants' risk-aversion level. Boldface indicates the top two diagnostic statistics. The probability in parenthesis after measures' names is the observed probability of positive excess returns when always taking the *long* positions at all horizons (ex-post).

3.4.5 Probability forecasts for 'extreme high' and 'extreme low' real oil prices

From the density forecasting evaluation, it is evident that the preferred specifications, including SV and CombineTVPSVsmss, differ from the conclusion in point forecasts, where BK dominates the other specifications. Moreover, the use of daily financial data indicates that the SV can generate a relative high probability of positive excess returns compared to other methods at all forecast horizons. Further, this chapter also extends the evaluation scheme to the probability forecasting of events, because in the crude oil market, participants, including investors and policy makers, have an interest in the probability of any extreme high or low real crude oil price. The subsequent paragraphs will evaluate the forecasts through a probability forecasting for extreme high and low oil prices, statistically analysed through the hit rate, Kuipers Score, a fair bet, as well as Brier Score and its decomposition.

The risk analysis in Alquist et al. (2013) and Wang et al. (2017) sets upper and lower thresholds for the extreme high and low nominal crude oil prices as being higher than \$80 per barrel and lower than \$45 per barrel, respectively. In

contrast, this exercise considers a dynamic attitude of the extreme events. I define the "extreme high price" as a price at time $T+h$, where h is the forecast horizon, that is higher than 95% of the previous observations during the period $T-24$ to T, which is the most recent two years' observations at time T. Symmetrically, the "extreme low price" of real crude oil prices is the observation at time $T+h$ that is lower than 95% previous of the observations in the most recent two years at time T.

The predicted probability of an extreme high crude oil price at time T for $T+h$, based on a specification M, is denoted as $P^{high}_{T+h|M}$, and calculated as:

$$P^{high}_{T+h|M} = 1 - \int_{-\inf}^{Q^{r^{oil}_{T-24,T}}(95\%)} p(u)du, \tag{3.13}$$

where $Q^{r^{oil}_{T-24,T}}(95\%)$ is the value for which the probability of obtaining values from $[r^{oil}_{T-24}, \ldots, r^{oil}_{T}]$ below $Q^{r^{oil}_{T-24,T}}(95\%)$ is 95%, and $p(.)$ is the probability generated from the density forecasts of $g(r^{oil}_{T+h|M})$.

Similarly, the probability forecast of an extreme low of crude oil price is calculated as:

$$P^{low}_{T+h|M} = \int_{-\inf}^{Q^{r^{oil}_{T-24,T}}(5\%)} p(u)du, \tag{3.14}$$

where $Q^{r^{oil}_{T-24,T}}(5\%)$ is the value for which the probability of obtaining values from $[r^{oil}_{T-24}, \ldots, r^{oil}_{T}]$ below $Q^{r^{oil}_{T-24,T}}(5\%)$ is 5%.

Figures 3.9 and 3.10 illustrate the probability forecasts of extreme high and low real crude oil prices at the 1-, 3-, 12-, and 24-month horizons over the 1992:01–2016:12 period, respectively. The 'ACTUAL' is the observations at $T+h$ in the post-report 2017:06 vintage. The x-axis shows vintages T, and $T \subseteq [1992{:}01, \ldots, 2016{:}12]$. Generally, SV is preferred to other models for forecasting the extreme high price, while the preferable model for the extreme low price is uncertain. In per Figures 3.9 and 3.10, I only plot the forecasts from TVP, SV, and BK for a clear demonstration, since the probability forecast performances of SV, TVPsmss, TVPSVsmss, and CombineTVPSVsmss are qualitatively the same. TVP is used for a relatively diffuse density forecast for comparison purposes. Meanwhile, the performances of VAR, BVAR, and BK are the same, with BK being the benchmark.

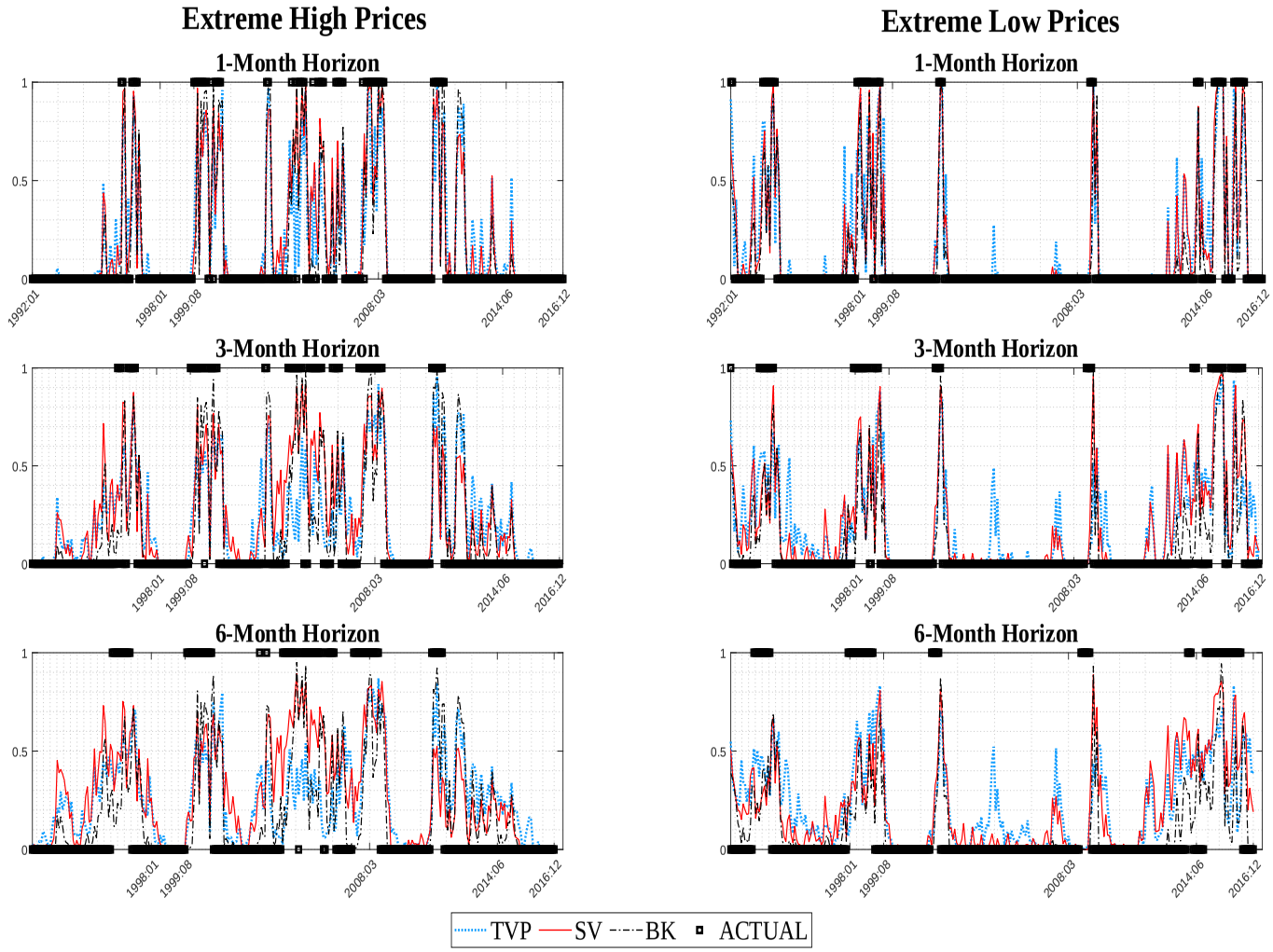

Figure 3.9: Probability forecasts for extreme high and low crude oil prices at the 1-, 3- and 6-month horizons

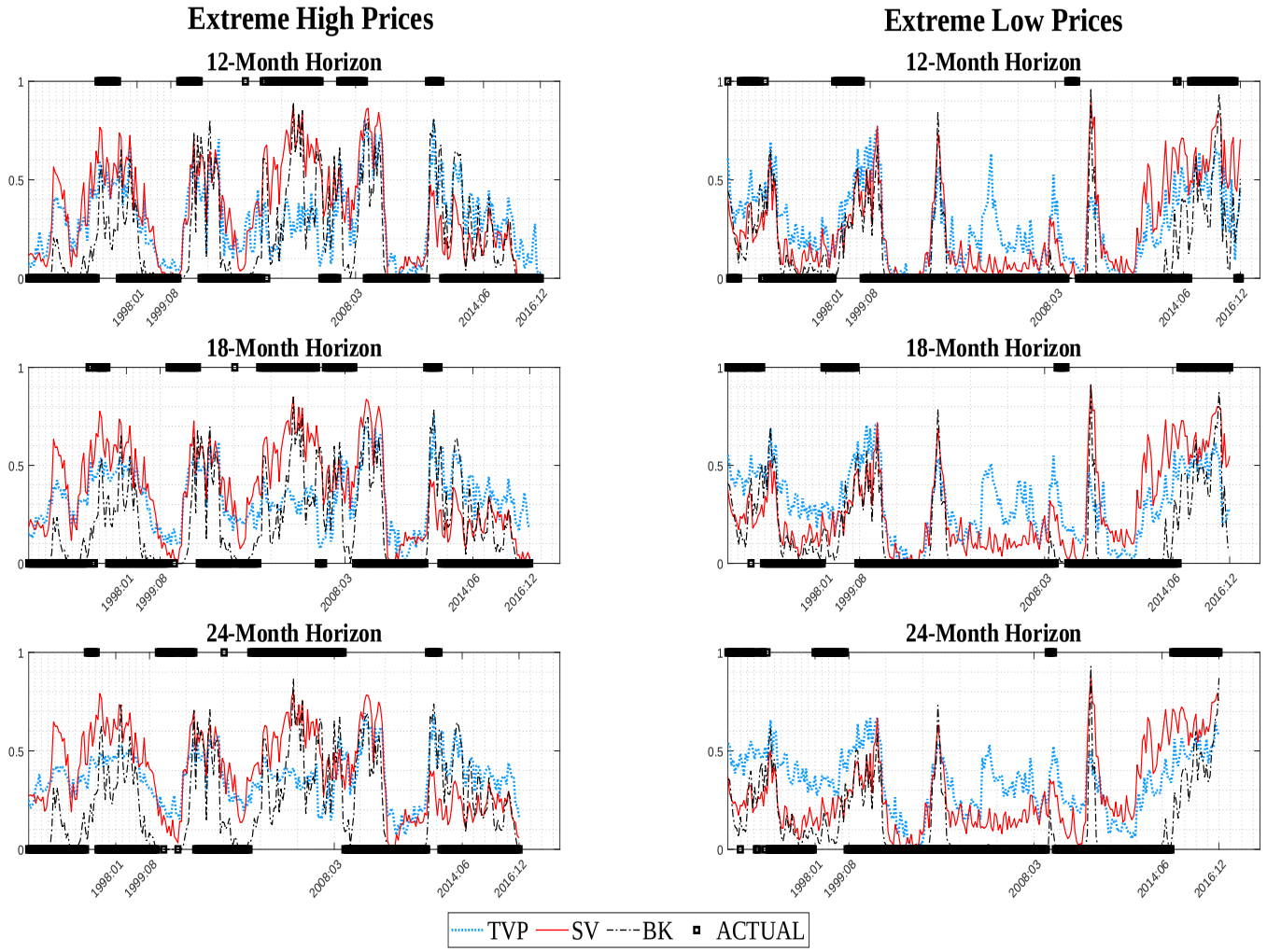

Figure 3.10: Probability forecasts for extreme high and low crude oil prices at the 12-, 18- and 24-month horizons

At the 1-, 3- and 6-month horizons, the probabilities for the extreme high and low crude oil prices forecasted by all three methods are able to meet the majority of the actually observed events. With the growth of forecasting horizons, the probabilities of extreme high and low prices are slightly converging to 50% for all three models. The performance of SV and BK are the same, but TVP shows less predicative power. For example, TVP achieves approximately 40% of the probabilities forecasted for extreme high price in terms of the events realised after the oil price crisis of 1998 and before the financial crisis in 2008 at the 3- and 6-month horizons, while BK and SV can provide a more certain answer with high predictive probabilities.

At the 12-, 18- and 24-month forecast horizons, as shown in per Figures 3.9 and 3.10, none of the models can demonstrate predictability as robustly as at the 1-, 3-, and 6- month forecast horizons. Meanwhile, the probability forecasts for extreme increases are far more accurate than for the extreme decreases. For example, the majority of the realised price extreme increases were predicted by SV and BK, but fewer decreases were forecasted by any of the models. Moreover, the probabilities for the extreme low prices predicted by SV can match the crude oil plunge in 2014. Furthermore, the probability forecasts of the two events from TVP generally fluctuate within the 20% to 60% bands, and are thus uninformative. The reason being that the forecasted density is too diffuse to give instructions.

Hit rate and Kuipers Score

To quantify the performance of events forecasts, there are a range of statistics to verify how often the specifications are able to predict whether the event will or will not manifest. Objective functions-based evaluation provides a broadly flexible environment for decision makers to evaluate the performance of the forecasts made from specifications (see Garratt et al. (2016) and Johnstone et al. (2013) for related analysis).

In this exercise, I illustrate the hit rate and Kuipers Score for a statistical analysis of the probability forecasts for the extreme high and low real crude oil prices (see a related discussion in Garratt et al., 2016). As per the hit rate in financial

markets, I calculate this as:

$$HR = (YY + NN)/(YN + NN + NY + YY),$$

where 'YY' indicates that the extreme high or low crude oil prices are forecast and occurs, while 'YN' indicates the event is forecast but not realised, and so on. Meanwhile, the Kuiper Score is $H - F$, where $H = YY/(YY + YN)$ and $F = NY/(NY + NN)$.

Fair bet

The forecast performance is measured by whether it can help market participants "generate profit" compared with a bet based on rival specifications. Following Garratt et al. (2016), I construct symmetric and asymmetric fair bet games to evaluate the forecasts in Tables 3.5 and 3.6. In symmetric games, if the specification can forecast the event, including extreme high and low prices, and it is observed in the post-report vintage, the specification gains $(s-1)$, but otherwise loses 1. Symmetrically, when the specification does not predict an event that is also not observed in the post-report vintage, then it will get $(s-1)$, but otherwise loses 1. This is summarised in Table 3.5. An asymmetric game is where there is nothing to win or lose if the specification does not foresee the event, and the payout contingencies are summarised in Table 3.6.

Table 3.5: Payout contingencies for the outcome of a symmetric fair bet

Event Forecast	**Event Occurs**	
	Yes	*No*
Yes	$s-1$	-1
No	-1	$s-1$

For a fair bet, the expected payout based on unconditional probability ρ should be zero:

$$(s-1)[\rho^2 + (1-\rho)^2] - 2\rho(1-\rho) = 0.$$

Therefore, $s = \dfrac{1}{2\rho^2 - 2\rho + 1}$. For the asymmetric game, the expected payout should be zero:

$$(s-1)\rho^2 - \rho(1-\rho) = 0.$$

Then, $s = \frac{1}{\rho}$, and the payout for success increases monotonically as $\rho \to 0$. Hence, the ex-post payout of the fair bet rate is conditional on the natural rate based on the post-report observation.

Table 3.6: Payout contingencies for the outcome of an asymmetric fair bet

Event Forecast	**Event Occurs**	
	Yes	*No*
Yes	$s-1$	-1
No	0	0

The payout of each specification M at horizon h in time T is calculated as:

$$W_{T+h|M} = \begin{cases} (s-1)I(R) - (1-I(R)) & \text{if the probability of event}_{T+h|M} > 0.5 \\ (s-1)(1-I(R)) - I(R) & \text{if the probability of event}_{T+h|M} < 0.5, \end{cases}$$

where, $I(R)$ indicates the ex-post observation of the event, which is equal to 1 or 0 depending on whether it is observed or not, respectively; the probability of $\text{event}_{T+h|M}$ is defined in Equations (3.13) and (3.14) for extreme high and low real crude oil prices respectively for specification M. Meanwhile, the payout of the asymmetric game is defined as:

$$W_{T+h|M} = \begin{cases} (s-1)I(R) - (1-I(R)) & \text{if the probability of event}_{T+h|M} > 0.5 \\ 0 & \text{if the probability of event}_{T+h|M} < 0.5. \end{cases}$$

Table 3.7 summarises the hit rates and Kuiper scores, as well as the cumulative pay-out of symmetric and asymmetric games over the 1992:01–2016:12 period, defined as $\Sigma_T W_{T+h|M}$, where $T \subseteq [1992{:}01, \ldots, 2016{:}12]$. The boldface indicates the top two statistics at the corresponding horizons across the specifications. It is clear that SV excels in terms of the priority for forecasting the extreme high real oil

Table 3.7: Forecasting 'extreme high' and 'extreme low' real crude oil prices, 1992:01–2016:12

	Forecast Horizons					
Specifications/MH	1-Month	3-Month	6-Month	12-Month	18-Month	24-Month
			Extreme High Prices			
Unconditional Prob.	19%	23%	26%	28%	30%	31%
			Hit Rate			
VAR	**0.88**	0.82	**0.79**	0.73	**0.69**	0.61
BVAR	**0.87**	**0.83**	**0.78**	**0.75**	0.67	0.60
SV	**0.88**	**0.85**	**0.79**	**0.76**	**0.69**	0.63
TVP	0.86	0.80	0.74	0.70	0.66	0.62
TVPsmss	**0.87**	0.80	0.75	0.69	0.62	0.63
TVPSV	**0.87**	0.82	0.75	0.74	0.66	0.61
TVPSVsmss	0.86	0.80	0.77	0.72	0.66	0.63
CombineTVPSVsmss	**0.87**	**0.83**	**0.78**	0.72	0.67	**0.64**
BK	**0.88**	0.82	0.75	0.73	**0.68**	**0.68**
			Kuipers Score			
VAR	**0.63**	0.49	**0.47**	0.29	0.19	0.01
BVAR	0.56	0.51	0.43	**0.36**	**0.20**	0.04
SV	**0.59**	**0.57**	**0.46**	**0.41**	**0.26**	**0.15**
TVP	0.55	0.43	0.29	0.15	0.08	-0.10
TVPsmss	**0.59**	0.44	0.31	0.10	-0.08	-0.02
TVPSV	0.57	0.49	0.31	0.32	0.10	-0.06
TVPSVsmss	0.56	0.45	0.44	0.22	-0.02	-0.17
CombineTVPSVsmss	**0.59**	**0.52**	**0.46**	0.23	0.04	-0.05
BK	**0.63**	0.48	0.34	0.30	0.16	**0.19**
			Returns to Fair Bet [Symmetric/Asymmetric]			
VAR	[**345.55/230.45**]	[320.14/161.80]	[317.34/136.42]	[273.89/89.89]	[**244.01**/61.16]	[190.28/34.32]
BVAR	[333.37/225.45]	[**327.77/194.56**]	[**309.46/141.25**]	[**287.23/137.08**]	[235.86/**86.93**]	[182.04/52.84]
SV	[**340.68/240.09**]	[**343.03/235.28**]	[**319.97/184.42**]	[**297.90/176.23**]	[**244.01/121.58**]	[201.28/**102.87**]
TVP	[330.94/207.00]	[309.96/157.80]	[280.55/87.17]	[247.21/51.77]	[222.28/35.90]	[193.03/4.59]
TVPsmss	[338.24/215.82]	[309.96/147.88]	[283.17/92.08]	[239.21/41.73]	[192.40/4.77]	[201.28/19.88]
TVPSV	[333.37/213.82]	[320.14/161.80]	[283.17/99.92]	[279.22/88.38]	[227.71/40.78]	[187.54/14.88]
TVPSVsmss	[330.94/195.36]	[309.96/128.04]	[304.20/96.17]	[265.89/41.19]	[227.71/9.14]	[201.28/-4.71]
CombineTVPSVsmss	[338.24/227.45]	[**327.77**/179.68]	[**309.46**/109.92]	[265.89/55.25]	[230.43/18.77]	[**209.53**/10.59]
BK	[**345.55/230.45**]	[320.14/181.64]	[288.43/121.50]	[273.89/96.92]	[**238.58**/53.41]	[**237.01/69.76**]
			Extreme Low Prices			
Unconditional Prob.	15%	17%	20%	22%	26%	25%
			Hit Rate			
VAR	**0.93**	**0.88**	**0.79**	**0.76**	0.72	0.75
BVAR	**0.93**	**0.88**	**0.79**	0.75	**0.73**	0.75
SV	**0.92**	0.86	**0.79**	0.75	**0.75**	0.77
TVP	0.91	**0.87**	**0.79**	**0.76**	**0.73**	0.74
TVPsmss	**0.93**	0.86	**0.78**	0.75	**0.75**	0.75
TVPSV	**0.93**	**0.87**	0.77	**0.76**	**0.73**	0.75
TVPSVsmss	**0.93**	0.85	**0.75**	0.75	**0.75**	**0.82**
CombineTVPSVsmss	**0.93**	**0.87**	0.77	**0.76**	**0.73**	**0.79**
BK	0.91	0.86	**0.79**	**0.77**	**0.73**	0.75
			Kuipers Score			
VAR	0.74	**0.57**	**0.32**	**0.31**	0.24	0.31
BVAR	0.75	**0.59**	**0.34**	**0.27**	**0.26**	0.32
SV	0.70	0.51	0.31	0.24	**0.31**	0.36
TVP	0.64	0.53	0.31	0.25	0.23	0.25
TVPsmss	0.73	0.51	0.29	0.24	**0.31**	0.27
TVPSV	0.73	0.55	0.23	0.26	**0.26**	0.29
TVPSVsmss	**0.78**	0.47	0.09	0.17	**0.31**	**0.62**
CombineTVPSVsmss	**0.80**	0.57	0.17	0.18	0.22	**0.49**
BK	0.73	0.56	0.27	0.25	0.21	0.20
			Returns to Fair Bet [Symmetric/Asymmetric]			
VAR	[**353.49/216.00**]	[**325.13/146.92**]	[280.22/**87.08**]	[**266.79/100.47**]	[248.92/**76.42**]	[261.39/**88.13**]
BVAR	[**351.15**/203.36]	[**327.52/147.92**]	[**285.18/85.17**]	[261.73/**80.89**]	[256.78/73.67]	[263.98/82.99]
SV	[344.15/194.55]	[315.58/128.04]	[280.22/83.17]	[259.21/65.83]	[**269.88/81.55**]	[274.34/86.99]
TVP	[337.15/191.55]	[317.97/134.00]	[280.22/79.25]	[**266.79**/54.77]	[256.78/53.53]	[251.04/62.62]
TVPsmss	[**351.15/209.18**]	[315.58/133.00]	[277.74/70.42]	[261.73/59.80]	[**272.50**/68.16]	[258.80/56.40]
TVPSV	[**351.15/209.18**]	[320.36/125.08]	[267.82/58.58]	[266.79/65.31]	[259.40/63.16]	[258.80/74.84]
TVPSVsmss	[**353.49**/198.55]	[308.42/110.16]	[250.47/24.17]	[259.21/34.19]	[**272.50**/48.03]	[**310.57/91.76**]
CombineTVPSVsmss	[**353.49**/192.73]	[322.74/126.08]	[267.82/31.17]	[264.26/29.16]	[262.02/32.52]	[**287.28**/58.18]
BK	[339.48/157.64]	[313.19/82.40]	[**282.70**/37.17]	[**274.37**/33.16]	[262.02/26.77]	[258.80/19.51]

Note: MH represents monthly forecast horizons. Unconditional probability of the extreme high and low prices during the 1992:01–2016:12 period are listed in the first row of each panel. Boldface indicates the top two diagnostic statistics.

price, which collects the majority of the highest hit rates, Kuipers Scores, and symmetric and asymmetric returns to fair bet at all the horizons presented. At the same time, there is no uniform conclusion on the relative performance in forecasting extreme low prices. TVPSVsmss and CombineTVPSVsmss can provide high hit rates, Kuipers Scores, and symmetric fair bet returns at the 1- and 24-month horizons, while VAR and BVAR perform better at the 3-, 6-, 12-, and 18-month horizons.

Brier Score and its decomposition

This subsection computes the Brier Score (Brier, 1950) to gain further insight into the forecast performance for extreme high and low real crude oil prices. A well-known decomposition of the Brier Score into *uncertainty*, *reliability* and *resolution* (Murphy, 1973), can be written as:

$$BS = \overline{o}(1-\overline{o}) + \frac{1}{N}\sum_{\kappa=1}^{K} n_\kappa (p_\kappa - \overline{o}_\kappa)^2 - \frac{1}{N}\sum_{\kappa=1}^{K} n_\kappa (\overline{o}_\kappa - \overline{o})^2, \tag{3.15}$$

where $\overline{o}$ is the observed frequency of a event over N evaluation periods and not the forecasts; N is the number of evaluation periods. The whole sample is split into verification groups, partitioned into deciles of forecast probability, such as the first group of 1 to 5% forecasts comparing with corresponding observations, the second contains 6 to 10% and so on. K is the number of groups, which we set 20 in this application; and, n_κ is the number of forecasts which fall into the κ_{th} group. The term p_κ denotes the forecast probability of the event, while ex-post event frequencies observed in the group κ are denoted as $\overline{o}_\kappa$.

There are three terms, measuring *uncertainty*, *reliability* and *resolution*. The calculation of *uncertainty* only uses the data to construct $\overline{o}$ and not the forecasts. Hence, this term is the same for all specifications. According to Galbraith & Norden (2012), *reliability* is the consistency between forecast probabilities p_κ and ex-post observed event frequencies $\overline{o}_\kappa$. Therefore, the lower is the second term indicates high consistency between these two. For example, suppose 15% is the probability of an extreme low real oil price, and the specification predicts that the real oil price falls below the 5%-quantile of the last two years' prices, which is the definition of the extreme low oil price in this application, for 15% of the observations, then I have a perfect match with the data, and *reliability* is perfect, with this term being zero.

The third term *resolution* measures the difference between the observed frequency $\overline{o}_\kappa$ in the κ^{th} group and the overall sample average frequency $\overline{o}$. The *resolution* enters negatively into the decomposition this term measures the ability to distinguish between relatively high-probability and low-probability events. If it is high, the conditional expectation of the outcomes will differ largely from its unconditional mean. Hence, high *resolution* lowers the Brier Score. Low Brier Scores are preferred.

Table 3.8 reports the Brier Score and it's decomposition for all model specifications, using as thresholds the extreme high and low real oil prices as in Table 3.7. The SV and CombineTVPSVsmss have approximately the same Brier Score at all forecast horizons for forecasting extreme high prices, and both outperform the benchmark VAR and BK at the 1-, 3-, 6-, 12-, and 18-month horizons. That is, the Brier Score is the highest for the VAR in the upper panel of Table 3.8. Comparing with the VAR, the major contribution to the superior performance comes from the second term at all horizons, measuring *reliability*, while the contribution becomes positive (reducing the Brier Score) from the third term (*resolution*) at the 12-, 18-, and 24-month horizons. This is because SV's density forecasts are not as sensitive as VAR's to the change of forecast horizons, as shown in Figure 3.7. Hence, the reduction of VAR's ability to distinguish between relatively high-probability and low-probability events is larger than that of SV's. Consequently, the *resolution* for VAR gradually reduces from 0.096 at the 1-month to 0.026 at the 24-month horizons, while it for SV decreases from 0.085 to 0.029. Therefore, the longer horizons of SV perform better than the benchmark VAR by the way of both *reliability* and *resolution*, and as opposed to *reliability* only at the shorter horizons.

For forecasting extreme low real oil prices in the lower panel of Table 3.8, the model specifications using the MKLD produced lower Brier Scores than the VAR at all forecast horizons. It is necessary to note that CombineTVPAVsmss performed slightly better than SV at the 12-, 18-, and 24-month horizons. The TVP, TVPSV, and TVPSVsmss, which allow parameters be more time-varying than both SV and CombineTVPSVsmss do, performed slightly better than all other specifications at the 12-, 18- and 24-month horizons for forecasting the low price, respectively.

Table 3.8: Brier Score and decomposition for 'extreme high' and 'extreme low' real crude oil prices forecasts, 1992:01–2016:12

	Forecast Horizons					
Specifications/MH	1-Month	3-Month	6-Month	12-Month	18-Month	24-Month
Extreme High Prices						
VAR	0.089[*0.152; 0.032; 0.096*]	0.136[*0.176; 0.037; 0.077*]	0.174[*0.193; 0.046; 0.065*]	0.213[*0.200; 0.061; 0.047*]	0.273[*0.209; 0.086; 0.022*]	0.336[*0.214; 0.148; 0.026*]
BVAR	0.097[*0.152; 0.026; 0.080*]	0.123[*0.176; 0.024; 0.077*]	0.150[*0.193; 0.017; 0.060*]	0.177[*0.200; 0.033; 0.056*]	0.210[*0.209; 0.029; 0.028*]	0.238[*0.214; 0.055; 0.031*]
SV	**0.083**[*0.152; 0.016; 0.085*]	**0.110**[*0.176; 0.008; 0.074*]	**0.135**[*0.193; 0.006; 0.064*]	**0.158**[*0.200; 0.013; 0.056*]	**0.196**[*0.209; 0.023; 0.036*]	**0.222**[*0.214; 0.037; 0.029*]
TVP	0.096[*0.152; 0.016; 0.072*]	0.128[*0.176; 0.008; 0.056*]	0.167[*0.193; 0.021; 0.047*]	0.193[*0.200; 0.022; 0.029*]	0.220[*0.209; 0.031; 0.019*]	0.239[*0.214; 0.039; 0.015*]
TVPsmss	0.095[*0.152; 0.014; 0.071*]	0.129[*0.176; 0.006; 0.054*]	0.161[*0.193; 0.024; 0.056*]	0.190[*0.200; 0.023; 0.033*]	0.223[*0.209; 0.038; 0.024*]	0.237[*0.214; 0.042; 0.019*]
TVPSV	0.090[*0.152; 0.011; 0.073*]	0.127[*0.176; 0.010; 0.059*]	0.162[*0.193; 0.016; 0.047*]	0.187[*0.200; 0.022; 0.035*]	0.234[*0.209; 0.044; 0.018*]	0.253[*0.214; 0.055; 0.016*]
TVPSVsmss	0.091[*0.152; 0.015; 0.077*]	0.134[*0.176; 0.016; 0.057*]	0.164[*0.193; 0.016; 0.045*]	0.186[*0.200; 0.016; 0.030*]	0.224[*0.209; 0.030; 0.015*]	0.253[*0.214; 0.070; 0.031*]
CombineTVPSVsmss	**0.087**[*0.152; 0.018; 0.083*]	**0.116**[*0.176; 0.009; 0.070*]	**0.141**[*0.193; 0.005; 0.058*]	**0.164**[*0.200; 0.023; 0.060*]	**0.198**[*0.209; 0.025; 0.036*]	0.226[*0.214; 0.023; 0.011*]
BK	0.094[*0.152; 0.021; 0.079*]	0.132[*0.176; 0.038; 0.082*]	0.161[*0.193; 0.026; 0.058*]	0.182[*0.200; 0.023; 0.041*]	0.212[*0.209; 0.028; 0.025*]	**0.225**[*0.214; 0.033; 0.022*]
Extreme Low Prices						
VAR	0.051[*0.125; 0.012; 0.086*]	0.112[*0.140; 0.023; 0.051*]	0.187[*0.162; 0.061; 0.036*]	0.227[*0.172; 0.078; 0.024*]	0.264[*0.191; 0.091; 0.019*]	0.240[*0.185; 0.079; 0.024*]
BVAR	0.056[*0.125; 0.010; 0.079*]	0.102[*0.140; 0.016; 0.054*]	0.160[*0.162; 0.038; 0.039*]	0.175[*0.172; 0.036; 0.034*]	0.201[*0.191; 0.041; 0.032*]	0.188[*0.185; 0.047; 0.045*]
SV	**0.050**[*0.125; 0.011; 0.086*]	**0.097**[*0.140; 0.016; 0.059*]	**0.138**[*0.162; 0.030; 0.053*]	0.156[*0.172; 0.026; 0.042*]	0.175[*0.191; 0.024; 0.040*]	0.170[*0.185; 0.025; 0.040*]
TVP	0.058[*0.125; 0.009; 0.076*]	0.102[*0.140; 0.010; 0.048*]	0.146[*0.162; 0.009; 0.025*]	**0.150**[*0.172; 0.026; 0.048*]	0.169[*0.191; 0.024; 0.046*]	0.167[*0.185; 0.024; 0.042*]
TVPsmss	0.052[*0.125; 0.012; 0.086*]	0.098[*0.140; 0.017; 0.059*]	0.143[*0.162; 0.015; 0.034*]	0.157[*0.172; 0.020; 0.036*]	**0.163**[*0.191; 0.013; 0.042*]	0.162[*0.185; 0.014; 0.037*]
TVPSV	0.049[*0.125; 0.009; 0.085*]	0.099[*0.140; 0.018; 0.059*]	0.150[*0.162; 0.024; 0.036*]	0.159[*0.172; 0.015; 0.029*]	0.180[*0.191; 0.018; 0.030*]	0.172[*0.185; 0.020; 0.033*]
TVPSVsmss	0.052[*0.125; 0.005; 0.078*]	0.104[*0.140; 0.010; 0.045*]	0.160[*0.162; 0.029; 0.030*]	0.159[*0.172; 0.017; 0.030*]	0.168[*0.191; 0.007; 0.031*]	**0.151**[*0.185; 0.020; 0.054*]
CombineTVPSVsmss	**0.050**[*0.125; 0.009; 0.084*]	**0.097**[*0.140; 0.011; 0.054*]	**0.139**[*0.162; 0.020; 0.043*]	**0.151**[*0.172; 0.016; 0.037*]	**0.166**[*0.191; 0.012; 0.037*]	**0.154**[*0.185; 0.007; 0.038*]
BK	0.057[*0.125; 0.010; 0.078*]	0.105[*0.140; 0.019; 0.054*]	0.143[*0.162; 0.044; 0.063*]	0.154[*0.172; 0.039; 0.058*]	0.171[*0.191; 0.041; 0.062*]	0.169[*0.185; 0.042; 0.058*]

Note: Brier Score, followed in parenthesis by *Uncertainty*, *Reliability* and *Resolution*, respectively. MH represents monthly forecast horizons. Boldface indicates the lowest two Brier Scores.

3.5 A brief discussion for the results so far

The in-sample statistics have illustrated that the prior SMSS can efficiently restrict time-varying parameters to be time invariant. The allowance of stochastic volatility, comparing TVPSVsmss to TVPsmss, will further reduce the time variation of parameters under the prior of SMSS. Meanwhile, the use of LOP, such as CombineTVPSVsmss, will give a relative high and consistent weight on the TVPSVsmss with one VAR lag length, and hence, restrict the majority of time-varying parameters in TVPSVsmss to be constant. Therefore, an out-of-sample forecasting comparison for the specifications listed in Table 3.1 amounts to a hierarchical model comparison for time-varying parameters.

For the 1992:01–2016:12 evaluation period, BK dominates the new model space using MKLD considering time-varying parameters, stochastic volatility, SMSS and LOP, respectively, based on recursive MSPE and success ratios for point forecasts. However, the evaluations of density forecasts, including *pits* based calibrations and CRPS, indicate that both the SV and CombineTVPAVsmss are better calibrated than the VAR and BK benchmarks.

The Subsections 3.4.4 and 3.4.5, therefore, evaluate the probability that forecasts can generate positive excess returns using daily oil futures and the event forecasts of extreme high and low real oil prices, respectively. Incorporating high frequency financial data, using the point forecast, where $\pi^c = 50\%$, increases the probability of positive excess returns from 49% (51%) when an investor always takes *long* position to a maximum of 53% (53%) across specifications for the WTI (Brent) measure. However, the density forecasts are more flexible, as it enables investors to choose a risk-aversion level (π^c). When ($\pi^c = 95\%$), the probabilities of positive excess returns are as high as 67% and 71% based on the density forecasts of SV for WTI and Brent, respectively. The results of the probability for positive excess returns indicate that the SV is well calibrated in both two price measures, and CombineTVPAVsmss is also an improvement for the Brent measure.

For event forecasts, the Hit rates, Kuipers Score and Fair bet metrics suggest that the SV is better calibrated than other specifications at a majority of forecast horizons for extreme high oil prices. In contrast, these statistics do not support SV for forecasting extreme low oil prices. To gain further insights to the event

forecasts, Brier Scores are calculated, and indicate that both SV and CombineTVPAVsmss are well calibrated for the extreme high and low events, whereas the VAR and BK benchmarks are not, despite this the BK benchmark is tougher to beat at the 24-month horizon. Overall, there is strong empirical evidence supporting the SV and CombineTVPSVsmss models, which are top two specifications restricting time-varying parameters. In short the gain largely arises due to the use of stochastic volatility.

3.6 Concluding remarks

I have developed the MKLD for improving highly parameterised models' out-of-sample density forecast accuracy through minimising one-step-ahead in-sample posterior estimates' Kullback–Leibler 'distance', followed by an elimination process for extreme forecasts. Moreover, this chapter has extended the evaluation of the model comparisons from a standard statistical evaluation of point and density forecasts to one that selects specifications that are more valuable in terms of excess returns when trading in the crude oil futures market, and forecasting the likelihood of extreme high and low real crude oil prices.

In order to determine the time variation of the parameters in the real crude oil price forecasts, this chapter has explored the vast model space of hybrid methods, which includes stochastic volatility, constant and time-varying parameters, as well as specifications allowing some of the VAR coefficients that are constant, while others that are time varying, using MKLD and the extended evaluation scheme. Over the 1992:01–2016:12 period, I find overwhelming support for models employing stochastic volatility and SMSS versus the conventional VAR and BK. However, the majority of the gains result from stochastic volatility.

From the point forecasting perspective, BK dominates those generations using the time-varying MKLD model over the 1992:01–2016:12 period. However, SV and CombineTVPSVsmss with the benefit of MKLD can provide enhanced calibrated density forecasts compared to BK at all horizons based on the Giacomini & White (2006) test utilising CRPS as a loss function. From a realistic *long-short* stratagem using daily crude oil futures, SV can provide higher probability of positive excess returns than other models considered in this exercise by taking *long* or

short positions during the evaluation period. SV using MKLD provides the highest hit rate, Kuipers Score and a fair bet, as well as at least the second smallest Brier Scores in order to forecast the extreme high real crude oil price at all forecast horizons. CombineTVPSVsmss provides the smallest Brier Scores in the extreme low real crude oil price forecasting at all forecast horizons

Moreover, the *long–short* stratagem used for evaluation purpose can be applied in practice. Firstly, it offers a more flexible platform to construct portfolio in the oil-futures market. Secondly, policy makers can utilise the process as a monitor to track the price of crude oil to facilitate policy reactions to predictions.

Chapter 4

Constrained Bayesian SVARs: Two Specifications in the World Crude Oil Market Modelling

Structural vector autoregressions (SVARs) have been widely used for Macroeconomic inferences since Sims (1980). A trade off between the computation efficiency and the complexity of the identification becomes a central issue. An increasingly popular way of identifying structural shocks is through a set of identification restrictions, which includes sign and other qualitative restrictions when shocks happen and in their aftermath. Computation using the traditional accept-reject approach can take a very long time since many draws can be rejected. An alternative way to identify only sign-restricted VARs using the Bayesian method, proposed by Baumeister & Hamilton (2015), does not run into this computational problem. However, this method limits identification restrictions to the impact matrix and, in practice, can lead to imprecise estimates of structural parameters. This chapter proposes a parallel Metropolis–Hastings algorithm to identify and compute Bayesian SVARs, which extends Baumeister & Hamilton's (2015) method from sign to a set of identification restrictions. Two specifications commonly used in the world crude oil market modelling are employed to illustrate that the new method offers improvements in terms of computational efficiency and a more precise estimation on crude oil demand elasticities. Moreover, I argue that the novel method proposed in this chapter can identify uncertainty of restrictions on non-linear structural parameters, for example demand and supply elasticities. This chapter provides empirical evidences that the uncertainty of restrictions leads to precise estimates of the non-linear parameters of interest.

4.1 Introduction

Structural vector autoregressions (SVARs) have provided valuable insights in the field of applied macro-econometrics since Sims (1980). The issue of identification becomes central for precise estimates of structural parameters. However, the traditional accept-reject method for SVARs is computationally inefficient when imposing restrictions, such as the dynamic sign restrictions in Kilian & Murphy (2014) and ranking restrictions in Amir-Ahmadi & Drautzburg (2019), referred as a 'set identification' in SVARs literature. Often only a low number of admissible draws are available when computing standard errors around point estimates, making for imprecise inference. For example, the sifting of 16 models in Kilian & Murphy (2014) required 5,000,000 posterior draws using their method from Journal of Applied Econometrics data archive. The econometric literature, on the one hand, has concentrated on improving the efficiency of processes for sampling the orthogonal rotation matrix using Bootstrapping or Bayesian methods (see Uhlig, 2005; Rubio-Ramirez et al., 2010; Inoue & Kilian, 2013; Kilian & Murphy, 2014; Arias et al., 2018; Amir-Ahmadi & Drautzburg, 2019; and Arias et al., 2019). Nevertheless, these methods still do not elevate the impact caused by the complexity of identification restrictions. (See a detailed discussion in Section 4.2.) On the other hand, Baumeister & Hamilton (2015, 2018, 2019) and Waggoner et al. (2016) sample structural parameters directly, but (1) they are unable to introduce the set restrictions after the impact of shocks; (2) they require a specific formation for endogenous variables aiming to identify structural parameters directly. Hence, the second approach changes both the method and the formation of the endogenous variables relative to the corresponding literature, which leads to difficulties in terms of drawing comparisons between them.

Mindful of the gap in SVARs' literature, this chapter proposes a novel identification and computation for Baumeister & Hamilton's (2015) Bayesian SVARs. Baumeister & Hamilton (2015, 2018, 2019) only allow for sign restrictions on the contemporaneous relations matrix of variables, using positive (or negative) truncated Student t distributions. In contrast, this chapter identifies parameters contained in the inverse contemporaneous relations matrix, which are the impulse response functions (IRFs) at the zero horizon, using a truncated Student t distributions for the sign restrictions. Moreover, a set identification is enabled, which include a broad range, such as zero, sign and dynamic sign restrictions on IRFs, as well as bounds (and their uncertainty) on elasticities. This is achieved through a (parallel) random-walk Metropolis–Hastings (MH) sampler, which samples estimates constrained by

the set identification. I refer to this as constrained Bayesian SVARs, thereafter denoted as C-BSVARs in this chapter.

For the computation, C-BSVARs reform Baumeister & Hamilton's (2015) Bayesian SVARs in the following ways. (1) C-BSVARs numerically solves a conditional arg-minimisation for the initial values of the parameters for Baumeister & Hamilton's (2015) Markov chain Monte Carlo (MCMC) iteration, which is strictly restricted by the set identification. To improve the computation efficiency, the log-likelihood function will be multiplied by an indicator, signalling that the estimates satisfy the set of identification restrictions, in the arg-minimisation step. (2) I keep the acceptance rate within a range 30%-40%, and let the MCMC searching step length adjust automatically through a scale factor. (3) I independently draw the random–walk MH samplers under different restrictions by adding the uncertainty of the threshold lower (or higher) bounds. This is because of the weakness of the standard MH algorithm, which draws samples only within a local maximum when the posterior is a multimodal distribution, highlighted in Waggoner et al. (2016).[1] As a result, independent MCMC draws can be sampled using a parallel calculation for improving the computational efficiency.[2]

In a special case, Kilian & Murphy (2014) propose global crude oil market SVARs, which explicitly accounts for shocks to the speculative demand for oil as well as shocks to the flow demand and flow supply. There were two major contributions in their paper. First, speculation was not an explanation for the 2003–2008 oil price surge, driven by flow oil demand shocks, while the speculation-demand shock played an important role in the years 1979, 1986, and 1990. Second, Kilian & Murphy (2014) distinguish and provide estimates of the short-run oil demand elasticity for use and production. Meanwhile, the absolute value of demand elasticity for use is smaller than that for production due to the cushion of oil inventories for use, but both of them are significantly larger than zero. Consequently, speculation will always accompany a change in the oil inventory. In other words, there is no evidence

[1] It is necessary to note that this does not mean the weakness of the standard MH algorithm biases Baumeister & Hamilton's (2015, 2018, 2019) Bayesian inferences, because their identifications are only applied to structural parameters directly, whose prior distributions are from Student t which is clearly unimodal. However, C-BSVARs consider a more complex set identification, such as the lower bound of oil demand elasticity calculated jointly through IRFs at the 0-horizon. Hence the shape of its prior or posterior is of ambiguity and not necessarily a perfect unimodal distribution.

[2] It is clear that C-BSVARs do not apply to the traditional orthogonal reduced-form parameterisation, but do use Bayesian method for structural inferences straightly.

for oil price endogeneity.[3]

Nevertheless, regarding the computational burden, only 150 of posterior models were deemed admissible for their global crude oil market SVARs (Kilian & Murphy, 2010), which is statistically unreliable.[4] The size of admissible models is of importance to get a precise inference. Baumeister & Hamilton (2017, p. 21-22), for example, claimed that the results in Kilian & Murphy (2014) are sensitive to the choice of seeds for replicating random number generation process. More specifically, I illustrate how the sample size influences the precision of the estimates' using Monte Carlo sampling of a standard normal ($N(0,1)$) distribution. Figure 4.1 plots the 'uncertainties' (left y-axis) of the 16^{th}, 50^{th}, and 84^{th} percentiles of a standard normal. The 'uncertainties' are presented as Kernel fitted probability distributions based on 1,000 'estimates' of the three percentiles. The 'estimates' of the three percentiles, using standard Normal Monte Carlo sampling, use sample sizes of 100, 1,000, 10,000, 100,000. It is clear that Monte Carlo sample size matters for the precision of Bayesian inference, as increasing sample size narrows the uncertainties of percentiles.

In this chapter, using two specifications in the world crude oil market, I illustrate how C-BSVARs provide a practical solution to this problem enabling larger numbers of posterior draws to be used for economic inference. The first specification utilises the same identification as the oil-market structural model in Kilian & Murphy (2014), while the second allows for uncertainty in the short-run crude oil demand elasticity lower bound for oil use, in contrast to Kilian & Murphy (2014) who fix the lower bound at -0.8 (see Kilian & Murphy, 2014, p. 462).

In the first specification, aiming to replicate Kilian & Murphy's (2014) identification, I restrict C-BSVARs using parallel MH samplers, and use one long Markov chain for the estimates. The effect of using C-BSVARs is to sharpen the distribu-

[3]The previous two chapters emphasis the importance of the time variation in parameters (TVP) and stochastic volatility (SV), while Kilian & Murphy's (2014) structural model is homoscedastic and has constant coefficients. For the oil-market structural analysis, there are strong empirical evidences showing that the SVAR, under suitable identification, is still valid producing economic insights, see among others, Baumeister & Kilian (2016c), Baumeister & Kilian (2016a), Baumeister & Kilian (2016b), and Baumeister & Hamilton (2019). Considering heteroskedastic SVARs with TVP are of interest for the oil market structure modelling, but it would be far more complex in both identification and computation than C-BSVARs given a large dimensional model. Therefore, I leave this for the further research.

[4]Kilian & Murphy (2010, p. 34) indicate the size of their admissible models, which is the working paper version of Kilian & Murphy (2014).

tion of impulse response functions, whilst confirming that Kilian & Murphy's (2014) economic findings, including relative importance of shocks accounting for the surge of the real oil price. Moreover, I broadly confirm Kilian & Murphy's (2014) result that the difference in posterior distributions between the elasticities for use and production, as well as the size being significantly larger than zero.

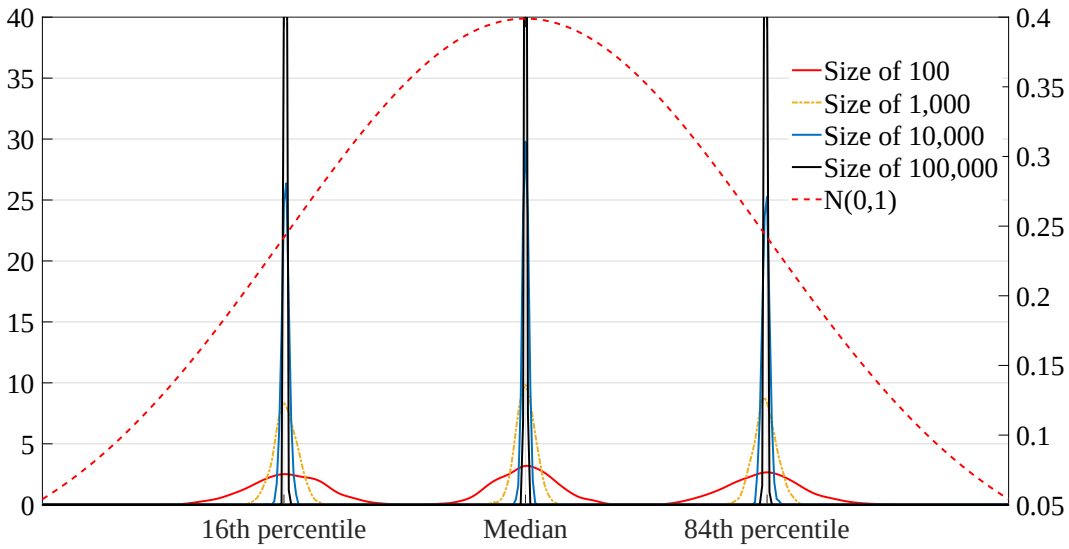

Figure 4.1: Kernel fitted probability distribution functions

Note: The kernel fitted probability distribution functions, which use the Normal method and the bandwidth that optimal for estimating Normal densities, are the percentiles generated from a standard normal distribution with respected to the size of Monte Carlo sampling, namely, 100, 1,000, 10,000, and 100,000, are measured by *left y-axis*. The dotted red line is a standard normal distribution (*right y-axis*).

However, this raises two concerns. First, the multimodal posterior distributions of oil demand elasticity, estimated by C-BSVARs, are diffuse lying within the truncated bounds of [-0.8, 0], resulting in a uniform-like distribution (using MATLAB seed 316). Therefore, it is difficult to draw a conclusion on the price endogeneity using the estimates of the 68% confidence sets. Second, the elasticities are sensitive to the choice of seed for replicating random numbers. This is problematic as an accurate inference of crude oil demand elasticities from a structural modelling perspective, as their magnitude reflects not only the existence of crude oil price endogeneity (Kilian & Murphy, 2014), but has implications for the relative importance of supply and demand factors in explaining oil market fluctuations (Caldara et al.,

2019).

There are two possible reasons for the sensitivity to the choice of seeds: (1) the MCMC algorithm has failed to converge; (2) the latent distribution is not perfectly unimodal. If it is caused by reason (1), we simply increase the sample size for MCMC convergence. However, if it is caused by reason (2), as documented in Waggoner et al. (2016), due to the weakness of MH sampler that the posterior will be sampled within a local maximisation, it is to my knowledge impossible to find a global maximisation using finite MCMC iterations. As a result, simply increasing the sample size would not be an appropriate idea, as there is no method can diagnose an efficient sample size. However, it is possible to jointly use Bayesian convergence diagnostics to give a rough guide for specifying which one of the two reasons causes the sensitivity. (See reviews and examples in Cowles & Carlin (1996) and Roy (2019).) Section 4.4.2 illustrates the sensitivity to the choice of seeds is highly likely to be caused by reason (2).

To solve the limitations above, the second specification of C-BSVARs proposes an additional restriction on the 'short-run crude oil demand elasticity lower bound for use', which is randomly sampled from a prior distribution calibrated according to a large range of estimates for gasoline demand elasticities taken from the literature. The rationale underlying is that the demand elasticity of crude oil should be smaller than that for gasoline, as it is a product of crude oil. This additional restriction enables C-BSVARs to use parallel MH samplers for the estimation, where the Markov chains start with different values. As a result, the estimates of the elasticities are diagnosed as converged using all the tests, which are not sensitive to the seeds in this case. The inference of oil demand elasticities via C-BSVARs have narrower distributions and are closer to zero than the estimates calculated where the restriction is not imposed. However, they are still larger than the estimates in literature, such as -0.06 reported in Hamilton (2008), -0.05 in Dahl (1993) and -0.07 in Cooper (2003). Hence, C-BSVARs confirm Kilian & Murphy's (2014) conclusion that there is no crude oil price endogeneity. Moreover, I illustrate that the economic findings in Kilian & Murphy (2014) are similar.

Although the sensitivity of Kilian & Murphy's (2014) inferences to the choice of seeds may be mitigated by increasing the number of admissible posteriors to 100,000, this is impractical using their approach. Applying the C-BSVARs approach makes this practicable, and inference is not sensitive to the choice of seeds. To my

knowledge, there is no econometric method in the literature which can implement the range of restrictions considered in this chapter. Finally, Caldara et al. (2019) and Herrera & Rangaraju (2018) claim that the relative importance of supply and demand shocks in terms of driving the real oil price is dominated by the restriction of the oil supply elasticity. However, this exercise somewhat rules out the former's argument through a comparison of different upper bounds of the oil supply elasticity.

The remainder of this chapter is structured as follows. After the literature review in the following section, Section 4.3 introduces the C-BSVARs with the MH sampler. Under the same section, I formulate a corresponding impulse response function and a historical decomposition. Then, two specifications of the world crude oil market will be illustrated in Section 4.4. Finally, Section 4.5 concludes this chapter.

4.2 Literature review

This section briefly reviews how SVARs have been used to make structural inferences in the macro-econometrics literature.[5] There are two main approaches for solving SVARs. First as described by Uhlig (2005), SVARs are written as the product of the reduced-form parameters and the set of orthogonal matrices, used to derive the so-called 'orthogonal reduced-form parameterisation'. Second as described in Baumeister & Hamilton (2015, 2018, 2019), the structural parameters are sampled directly using posterior inferences.

The traditional orthogonal reduced-form parameterisation typically uses a Normal–inverse–Wishart–Uniform (NiWU) prior providing for a simpler analysis of the posterior distribution (see Uhlig, 2005; Rubio-Ramirez et al., 2010; Kilian & Murphy, 2014; Arias et al., 2018; and Amir-Ahmadi & Drautzburg, 2019).[6] These algorithms draw from a conjugate posterior distribution over the orthogonal reduced-form parameterisation. The method only accepts draws when the restrictions hold, and then transforms the accepted draws into the structural parameterisation using the orthogonal matrix. For example, Arias et al. (2018) propose an importance sampler for orthogonal matrices, which imposes zero restrictions on the structural

[5] This section reviews SVARs literature after Uhlig (2005), see Kilian & Lütkepohl (2017) for a more detailed review.

[6] Inoue & Kilian (2013) critique the way these methods present the inference of estimates, such as the median of the estimates, and they propose a diagnostic test indicating the most likely admissible model.

parameters. Arias et al. (2019) employ Arias et al.'s (2018) method to analyse the systematic component of monetary policy, and Antolín-Díaz & Rubio-Ramírez (2018) utilise the same method to emphasise the narrative sign restrictions on historical decomposition applied in Kilian & Murphy's (2012) three-variable oil market model. Arias et al. (2018) claim that their choice of conjugate density over the set of orthogonal matrices conditional on the reduced-form parameters is uniform. However, Baumeister & Hamilton (2015) emphasise that the method based on NiWU cannot transform the orthogonal matrices to a uniform distribution over the identified set of the structural parameters, as also observed in Amir-Ahmadi & Drautzburg (2019).

The inefficiency of this accept–reject method for SVARs has been documented in the applied econometrics literature. Amir-Ahmadi & Drautzburg (2019), for an example, develop the orthogonal reduced-form parameterisation, aiming to address the inefficiency, as evidenced by a diagnostic test for whether the identified set is empty, which avoids sampling orthogonal matrices based on the unrestricted models that cannot be rotated to an admissible set. Responding to the criticism from Baumeister & Hamilton (2015) on uniformity, Amir-Ahmadi & Drautzburg (2019) also propose a prior-robust algorithm to sample the posterior bounds of the identified set for acceptable reduced models. Bruns & Piffer (2019), in another example to improve the efficiency of orthogonal reduced-form parameterisation, allow for prior flexibility as in Baumeister & Hamilton (2015) at no additional computational cost, then employ the importance sampler using the distribution proposed by NiWU in Rubio-Ramirez et al. (2010) for structural reparameterisation.

The methods, nonetheless, are still constrained by the natural accept–reject ratio, conditional on the strength of identifying restrictions, or what Amir-Ahmadi & Drautzburg (2019) refer to as ranking restrictions. In the literature, the restrictions are based on evidence from micro data, theory (e.g. elasticities) and dynamics. The more stringent the restrictions are leads the algorithm to become increasingly inefficient when accepting the unrestricted models. For example, the algorithm's structural reparametrisation in Amir-Ahmadi & Drautzburg (2019) is conditional on only 500 unrestricted estimations, whose identified set is not empty based on their diagonal test. Nevertheless, a 500-size is not statistically reliable. Moreover, Amir-Ahmadi & Drautzburg (2019) retain any unrestricted model whose parameters' maximum absolute eigenvalue is < 1.03 without a detailed explanation for the choice. However, for a stable VAR system, the corresponding eigenvalue should be

< 1 (see Hamilton, 1994, p. 260–311). Further, exact zero restrictions, identifying contemporary IRFs equal to zero, are not allowed in Amir-Ahmadi & Drautzburg (2019), despite their use of soft zero restrictions as a substitute for the exact zero restriction. As for the example in Bruns & Piffer (2019), the proposal density of the importance sampler is NiWU, and hence, its efficiency is limited by the strength of restrictions imposed.

Some recent papers in this literature have focused instead on drawing structural parameters directly through Bayesian methods, for example see Baumeister & Hamilton (2015, 2018, 2019) and Waggoner et al. (2016). Using the structural parameterisation can not only define priors' densities directly through economically interpretable structural parameters, but also this is not limited by the natural accept–reject ratio when reparametrising the unrestricted model for structural interpretation using orthogonal matrices. A random-walk MH sampler is utilised in Baumeister & Hamilton (2015, 2018, 2019). Considering that the MH sampler is limited by only using sample posterior draws within one local peak when the posterior is not a perfectly unimodal distribution, Waggoner et al. (2016) propose a sequential approach for drawing high-dimensional posteriors, which they call 'the striated MH sampler'. The multimodal-posterior assumption also used in this exercise, motivated by the fact that parameters (elasticities) of interest is non-linear.

However, Kilian & Zhou (2018) criticise the method of sampling the structural parameters directly as it is unable to use the restrictions after the impact, and utilises non-linear restrictions. Moreover, directly drawing structural parameters requires a set pattern of corresponding endogenous variables. For example, due to the definition of an elasticity relating to percentage changes, the endogenous variables should be formed as the change in log levels. In other words, the model not only changes the method for imposing restrictions, but also the transformations of the variables. Hence, it is hard to draw comparisons of the new model with respect to the existing literature. Further, the striated MH sampler in Waggoner et al. (2016) is not only subject to the criticisms above, but is also inefficient (Bruns & Piffer, 2019). This is because the sampler transforms the posterior distribution by tempering the likelihood from sequential stages. Each stage runs a standard random-walk Metropolis algorithm with a tempered posterior kernel, and the starting value does not come from the maximization, but from a proposal draw at the previous stage using importance weights.

In contrast to the existing literature, this chapter finds an admissible solution numerically, and sampling the structural parameters undertaken a random-walk MH algorithm starting with the admissible solution. Inferences from C-BSVARs are based on parallelly calculation of the above process for addressing the multimodal-posterior assumption. Following sections will detail the method, and using two specifications in oil market modelling to illustrate that the approach can address both the efficiency and implementations of complex restrictions used in SVARs.

4.3 The constrained Bayesian structural vector autoregression

A structural VAR model can be written as:

$$Ay_t = BX_{t-1} + u_t, \tag{4.1}$$

where y_t is a $(n \times 1)$ vector, and X_{t-1} is defined as $[1, y_{t-1}^{'}, ..., y_{t-p}^{'}]^{'}$, in which p is the VAR lag length. Hence, X_{t-1} is a $(n \times p + 1)$ vector, and A is a matrix that contains all structurally identified restrictions, being referred to as an identification matrix. The parameter matrix capturing the lagged structural coefficients, B, is a $(n, n \times p + 1)$ matrix, u_t is a $(n \times 1)$ vector of structural disturbances distributed as $u_t \sim N(0, \Sigma_u)$, where the corresponding variance matrix (denoted Σ_u) is diagonal.

Traditional accept-reject methods for SVARs have to sacrifice efficiency to ensure the effectiveness of the identifications. Estimation employing traditional methods utilises two steps: (1) Estimate the corresponding reduced-form model:

$$\begin{aligned} y_t &= \Phi X_{t-1} + \epsilon_t \\ &= C + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t, \end{aligned} \tag{4.2}$$

where

$$\Phi = A^{-1}B,$$

$$\epsilon_t = A^{-1}u_t, \quad \epsilon_t \sim N(0, \Sigma_\epsilon),$$

and $\Sigma_\epsilon = A^{-1}\Sigma_u(A^{-1})^{'}$, through the conventional maximum likelihood/least squares estimator. (2) Utilise the Cholesky decomposition for A^{-1}, which is then rotated via

an orthogonal matrix sampled randomly or from NiWU to identify the admissible models, utilising the estimated covariance matrix Σ_ϵ (see Kilian & Lütkepohl (2017) for a review). In the literature, this method is referred to as the orthogonal reduced-form parametrisation. As described in the literature review, however, as the degree of stringency, complexity and number of restrictions implied by the microeconomic literature increases, the probability of accepting a model as admissible from the step (2) decreases sharply.

Differently, Baumeister & Hamilton (2015, 2018, 2019) outline a Bayesian method, which considers the uncertainty underlying identification in advance. They identify the model using the A matrix as opposed to the A^{-1} in Equation (4.1), while the matrices B and Σ_u are conditional on A. Waggoner et al. (2016) also identify the A matrix directly. However, there are two main shortcomings to reproduce the SVARs of orthogonal reduced-form parameterisation using their Bayesian SVARs, as mentioned in the literature review and described below.

First, the method in Baumeister & Hamilton (2015, 2018, 2019) is limited to the complicated restrictions following the impact of shocks, which are based on the impulse response functions. A joint set of sign restrictions, in Kilian & Murphy (2014) for example, are such that responses of the oil production and the global real activity to an unanticipated flow supply disruption are negative for the first 12 months, while the response of the real price of oil is positive. The joint set is necessary because the restriction rules out specifications in which unanticipated flow supply disruptions lead to a decline in the real price of oil below its starting level, which would be at odds with the conventional views of the effects of unanticipated oil supply disruptions (Kilian & Murphy, 2014). It is impossible to identify the joint set with the A matrix directly, despite the influence of identified priors does not vanish asymptotically under the system of Baumeister & Hamilton (2015).

Second, identifying the A matrix from the microeconomic literature requires a certain formation of endogenous variables, such as the first difference of log level. More specifically, Baumeister & Hamilton's (2019) oil market structural model utilises the first difference of log-level real crude oil price as one of the endogenous variables, but not the log-level employed in the crude oil literature (e.g. Kilian, 2009; Kilian and Murphy, 2012; 2014). The reason being that Baumeister & Hamilton (2019) identify the short-run crude oil supply and demand elasticities in A, where the definition of elasticities is the percentage change of supply or demand

caused by the percentage change in the price.

In contrast to Baumeister & Hamilton (2015, 2018, 2019), C-BSVARs identify A^{-1} with positively or negatively truncated Student t distributions for sign restrictions. A closed-form likelihood function based on A (conditional on A^{-1} in C-BSVARs) is demonstrated in Baumeister & Hamilton (2015). Aiming for a more precise estimation, identification beyond the signs, such as restrictions after imposing the shocks, namely dynamic sign restrictions, are proposed in structural analysis literature. C-BSAVRs enable these broader identifications beyond Baumeister & Hamilton's (2015) sign restrictions through searching to maximise the likelihood function using all the constraints. Additionally, I add two more constraints for all specifications using C-BSVARs. I measure the performance using the log-level marginal likelihood (denoted as LogML). (1) The posterior draws have higher or equal LogML relative to the estimates of a reduced-form model (Equation 4.2) using the maximum likelihood/least squares estimator. And (2) I require a stable VAR system, avoiding explosive impulse responses.

The motivation for the additional constraints described in (1) is to rule out estimates that are not consistent with the data. The traditional method for structural interpretation is based on the orthogonal rotation of unrestricted VAR, which has been referred to as being 'observationally equivalent', (e.g. Kilian & Murphy, 2014; Amisano & Giannini, 2012; Kilian & Lütkepohl, 2017; and Arias et al., 2018). Differently, C-BSVARs samples A^{-1} directly but is not orthogonally rotated from unrestricted model. The potential sample of A^{-1}, whose LogML is less than the unrestricted VAR's, is the one to avoid. The marginal empirical fit of a model in this chapter is measured through LogML, which is widely utilised for measuring the models' performance in the out-of-sample forecasting literature (e.g. Garratt et al., 2011; Garratt et al., 2014; Chan, Eisenstat, et al., 2018; Mitchell & Wallis, 2011), and is referred to as the logarithmic score, which is usual justified from the perspective of Kullback-Leibler information criterion.

To resolve the constrained maximisation problem, I propose an MH algorithm to enable estimation of the parameters in A^{-1} through the following three steps: (I), estimate the reduced-form and calculate its LogML for the first restriction; (II), numerically determine a starting point that satisfies all restrictions; and (III), use a random-walk MH process to estimate A^{-1}. Accordingly, the rest of this section details the three steps of the estimator in three subsections.

4.3.1 Step I: estimation of the reduced-form model and measuring its empirical fit

C-BSVARs use Baumeister & Hamilton's (2015) MCMC chain, but identifies A^{-1} for signs and considers identifications beyond signs of the contemporary correlation matrix, such as dynamic sign restrictions. I add one additional restriction, that the LogML of each accepted MCMC draw using C-BSVARs is equal or higher than the LogML of unrestricted VARs, estimated via least squares estimator, which is equal to the LogML for all the models deemed as admissible using the traditional accept-reject methods.[7] The estimation of a reduced-form model creates a benchmark measure of the estimated fit with respect to the data (Kilian & Murphy, 2014).

Since the shocks in Equation (4.2) are assumed to be locally Gaussian at each time $t = [1, \ldots, T]$, the log-level marginal likelihood of the reduced-form model ($LogML_{obs}$) is calculated as:

$$LogML_{obs} = \sum_{t=1}^{T} log(f(y_t^{'}, (\Phi X_{t-1})^{'}, \Sigma_{\epsilon})), \tag{4.3}$$

where the joint probability, $f(y_t^{'}, (\Phi X_{t-1})^{'}, \Sigma_{\epsilon})$, is defined as

$$f(y_t^{'}, (\Phi X_{t-1})^{'}, \Sigma_{\epsilon}) = \frac{1}{\sqrt{|\Sigma_{\epsilon}|(2\pi)^n}} e^{-\frac{1}{2}(y_t^{'}-(\Phi X_{t-1})^{'})\Sigma_{\epsilon}^{-1}(y_t^{'}-(\Phi X_{t-1})^{'})^{'}}. \tag{4.4}$$

Additionally, the estimation of the log-level marginal likelihood in C-BSVARs is an alternative criterion for specifying the most likely admissible model relative to the diagnostic test proposed in Inoue & Kilian (2013).

[7] The estimation of the reduced-form model can utilise either the maximum likelihood/least squares estimator, or the Bayesian method defined in Giannone et al. (2015). In this exercise, I employ the least squares estimator, following Kilian & Murphy (2014).

4.3.2 Step II: numerically determine the starting point for the MH algorithm

This subsection proposes a non-linear arg-minimisation method, based on a closed log likelihood function, to identify the starting point for the constrained maximization issue with the MH algorithm. Initially, instruction is presented for mapping the identification matrix A^{-1} to A. Then, I briefly introduce the closed-form likelihood conditional on the A matrix, as proposed by Baumeister & Hamilton (2015, 2018, 2019). Finally, an arg-minimisation problem is illustrated.

The structural prior information is given in matrix A^{-1}, represented from a Bayesian perspective, in the form of a density $p(A^{-1})$. The analytical impulse response functions are derived via its reduced-form moving average formation, and the conventional sign restrictions are identified through the impulse response functions at horizon zero. Continuing with the model of Equation (4.2), the $(n \times n)$ non-orthogonalised impulse–response matrix at horizon s is:

$$\begin{aligned} \Psi_s &= \frac{\partial y_{t+s}}{\partial \epsilon_t^{'}} \\ &= jF^s j^{'}, \end{aligned} \tag{4.5}$$

where j and F_t are:

$$j = [I_n \quad \mathbf{0}_{(p-1)\times n}],$$

and

$$F = \begin{bmatrix} \Phi_1 \dots \Phi_{p-1} & \Phi_p \\ I_{n\times(p-1)} & 0_{n(p-1)\times n} \end{bmatrix}.$$

Hence, the dynamic effects of structural shocks at horizon s are given by

$$\begin{aligned} H_s &= \frac{\partial y_{t+s}}{\partial u_t^{'}} \\ &= \Psi_s A^{-1}, \end{aligned} \tag{4.6}$$

based on the chain rule since $\epsilon_t = A^{-1}u_t$, (see Hamilton, 1994, p. 260–311). As $u_t \sim N(0, \Sigma_u)$, where Σ_u is diagonal, the dynamic effect of standardised shocks $e_t \sim N(0, I_n)$ is

$$\begin{aligned} H_s &= \frac{\partial y_{t+s}}{\partial e_t^{'}} \\ &= \Psi_s A^{-1} chol(\Sigma_u)'. \end{aligned} \tag{4.7}$$

Then the dynamic effects of the structural shocks can be decomposed through:

$$\begin{aligned} y_t &= \hat{y}_{t|t-r} + \sum_{s=0}^{r-1} H_s e_{t-s} \\ &= \hat{y}_{t|t-r} + \sum_{s=0}^{r-1} \Psi_s A^{-1} chol(\Sigma_u)' e_{t-s}, \end{aligned} \tag{4.8}$$

where $\hat{y}_{t|t-r}$ is the r-period-ahead forecast using Equation (4.2). The standard error e_t can be computed as $e_t = (chol(\Sigma_u)')^{-1} A y_t - (chol(\Sigma_u)')^{-1} B X_{t-1}$.

When $s = 0$, the sign of impulse response is defined by the non-diagonal elements in A^{-1} from Equation (4.7), because Ψ_0 is identical and the variance matrix Σ_u is positive definite. Additionally, the VAR system is considered to be stable when the maximum absolute eigenvalue denoted as ϱ^{max} for F, is strictly smaller than 1 (see Hamilton, 1994, p. 260–311).

To ensure that the A (or A^{-1}) matrix is invertible, it is necessary to establish a prior belief of $h_1 = det(A) > 0$, as presented in the asymmetric t distribution, where $p(h_1) = k_1 \sigma_1^{-1} \widetilde{\phi}_{\nu_1}((h_1 - \mu_1)/\sigma_1) \Phi(\lambda_1 h_1/\sigma_1)$, introduced by Baumeister & Hamilton (2018, 2019). $\widetilde{\phi}_{\nu_1}((h_1 - \mu_1)/\sigma_1)$ is a probability density function of a standard Student t with ν_1 degree of freedom evaluated at the point $(h_1 - \mu_1)/\sigma_1$, while $\Phi(\lambda_1 h_1/\sigma_1)$ is the cumulative distribution function for a standard normal, and k_1 is a constant for integrating the density to unity. The parameter λ_1 determines how strongly the distribution of h_1 skews to positive. Under the set $\mu_1 = 0.6$, $\sigma_1 = 1.6$, $\nu_1 = 3$, and $\lambda_1 = 2$, the prior probability of $h_1 > 0$ is 91.2% from the simulation practice.

More plausible a prior are larger values for $p(A^{-1})$, whilst $p(A^{-1}) = 0$ for any values of A^{-1} are completely ruled out. Information may pertain to individual elements of A^{-1} or to non-linear combinations such as specified elements of the impulse response functions $\Psi_s A^{-1} chol(\Sigma_u)'$, which represent the equilibrium effects of structural shocks. The application in this chapter draw on both sources of prior information. Implementation of the procedure requires only that $p(A^{-1})$ be a proper density that integrates to unity, where any distribution can be utilised to present the prior information in A^{-1}.

Prior information regarding the other parameters can be represented by par-

ticular families of parametric distributions that enable many features of the Bayesian posterior distribution to be calculated with closed-form analytic expressions, such as in Baumeister & Hamilton (2015, 2018, 2019). Specifically, prior information regarding Σ_u conditional on A can be represented using gamma distribution $\Gamma(\kappa_i, \tau_i)$:

$$p(\Sigma_u|A) = \prod_{i=1}^{n} p(\sigma_{ii}|A), \tag{4.9}$$

$$p(\sigma_{ii}|A) = \begin{cases} \frac{\tau_i^{\kappa_i}}{\Gamma(\kappa_i)}(\sigma_{ii}^{-1})^{\kappa_i - 1} exp(-\tau_i \sigma_{ii}^{-1}) & \text{for } \sigma_{ii}^{-1} \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where σ_{ii} is the row i and column i of Σ_u. Since this is a gamma distribution, the mean and variance of σ_{ii} are κ_i/τ_i and κ_i/τ_i^2, respectively. Theoretically, both κ_i and τ_i can be conditional on A. However, while following Baumeister & Hamilton (2019), I only allow τ_i conditional on A, which determines the tightness of prior beliefs. (For further details on the choice of κ_i and τ_i, see Appendix C.1 or Baumeister & Hamilton, 2019.) The conditional Gaussian distribution of $B|A, \Sigma_u$ is:

$$p(B|A, \Sigma_u) = \prod_{i=1}^{n} p(b_i|A, \Sigma_u), \tag{4.10}$$

where $p(b_i|A, \Sigma_u) \sim N(m_i, \sigma_{ii} M_i)$, and the vector m_i denotes the best possible prediction for the value of $b_i^{'}$ prior to appraising the data, where $b_i^{'}$ denotes row i of B, that is, b_i contains the lagged coefficients for the ith structural equation, while the matrix M_i characterises the confidence in this prior information. A large variance would represent high uncertainty, while having no useful prior information could be regarded as the limiting case when M_i^{-1} returns as zero. The application in Baumeister & Hamilton (2019) allows m_i to depend on A, but assumes that M_i does not. (For more details on the specifications of m_i and M_i, see Appendix A of Baumeister & Hamilton, 2019.)

Then, the overall prior is:

$$p(A, B, \Sigma_u) = P(A) \prod_{i=1}^{n} [p(\sigma_{ii}|A) p(b_i|A, \Sigma_u)]. \tag{4.11}$$

With Gaussian residuals, the likelihood function in closed-form is given by:

$$p(Y_t|A,B,\Sigma_u) =$$
$$(2\pi)^{-Tn/2}|det(A)|^T|\Sigma_u|^{-T/2} \times exp[-(1/2)\sum_{t=1}^{T}(Ay_t - Bx_{t-1})'\Sigma_u^{-1}(Ay_t - Bx_{t-1})], \tag{4.12}$$

where $|det(A)|$ denotes the absolute value of the determinant of A.

Starting with the conjugate prior, the posterior distribution is:

$$p(A,B,\Sigma_u|Y_t) = p(A|Y_t)p(\Sigma_u|A,Y_t)p(B|A,\Sigma_u,Y_t), \tag{4.13}$$

which summarises the researcher's uncertainty regarding the parameters. (For further details of the posterior sampling of $p(A|Y_t)$, $p(\Sigma_u|A,Y_t)$, and $p(B|A,\Sigma_u,Y_t)$, see Appendix C.2 or Baumeister & Hamilton 2015; 2018; 2019.)

The structural inference is conditional on the reduced-form estimation in Equation (4.2). This differs from the random sampling of the orthogonal rotation matrix for observationally equivalent structural analysis. In this chapter, I numerically determine a conditional local minimisation, aiming to identify the starting value for the MH sampler:

$$\begin{aligned}
\min_{\boldsymbol{\phi}} \quad & -(log(p(Y_t|\boldsymbol{\phi})) + log(p(A_1^{-1})|\boldsymbol{\phi})) \times g(\Theta) \\
\text{s.t.} \quad & log(f(y_t, A_1^{-1}BX_{t-1}, (A_1'\Sigma_u A_1)^{-1})) \geq LogML_{obs} \\
& \varrho^{max} < 1 \\
& \underline{\boldsymbol{\phi}} < \boldsymbol{\phi} < \overline{\boldsymbol{\phi}} \\
& \text{A set of structural } \textit{identifications} \text{ depending on the application,}
\end{aligned} \tag{4.14}$$

where $\boldsymbol{\phi}$ is a vector that contains all the elements in A_1^{-1}, which are the initial values of A^{-1} for the MH process, as well as one hyper parameter for the tightness of the prior density of the largest uncertain element in A_1^{-1}. The inclusion of the hyper parameter decides the relative tightness for priors in general. The rationale of including it is that the model is sensitive to the relative tightness for prior densities (e.g. Bańbura et al., 2010; Giannone et al., 2015; and Chan, Jacobi, & Zhu,

2018). It is also demonstrated in practice that including the measure of a relative tightness helps to solve the arg-minimisation and can accelerate the computational speed. Hence, it is straightforward that the probability of $p(A^{-1})$ is conditional on $\boldsymbol{\phi}$.

The ϱ^{max} is the maximum absolute eigenvalue, and $\varrho_t^{max} < 1$ indicates the VAR system's stationarity. The restriction $log(f(y_t, A_1^{-1}BX_{t-1}, (A_1'\Sigma_u A_1)^{-1})) \geq LogML_{obs}$ results in the structural formation locally and observationally having a higher or equal marginal likelihood than the reduced form. The upper and lower bounds of structural identifications are given in vectors $\underline{\boldsymbol{\phi}}$ and $\overline{\boldsymbol{\phi}}$ with the same length of vector $\boldsymbol{\phi}$, respectively. The inclusion of these bounds avoids searching the minimisation from infinities. The last restriction is with respect to the equilibrium effects of structural shocks, such as the dynamic sign restrictions. The vector Θ contains all indicators (θ_i) signalling that the i^{th} system restriction is not satisfied, denoted as 1, and 0 otherwise. The indicator $g(\Theta)$ is defined as:

$$g(\Theta) = \begin{cases} 1 & \text{if } \sum_i \theta_i > 0 \\ 0 & \text{if otherwise.} \end{cases} \tag{4.15}$$

Furthermore, the constrained posterior densities, where $\sum_i \theta_i = 0$, are denoted as $p^R(A|Y_t)$, $p^R(\Sigma_u|A, Y_t)$, and $p^R(B|A, \Sigma_u, Y_t)$. The $p^R(A^{-1})$ denotes the prior density of A^{-1} restricted by $\sum_i \theta_i = 0$.

4.3.3 Step III: the MH algorithm

Commencing with the arg-minimisation, a random-walk MH algorithm maximises the restricted likelihood function (4.12), denoted as $p^R(Y_t|A^{-1})$ since A, B and Σ_u are conditional on A^{-1}. The numerical starting values are contained in the vector $\widehat{\boldsymbol{\phi}}$, which provides a guess for the posterior mean of $\boldsymbol{\phi}$. Meanwhile, the numerically calculated matrix of second derivatives may contain the curvature of the posterior distribution:

$$\widehat{\Lambda} = -\frac{\partial^2 q(\boldsymbol{\phi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\bigg|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}},$$

where $q(\boldsymbol{\phi}) = -(log(p(Y_t|\boldsymbol{\phi})) + log(p(\boldsymbol{\phi}))) \times g(\Theta)$.

Thereafter, the guess above is employed to inform a random-walk MH algo-

rithm to generate candidate posterior draws for $vec(A^{-1})$, including all elements in A^{-1}. As a result of draw ι, I have generated values for $vec(A_{\iota+1}^{-1})$:

$$\widetilde{vec}(A_{\iota+1}^{-1}) = vec(A_{\iota}^{-1}) + \xi(\widehat{Q}^{-1})' v_{\iota}, \tag{4.16}$$

where the starting value A_{ι}^{-1} is denoted as the A_1^{-1} contained in $\widehat{\boldsymbol{\phi}}$.[8] The white noise error v_{ι} is a vector that has the same number of elements in A^{-1}, which is generated from a fat-tailed Student t with 2 degrees of freedom. $\widehat{Q}\widehat{Q}' = \widehat{\Lambda}^{A^{-1}}$, where $\widehat{\Lambda}^{A^{-1}}$ is the second derivative with respect to the elements in A^{-1}, with $\widehat{Q}$ lower triangular. However, the numerical minimisation will always stop at $g(\Theta) = 0$. Hence, the second derivative is not always meaningful, in which case $\widehat{Q}$ is set to be an identical matrix.

Algorithm 1 - Structural posterior sampling

Step I: Reduced-form model estimation

- Utilise the least square estimator to estimate Φ and Σ_{ϵ}
- Calculate the $LogML_{obs} = \sum_{t=1}^{T} log(f(y_t', (\Phi X_{t-1})', \Sigma_{\epsilon}))$

Step II: Numerically determine a starting point

$$\min_{\boldsymbol{\phi}} \quad -(log(p(Y_t|\boldsymbol{\phi})) + log(p(A_1^{-1})|\boldsymbol{\phi})) \times g(\Theta)$$

$$\text{s.t.} \quad log(f(y_t, A_1^{-1}BX_{t-1}, (A_1'\Sigma_u A_1)^{-1})) \geq LogML_{obs}$$

- $\varrho^{max} < 1$

  $\underline{\boldsymbol{\phi}} < \boldsymbol{\phi} < \overline{\boldsymbol{\phi}}$

  A set of structural *identifications* depending on an application

Step III: the random-work MH algorithm for $\boldsymbol{A}_{\iota}^{-1}$, where $\boldsymbol{\iota}$=[1, ..., $\boldsymbol{N}_1$, ..., $\boldsymbol{N}_2$]

- Sample $A_{\iota+1}^{-1} \sim p(A_{\iota+1}^{-1}|A_{\iota}^{-1})$, where A_1^{-1} is solved in Step II
- Accept $A_{\iota+1}^{-1}$ substituting A_{ι}^{-1} with probability:
  $$exp[q^R(Y_t|A_{\iota+1}^{-1}) - q^R(Y_t|A_{\iota}^{-1})]$$
- Gather A_{ι}^{-1} constrained by restrictions for ι=[N_1, ..., N_2]

Then, repeat Steps II and III until N_3 posterior draws are saved

The tuning scalar, ξ, is automatically selected, which is based on every 100 random draws, so that approximately 30%–40% of the newly retained $vec(A_{\iota}^{-1})$

[8] In MH process, C-BSVARs only allow the elements in A^{-1} following the random work, but not for the elements of the strength of shrinkage contained in $\boldsymbol{\phi}$. This is because I want to keep the likelihood function consistent over the MH process.

meets the condition through:

$$vec(A_{\iota+1}^{-1}) = \begin{cases} vec(A_{\iota}^{-1}) & \text{with probability } 1 - exp[q^R(Y_t|A_{\iota+1}^{-1}) - q^R(Y_t|A_{\iota}^{-1})] \\ \widetilde{vec}(A_{\iota+1}^{-1}) & \text{otherwise,} \end{cases} \tag{4.17}$$

where I denote $q^R(Y_t|A_{\iota}^{-1}) = log(p^R(Y_t|A_{\iota}^{-1})) + log(p^R(A_{\iota}^{-1}))$. If the acceptance ratio is too high, the algorithm will converge slowly to $P^R(A|Y_t)$. Otherwise, if the acceptance ratio is too low, then $P^R(A|Y_t)$ may arrive to the low-probability region of the true posterior distribution.

In the algorithm, summarised as Algorithm 1, I use N_1 draws for burn-in, and $N_2 - N_1$ draws for gathering the admissible posterior draws in the MH process (Step III). The algorithm parallelly calculates both the II and III steps until N_3 draws, which is the size desired for the posterior draws selected by researchers, are achieved.

4.4 Two world crude oil market specifications

In this section, I illustrate how to apply C-BSVARs with two empirical specifications of the world crude oil structural model. The first specification utilises the world crude oil market model under the exact identification of Kilian & Murphy (2014), while the second imposes 'lower bound uncertainties of the short-run crude oil demand elasticity for use', as an additional restriction.

The multivariate time series utilised in their paper include the U.S. refiners' acquisition cost (RAC) for crude oil imports, the global crude oil production, a business cycle index of global real activity proposed in Kilian (2009), the U.S. consumer price index (CPI), and global crude oil inventories. The data resource is taken from the Journal of Applied Econometrics data archive for Kilian & Murphy (2014) (available at http://qed.econ.queensu.ca/jae/2014-v29.3/kilian-murphy/). The original resources of the data include the U.S. Energy Information Administration, Bloomberg, Drewry Shipping Consultants Ltd., the Federal Reserve Economic Data (FRED) and the Federal Reserve Archival System for Economic Research (FRASER) database of the Federal Reserve Bank of St. Louis.

As in Kilian & Murphy (2014), the endogenous variables $y_t = [\%\Delta prod_t, rea_t, r_t^{oil}, \Delta inv_t]'$ are a 4×1 vector ($n = 4$), where $\%\Delta prod_t$ is the percentage change of oil production times 100; rea_t is the measure of fluctuations in global real activity; r_t^{oil} represents the log-levels of RAC deflated by the U.S. CPI times 100; and Δinv_t is the first difference of oil inventories in levels. The reduced-form model allows for up to two years of lags, $p = 24$, which is consistent with evidence presented in Hamilton & Herrera (2004), Kilian (2009), and Kilian & Murphy (2012, 2014) on the importance of allowing for long lags in the transmission of oil price shocks and in modelling business cycles in commodity markets. (For more detailed motivations of the choice of endogenous variables, see Kilian & Murphy, 2014.)

Subsections 4.4.1–4.4.3 will first introduce the identification used in Kilian & Murphy (2014). Then, empirical results using C-BSVARs for the first specification are presented, which are followed by a discussion of the inference for oil demand elasticities. Finally, I illustrate the results under the second specification when considering the uncertainty of 'the short-run lower bound for oil demand elasticity for use'.

4.4.1 Specification I: identification restrictions in Kilian & Murphy (2014)

The structural VAR model in Kilian & Murphy (2014) is set-identified based on a combination of sign restrictions and bounds on the implied price elasticities of oil demand and supply, motivated by economic theory or extraneous information. There are four sets of restrictions: (1) the impact sign restrictions listed in Table 4.1 below; (2) the bounds on the impact price elasticity of oil supply, where $0 < \eta^{Supply} < 0.0258$; (3) the bounds on the impact price elasticity of oil demand for use, where $-0.8 \leqslant \eta^{O,Use} \leqslant 0$; and (4) dynamic sign restrictions — for the first 12 months, responses of the oil production and the global real activity to an unanticipated flow supply disruption are negative, while the response of the real oil price to the disruption is positive. (For a more detailed discussion on identifications, see Kilian & Murphy, 2014, p. 460–462.)

Table 4.1: Sign restrictions on impact responses in the VAR model

	Flow supply shock	Flow demand shock	Speculative demand shock
Oil production	−	+	+
Real activity	−	+	−
Real price of oil	+	+	+
Inventories			+

Note: All structural shocks have been normalised to imply an increase in the real price of oil. Missing entries indicate that no sign restriction is imposed. (The same table is available in Kilian & Murphy, 2014, p. 461.)

In C-BSVARs, the choice of prior distributions of the elements A^{-1} is selected from asymmetric, truncated or symmetric Student t distributions, as suggested in Baumeister & Hamilton (2019), while the choice is conditional on the aforementioned restrictions (1)–(4). Appendix C.3 provides detailed guidance on how to apply the identifications above to A^{-1} in C-BSVARs.

It is necessary to note that the prior density of oil inventory (the 4^{th} endogenous variable) response to a speculative demand shock (a shock from the 3^{rd} equation), A_{43}^{-1}, is Student $t(2, \sigma^{A_{43}^{-1}}, 3)$, truncated to be positive with a large mode. The positive truncation is due to the sign restriction in Table 4.1 (the last row and column). Here, I allow the uncertainty $\sigma^{A_{43}^{-1}}$ to be estimated through arg-minimisation within a range $[0,\ \overline{\sigma}^{A_{43}^{-1}}]$. This is because (1) the posterior is not sensitive to the location of the prior but rather to its relative shrinkage as discussed in Subsection 4.3.2; and (2) there is no empirical evidence in terms of the size for oil inventories' responses directly from the corresponding microeconomic literature.

Therefore, the prior of structural parameters, considering a positive determined A as discussed in Subsection 4.3.2, becomes:

$$\begin{aligned} p(A^{-1}) \propto & p(A_{21}^{-1})p(A_{31}^{-1})p(A_{41}^{-1})p(A_{12}^{-1})p(A_{32}^{-1})p(A_{42}^{-1})p(A_{13}^{-1})p(A_{23}^{-1})p(A_{43}^{-1})p(A_{14}^{-1}) \\ & p(A_{24}^{-1})p(A_{34}^{-1})p(A_{11}^{-1})p(A_{22}^{-1})p(A_{33}^{-1})p(A_{44}^{-1})p(h_1(A^{-1})). \end{aligned} \tag{4.18}$$

The non-linear restrictions (2–4) are unable to be formulated through A^{-1}. For example, Kilian & Murphy (2014) proposed the elasticity of oil demand for use, $\eta^{O,Use}$ as:

$$\eta^{O,Use} = \frac{1}{t-1}\sum_{i=1}^{t-1} \frac{\frac{Q_{t-1}\times A_{11}^{-1}/100 - A_{41}^{-1}}{Q_{t-1} - \overline{\Delta S}}}{A_{31}^{-1}/100}, \tag{4.19}$$

where Q_{t-1} is the quantity of oil produced in period $t-1$, and $\overline{\Delta S}$ is the observed mean of the change in crude oil inventories (see the derivation in Kilian & Murphy, 2014, p. 477-478), while A_{31}^{-1} and A_{41}^{-1} represent the oil price and inventory responses to a negative flow supply shock. Hence, the additional constraints are proposed in C-BSVARs for arg-minimisation:

$$\begin{aligned}
\min_{\boldsymbol{\phi}} \quad & -\left(log(p(Y_t|\boldsymbol{\phi})) + log(p(A_1^{-1})|\boldsymbol{\phi})\right) \times g(\Theta) \\
\text{s.t.} \quad & log(f(y_t, A_1^{-1}BX_{t-1}, (A_1'\Sigma_u A_1)^{-1})) \geq LogML_{obs} \\
& \varrho^{max} < 1 \\
& \underline{\boldsymbol{\phi}} < \boldsymbol{\phi} < \overline{\boldsymbol{\phi}} \\
& 0 < \eta^{Supply} < 0.0258 \\
& \underline{\eta}^{O,Use} \leqslant \eta^{O,Use} \leqslant 0 \\
& \sum_{i=0}^{s} H_{s,11} < 0, \ \forall \text{s} \leqslant 12 \\
& H_{s,21} < 0, \ \forall \text{s} \leqslant 12 \\
& H_{s,31} > 0, \ \forall \text{s} \leqslant 12,
\end{aligned} \tag{4.20}$$

where $\boldsymbol{\phi} = [A_{21}^{-1}, A_{31}^{-1}, A_{41}^{-1}, A_{12}^{-1}, A_{32}^{-1}, A_{42}^{-1}, A_{13}^{-1}, A_{23}^{-1}, A_{43}^{-1}, A_{14}^{-1}, A_{24}^{-1}, A_{34}^{-1}, A_{11}^{-1}, A_{22}^{-1}, A_{33}^{-1}, A_{44}^{-1}, \sigma^{A_{43}^{-1}}]'$ corresponding $\underline{\boldsymbol{\phi}} = [0, -10, -10, 0, 0, -10, 0, -10, 0, -10, -10, -10, -10, -10, -10, -10, 0]'$, and $\overline{\boldsymbol{\phi}} = [10, 0, 10, 10, 10, 10, 10, 0, 10, 10, 10, 10, 10, 10, 10, 10, \overline{\sigma}^{A_{43}^{-1}}]'$. Since there is no instruction for setting $\sigma^{A_{43}^{-1}}$ in Baumeister & Hamilton (2019), I allow $\sigma^{A_{43}^{-1}}$ and its upper bound $\overline{\sigma}^{A_{43}^{-1}}$ to be 0.8 and 2 for searching respectively. Aiming to numerically find feasibly constrained starting values of $\boldsymbol{\phi}$, $\sigma^{A_{43}^{-1}}$ and $\overline{\sigma}^{A_{43}^{-1}}$ will increase with 0.1 and 0.5 respectively in case that there is no constrained solution.[9] For a consistent MH process, A^{-1} follows a random walk for the posterior, but $\sigma^{A_{43}^{-1}}$ and $\overline{\sigma}^{A_{43}^{-1}}$ are kept constant at their arg-minimisation level. Under the exact Kilian & Murphy's (2014) identification, the lower bound of $\eta^{O,Use}$, denoted as $\underline{\eta}^{O,Use}$, is -0.8.

The responses of the oil production and the economic activity to an unanticipated flow supply disruption are negative for the first 12 months are denoted $\sum_{i=0}^{s} H_{s,11} < 0, \ \forall \text{s} \leqslant 12$ and $H_{s,21} < 0, \ \forall \text{s} \leqslant 12$, respectively. Moreover, the re-

[9] The choice of $\sigma^{A_{43}^{-1}}$ could be any number, whose purpose is calculating an admissible solution from the arg-minimisation (4.20). The reason that $\sigma^{A_{43}^{-1}}$ begins with 0.8 is because A_{43}^{-1} is more uncertain (using a diffuse prior) than the rest of elements in A^{-1} (using a variance of 0.2). (See a detailed discussion in Appendix C.3.)

striction — the responses of the real price to a supply disruption are positive within the first 12 months — is denoted as $H_{s,31} > 0$, $\forall s \leqslant 12$.

Estimation results

I use 1,000 draws for burnin ($N_1 = 1,000$), and 1,000,000 draws for gathering admissible posterior draws ($N_2 = 1,001,000$), until 100,000 admissible draws ($N_3 = 100,000$). The lower-bound of demand elasticity for use is set to -0.8, following Kilian & Murphy (2014), and hence the constraints under the arg-minimisation (4.20) will not change. Since the arg-minimisation (4.20) will always find the same starting value A_1^{-1} under the same constraints, I choose a large N_2 in this specification. In other words, the gathering of 100,000 admissible draws will not parallelly repeat Steps II and III in Algorithm 1. This is also because Kilian & Murphy (2014) have not identified any multimodal distribution assumption on the parameters of interest. Hence, the specification is designed for consistency with Kilian & Murphy (2014), despite the fact that it may be restricted by the standard weakness of the MH sampler that the posterior will be sampled within a local maximisation. To summarise, the results from the C-BSVARs approach confirm Kilian & Murphy's (2014) economic findings, and sharpen the impulse response functions. (See Appendix C.4 for the C-BSVAR results relating to Kilian & Murphy's (2014) economic findings.)

The structural impulse responses to the oil supply and demand shocks are illustrated in Figure 4.2 for the same sample period in Kilian & Murphy (2014), 1973:02–2009:08. The solid black lines are the Bayesian posterior median from the C-BSVARs, and the shaded regions are 68% posterior credible sets. I also draw the estimates utilising Kilian & Murphy's (2014) code available on the aforementioned journal's website for the comparison. The solid red lines are the impulse response estimates for Kilian & Murphy's (2014) model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity amongst the admissible structural models obtained, while the dotted blue lines are the corresponding 68% posterior error bands from Kilian & Murphy's (2014) estimates. Following Kilian & Murphy (2014), all shocks have been normalised to imply an increase in the real price of oil, and the flow supply shock refers to an unanticipated disruption. Primarily, the impulse responses are similar, but the key part is the narrowing of the distributions by utilising C-BSVARs relative to using the method in Kilian & Murphy (2014).

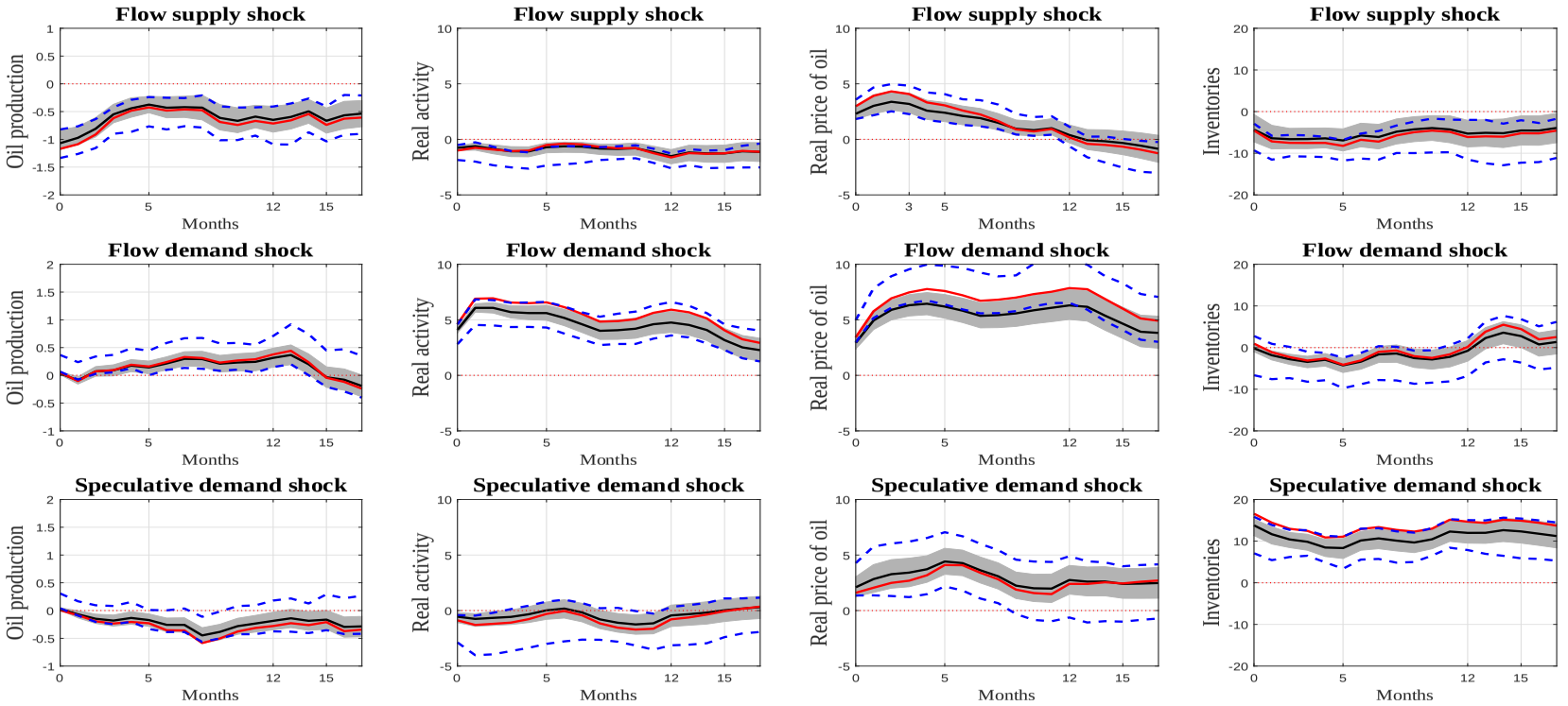

Figure 4.2: Structural impulse responses: 1973:02–2009:08 ($\eta^{Supply} \leq 0.0258$)

Note: *Solid black lines*: Bayesian posterior median; *shaded region*: 68% posterior credible sets; *solid red lines*: the impulse response estimates for the model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity amongst the admissible structural models obtained, conditional on the least-squares estimate of the reduced-form VAR model replicated with Kilian & Murphy's (2014) code; *dotted blue lines*: the corresponding point-wise 68% posterior error bands replicated with the aforementioned code.

For example, Kilian & Murphy's (2014) preferred responses of the oil production to the flow demand shock (red line) are close to the lower bound of its 68% confidence sets (the lower blue dotted lines). Meanwhile, the corresponding 68% posterior credible sets from C-BSAVRs are located in the lower half of Kilian & Murphy's (2014) 68% bands, and the median estimates from C-BSVARs cover the red line. Moreover, the real oil price's response to the flow demand shock is closer to zero when estimated via C-BSVARs than using the method in Kilian & Murphy (2014). However, the responses are still above zero, and persistent. Meanwhile, the most preferred admissible prediction by Kilian & Murphy (2014) (i.e. the red line) are covered by the 68% confidence sets estimated by C-BSVARs.

4.4.2 Discussion: what can we learn about oil demand elasticities from C-BSVARs relative to the oil literature?

As mentioned in Kilian & Murphy (2014), Hamilton's (2009) observation that the speculation drives up oil prices without the increase of oil inventory (the oil price endogeneity) is dependent upon the magnitude of the short-run price elasticity of demand. More specifically, if the oil demand elasticity is significantly larger than zero, the oil price endogeneity will not exist. The conventional estimates of oil demand elasticity are based on a smaller model, ignoring the role of inventories (Kilian & Murphy, 2014). Echoing Kilian & Murphy (2014), the oil consumption explored in this chapter is smoothed through the inclusion of inventories, which is a more appropriate definition for policy questions. Kilian & Murphy (2014) provided estimates of the short-run price elasticities of oil demand and their 68% confidence sets, which were significantly above zero. Consequently, they present no evidence for oil price endogeneity.

The corresponding estimates calculated by C-BSVARs are presented in Figure 4.3, where the black and blue X-tick labels are the 16^{th}, 50^{th}, and 84^{th} quantiles of $\eta^{O,Production}$ and $\eta^{O,Use}$ reported in Kilian & Murphy (2014), respectively. A reliable numeric estimator should not be sensitive to the choice of seeds for the random number generating process. As an illustration, I present the estimates with seed 316 in the upper panel, while the lower panel features seed 613 (in MATLAB). Due to their posterior sample size being only 150, it is impossible to replicate identical estimates of corresponding elasticities utilising C-BSVARs. However, it is clear

that the C-BSVARs broadly confirm that (1) the magnitude of $\eta^{O,Use}$ is smaller than that of $\eta^{O,Production}$, and (2) the boundaries of the 68% confidence sets exclude zero.

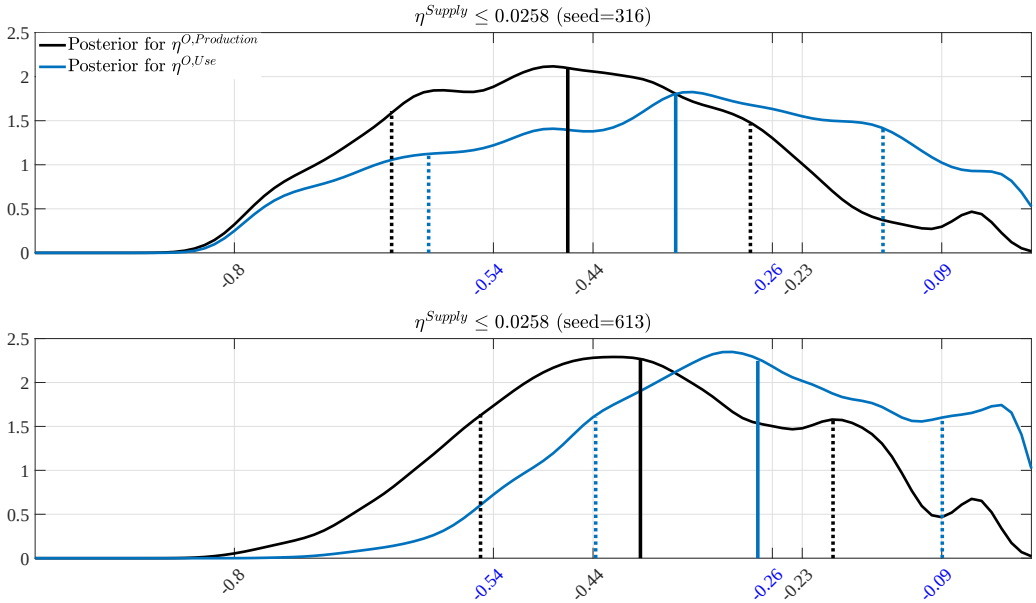

Figure 4.3: Posterior densities of the oil demand elasticity in production and use under a -0.8 lower bound

Note: $\eta^{O,Production}$ refers to the impact price elasticity of oil demand in production, $\eta^{O,Use}$ indicates the average impact price elasticity of oil demand in use, and η^{Supply} indicates the impact price elasticity of the oil supply. *Solid vertical lines*: medians of the posterior; dotted vertical lines: 68% posterior error bands. *Black X-tick labels*: 16^{th}, 50^{th}, and 84^{th} quantiles of $\eta^{O,Production}$, as reported in Kilian & Murphy (2014); *blue X-tick labels*: 16^{th}, 50^{th}, and 84^{th} quantiles of $\eta^{O,Use}$, as reported in Kilian & Murphy (2014). Posterior densities are Kernel fitted, using the Normal method and the bandwidth that optimal for estimating Normal densities.

Nevertheless, two concerns are raised. On the one hand, the distributions of oil demand elasticities, estimated by C-BSVARs, are diffuse within the truncated bounds of [-0.8, 0], which appear as a uniform distribution for $\eta^{O,Use}$ using seed 316. Moreover, there is a peak within the range of [-0.09 0] for $\eta^{O,Production}$. On the other hand, the elasticities are sensitive to the choice of seeds for replicating random numbers. For example, as shown in Figure 4.3, the median of $\eta^{O,Use}$ under seed 316 is -0.36, but for the seed 613 it becomes -0.27. Furthermore, the shape of the two elasticities under seed 613 is narrower than that for the seed 316.

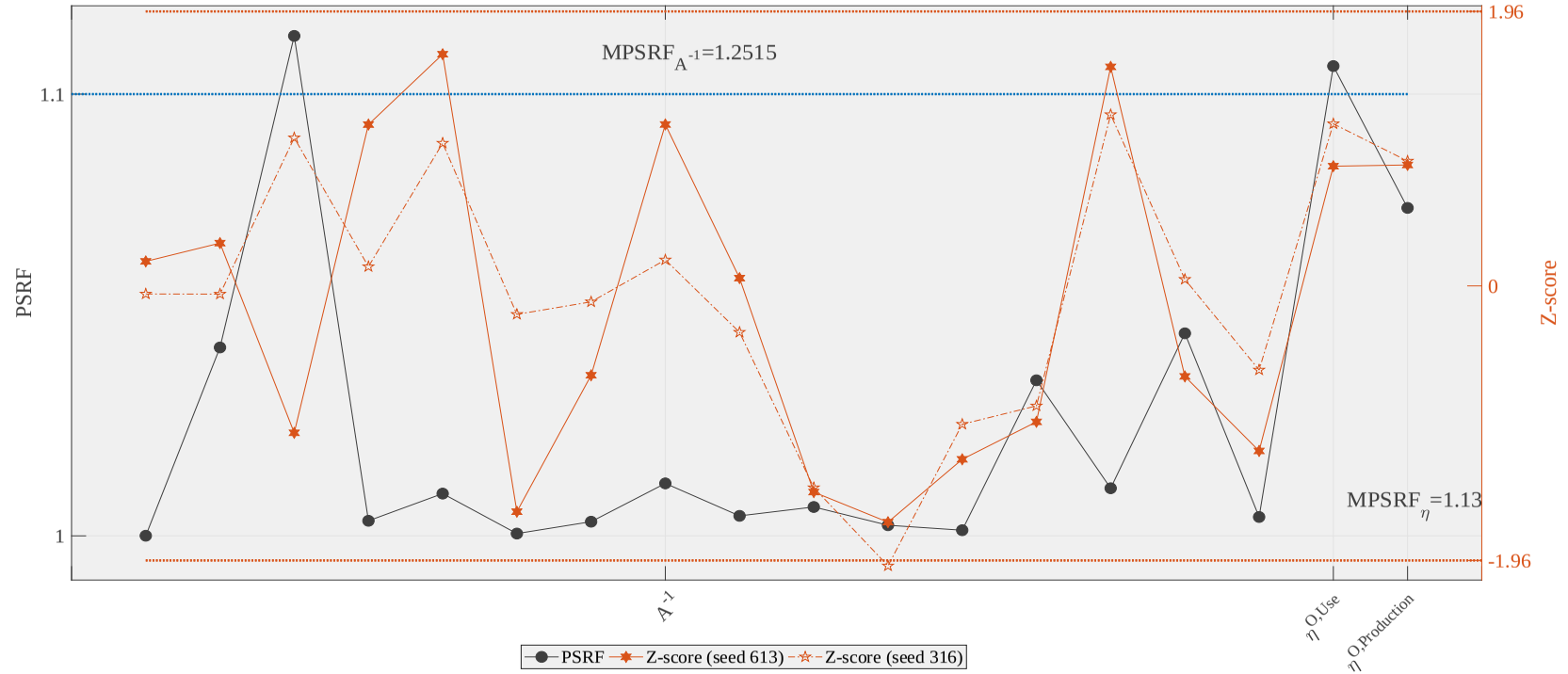

Figure 4.4: Convergence diagnostics for A^{-1}, $\eta^{O,Use}$ and $\eta^{O,Production}$ in Specification I

Note: *The right y-axis* presents Geweke's (1992) $Z-$score using methods from spectral analysis to compare the mean difference between the first 10% and the last 50% iterations of a retained MCMC chain. *Red dotted lines* are the critical values of 5% significance levels, ± 1.96. *The left y-axis* illustrates Gelman & Rubin's potential scale reduction factor (PSRF) based one the two parallel chains with seeds 316 and 613. *The blue dotted line* is the cutoff point suggested by Gelman & Rubin (1992). The parallel chains are diagnosed as converged when PSRF$<$ 1.1. Further, multivariate PSRF (MPSRF), proposed by Brooks & Gelman (1998), for A^{-1} and the two elasticities are reported. If the MPSRF is smaller than 1.1, the chains for multiple variables are jointly converged.

There are two possible reasons causing the sensitivity to seeds as mentioned in the introduction, the MCMC is unconverged and the method suffers the standard weakness of the MH sampler. This is detectable by using joint convergence diagnostics. In this exercise, I use Geweke's (1992) $Z-$score diagnosing the unconvergence for estimates in parallel MCMC chains independently, Gelman & Rubin's (1992) potential scale reduction factor (PSRF) considering the covariance among parallel chains and Brooks & Gelman's (1998) multivariate PSRF (MPSRF) conditional on both the covariances among estimates and MCMC chains. (Please see simulation examples in Cowles & Carlin (1996) and Roy (2019).) If all the tests diagnose the unconvergence, the sensitivity to seeds will simply mean the MCMC algorithm has failed to converge. If estimates are diagnosed as converged independently, but there are evidences for unconvergence considering the covariance among MCMC chains and estimates, the sensitivity would be likely to be caused by the standard weakness of the MH sampler.

Geweke (1992) recommended using methods from spectral analysis to compare the mean difference between the first 10% and the last 50% iterations of a retained MCMC chain. If the calculated $Z-$score is between the critical values at 5% significance levels, ± 1.96, the chain is accepted as being converged. Gelman & Rubin (1992) construct two estimators of the underestimated within-chain-variance and overestimated pool-variance estimates, based on parallel MCMC chains. The PSRF is calculated as the square-root of the ratio between pooled and within-chain variances, which is compared with 1. Gelman & Rubin (1992) set the cutoff value 1.1. If PSRF is smaller than 1.1, the parallel chains are converged. Brooks & Gelman (1998) suggested the use of MPSRF is in the case of multivariate estimates, built upon the PSRF. Hence, if the MPSRF is smaller than 1.1, the chains are converged. It is necessary to mention that all these diagnostics are rough guide for evidence of unconvergence, none of them can precisely indicates the convergence (Cowles & Carlin, 1996; Roy, 2019).

Figure 4.4 illustrates the convergence diagnostics in Specification I, considering two parallel MCMC chains with seeds 316 and 613. The first 16 diagnostics (x-axis) are for parameters in A^{-1} matrix, and the last two are $\eta^{O,Use}$ and $\eta^{O,Production}$ respectively. Geweke's (1992) $Z-$score (right y-axis) is calculated for the two MCMC chains independently. Hence, there are two statistics (orange hexagrams) for each parameter. It is clear that excepting A^{-1}_{24} using seed 316, all parameters from the two chains including the elasticities are converged significantly

at the 5% level. Meanwhile, the left y-axis measures Gelman & Rubin's (1992) PSRF for the two parallel chains jointly. For all the elements' in A^{-1} excepting A^{-1}_{41} are diagnosed as converged, because their PSRFs are smaller than 1.1. However, when considering the covariance among the estimates in A^{-1}, they could be unconverged, as the $\text{MPSRF}_{A^{-1}} = 1.2515$ which is larger than 1.1. Nether $\eta^{O,Use}$ nor the multiple test for it and $\eta^{O,Production}$ can be concluded as converged. This is understandable as the $\eta^{O,Use}$ is calculated through Function (4.19), which is jointly decided by A^{-1}_{11} and A^{-1}_{41}.

It is necessary to say that an accurate inference of crude oil demand elasticities from a structural model is of importance, because the magnitude of them decides not only the existence of crude oil price endogeneity (Kilian & Murphy, 2014), but also the conclusion on the relative importance of supply and demand factors in explaining oil market fluctuations (Caldara et al., 2019).

In the oil-market literature, the lower bound of crude oil demand elasticity is determined by the price elasticity for gasoline demand (e.g. Kilian & Murphy, 2014; Baumeister & Hamilton, 2019), since crude oil represents the main cost of refined products, and therefore a 10% increase in the price of crude oil should result in a $\leqslant$ 10% increase in the price of gasoline. Following this logic, the lower bound of short-run oil demand elasticity for use should be smaller than the short-run gasoline demand elasticity.

There is a broad number of papers devoted to the estimation of the demand elasticity for gasoline. The long-run gasoline demand elasticity is approximately -0.8, (see Hausman & Newey (1995) and a review in Espey (1998), who concluded at -0.58). Therefore, Kilian & Murphy (2014) set the lower bound of short-run oil demand elasticity for use at -0.8. Meanwhile, the short-run elasticity for gasoline demand is approximately -0.26 (see the survey in Dahl & Sterner (1991), utilising the U.S. data). The majority of the microeconomic literatures' estimates are in the range from -0.18 to -0.5, (see Gelman et al., 2016; and Coglianese et al., 2017). Hughes et al. (2006) concluded that the short-run gasoline demand price elasticity is as low as -0.08 for the U.S. data during the 2001–2006 period. Echoing the logic that the short-run elasticity of crude oil should be smaller than that for gasoline, the short-run lower-bound crude oil demand elasticity is uncertain. In summary, there is little consensus about the lower bound of the short-run oil price elasticity of demand.

In the SVARs literature, the importance of the uncertainty has been realised. Baumeister & Hamilton (2019) gives a prior for the demand elasticity as Student $t(-0.1, 0.2, 3)$, truncated to be negative, to address the importance of the uncertainty. However, the oil demand elasticity for use is non-linear, calculated using Equation (4.19) and conditional on the impulse responses at horizon zero. Therefore, this chapter proposes a novel approach to the uncertainty of non-linear lower bounds in SVARs literature, with the following subsection detailing another specification of C-BSVARs for the world crude oil market, which is restricted by the identifications in Kilian & Murphy (2014) as well as a lower-bound uncertainty of the crude oil demand elasticity for use.

4.4.3 Specification II: the uncertainty of the short-run lower bound for oil demand elasticity for use

In contrast to Baumeister & Hamilton (2019), this chapter identifies the parameters in A^{-1}; otherwise, it is impracticable to identify the demand elasticity through the direct use of the prior distribution. This is because the demand elasticity is non-linear, as described in Equation (4.19). Therefore, this chapter proposes modelling the short-run uncertainty of the lower-bound of oil demand elasticity for use, $\underline{\eta}^{O,Use}$, using a Student $t(-0.1, 0.2, 3)$ distribution truncated within the range [-0.8, -0.1]. Short-run lower bounds are randomly sampled from the truncated distribution, which enables infinite possibilities of constrains (or the set identification) for the arg-minimisation (4.20). Hence I can use parallel computing in Steps II and III in Algorithm 1, avoiding the possible weakness of the MH sampler by sampling the parameters only within a local sharp mode, as mentioned in Waggoner et al. (2016). Under the prior, the probability of a random lower bound within the range (-0.5, -0.1) is approximately 89%.

In practice, at each $\underline{\eta}^{O,Use}$, I utilise 1,000 ($N_1 = 1,000$) draws for burnin, and 1,000 ($N_2 = 2,000$) additional draws for gathering constrained posterior draws. The rationale being that the MH process converges quickly and normally within 500 draws. Since B and Σ_u are randomly sampled conditional on A^{-1} being accepted, not all posterior draws satisfy the constraints, and I only retain the constrained posterior draws as admissible posteriors. Then, Steps II and III are repeated with different $\underline{\eta}^{O,Use}$, which differs from Specification I whose $\underline{\eta}^{O,Use} = -0.8$. In this specification, the loop will continue until the size of admissible posteriors is equal

to 100,000 ($N_3 = 100,000$), which requires no more than 3,580,000 draws including burnin.[10] The following paragraphs describe the empirical results of the model.

Before the analysis of elasticities, I illustrate the convergence of Specification II based on two parallel chains with seeds 316 and 613 using Figure 4.5, which is the same as Figure 4.4. It is clear that all the diagnostic statistics indicate that there is no evidence for unconvergence of all elements in A^{-1} individually nor jointly. The crude oil demand elasticities for use and production, $\eta^{O,Use}$ and $\eta^{O,Production}$, are significantly converged.

As shown in the top-left subplot in Figure 4.6, the posterior density of $\eta^{O,Use}$ is skewed towards zero relative to the posterior distribution of $\eta^{O,Production}$. Comparing with the estimates from C-BSVARs without the uncertainty shown in Figure 4.3, the distributions for the elasticities are sharper or less diffuse and are close to zero. The posterior distribution's 84^{th} quantile of $\eta^{O,Use}$ is -0.06, close to the estimates in the literature (e.g. Hamilton, 2008; Dahl, 1993; and Cooper, 2003). However, these estimates in the literature are the oil demand elasticity for production. The 68% confidence set of $\eta^{O,Production}$ via C-BSVARs is [-0.40, -0.15] is significantly larger than the estimates presented in the literature.

Meanwhile, the economic findings in Kilian & Murphy (2014) and described in Appendix C.4 are again confirmed through the second specification with the uncertainty, which are illustrated in Appendix C.5. Therefore, adding the uncertainty of the short-run lower bound for oil demand elasticity for use confirms that Kilian & Murphy's (2014) economic findings are coherent, and modifies the estimations of oil demand elasticities. Further, a prior for key structural parameters, which reflects the corresponding uncertainty, is important. (For a more detailed discussion see Baumeister & Hamilton (2015, 2018, 2019), and Bruns & Piffer (2019).) The uncertainty over the oil demand elasticity provides an alternative means of implementing a prior for the key structural parameters by identifying restrictions.

The following subsection will demonstrate the C-BSVARs' sensitivity to the different upper bounds on oil supply elasticity including 0.0258 to 0.05, 0.1 and 0.5, as well as seeds for random numbers. Subsection 4.4.3 then breaks down the movement of the oil prices, that is, what proportion of the movement is attributable to

[10] Since this step is independent, the parallel calculation can help conserve computational time.

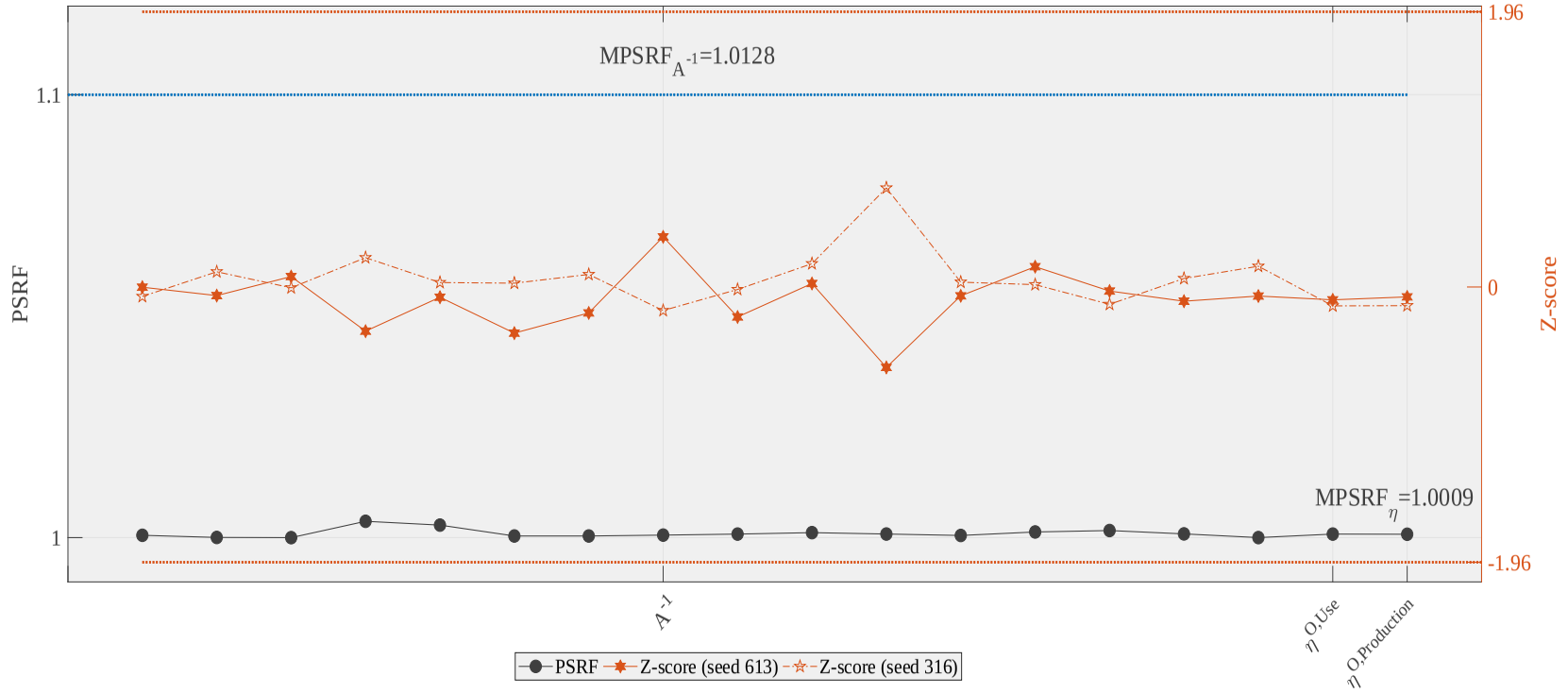

Figure 4.5: Convergence diagnostics for A^{-1}, $\eta^{O,Use}$ and $\eta^{O,Production}$ in Specification II

Note: *The right y-axis* presents Geweke's (1992) $Z-$score using methods from spectral analysis to compare the mean difference between the first 10% and the last 50% iterations of a retained MCMC chain. *Red dotted lines* are the critical values of 5% significance levels, ± 1.96. *The left y-axis* illustrates Gelman & Rubin's PSRF based one the two parallel chains with seeds 316 and 613. *The blue dotted line* is the cutoff point suggested by Gelman & Rubin (1992). The parallel chains are diagnosed as converged when PSRF<1.1. Further, MPSRF, proposed by Brooks & Gelman (1998), for A^{-1} and the two elasticities are reported. If the MPSRF is smaller than 1.1, the chains for multiple variables are jointly converged.

which shocks and at each point in time. Additionally, a comparison between prior and posterior distributions for the identified A^{-1} is illustrated in the first part of Appendix C.5.

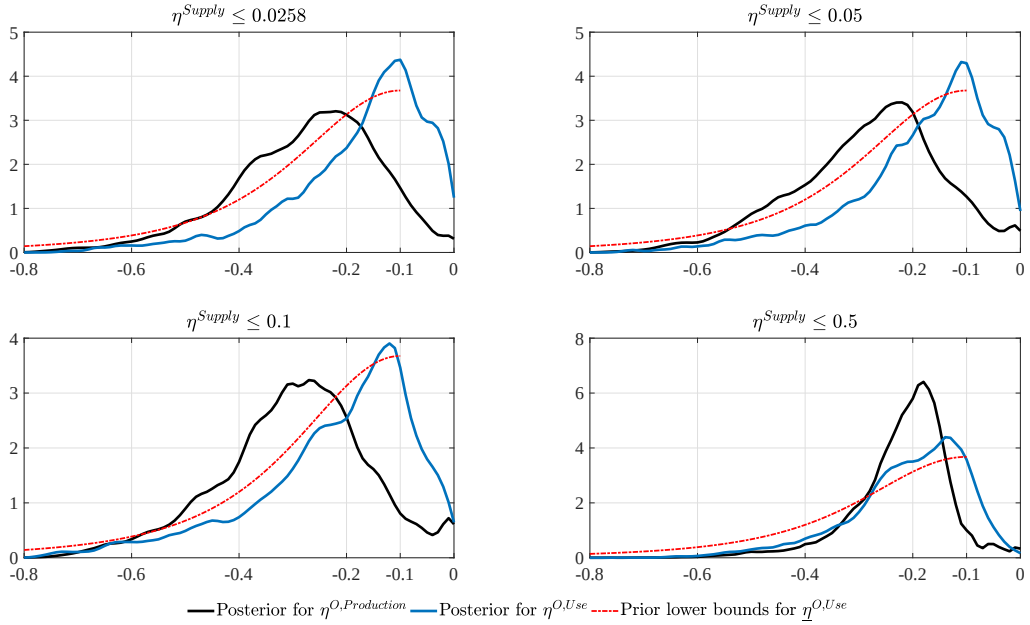

Figure 4.6: Posterior densities of the oil demand elasticity in production and use under an uncertain lower bound with different oil supply elasticity higher bounds, including 0.0258, 0.05, 0.1 and 0.5

Note: $\eta^{O,Production}$ refers to the impact price elasticity of oil demand in production, $\eta^{O,Use}$ to the average impact price elasticity of oil demand in use, and η^{Supply} indicates the impact price elasticity of oil supply. Posterior densities are Kernel fitted, using the Normal method and the bandwidth that optimal for estimating Normal densities. *The dashed red line* represents the prior density for the random lower bounds of oil demand elasticity.

Model sensitivity for the short-run price elasticity of oil demand

Different upper bounds of oil supply elasticity such as 0.0258, 0.05, and 0.1 in Kilian & Murphy (2014) are utilised for the sensitivity check. Meanwhile, Baumeister & Hamilton (2019) argue against the estimates of Kilian & Murphy (2014), which are sensitive to the choice of seeds for random numbers, and assert for that the upper bound of oil supply elasticity to be extended to 0.5. Therefore, this subsection also verifies whether C-BSVARs are sensitive to random number generation seeds and an additional 0.5 upper oil supply elasticity bound.

Figure 4.6 plots the posterior densities of the oil demand elasticity in production, denoted by $\eta^{O,Production}$, the price elasticity of oil demand in use, $\eta^{O,Use}$, and a prior density of random lower oil demand bound for posterior simulation under restrictions of $\eta^{Supply} \leq 0.0258$, $\eta^{Supply} \leq 0.05$, $\eta^{Supply} \leq 0.1$, and $\eta^{Supply} \leq 0.5$, respectively. The absolute estimates for $\eta^{O,Production}$ are clearly larger than the prior given in the literature. Posterior densities of $\eta^{O,Use}$ skew to the left of $\eta^{O,Production}$ for oil supply upper bounds of 0.0258, 0.05, and 0.1. However, when increasing the upper bound of the oil supply elasticity to 0.5, as proffered by Baumeister & Hamilton (2019), the differences of the quantiles between the two distributions of demand elasticities for production and use will vanish. Meanwhile, $\eta^{O,Use}$ has a fatter tail within the range (-0.1, 0) relative to the density of $\eta^{O,Production}$.

Table 4.2: Posterior distributions of the short-run price elasticities of demand for crude oil

		$\eta^{O,Production}$					$\eta^{O,Use}$				
	Seed	2.5^{th}	16^{th}	50^{th}	84^{th}	97.5^{th}	2.5^{th}	16^{th}	50^{th}	84^{th}	97.5^{th}
$\eta^{Supply} \leq 0.0258$	**316**	-0.58	-0.40	-0.26	-0.15	-0.05	-0.53	-0.30	-0.14	-0.06	-0.01
			[-0.80]	[-0.44]	[-0.23]			[-0.54]	[-0.26]	[-0.09]	
	613	-0.58	-0.41	-0.27	-0.15	-0.05	-0.51	-0.30	-0.15	-0.07	-0.02
$\eta^{Supply} \leq 0.05$	**316**	-0.58	-0.42	-0.27	-0.15	-0.04	-0.54	-0.31	-0.15	-0.06	-0.01
			[-0.80]	[-0.45]	[-0.29]			[-0.57]	[-0.27]	[-0.09]	
	613	-0.60	-0.43	-0.28	-0.15	-0.03	-0.56	-0.32	-0.16	-0.07	-0.01
$\eta^{Supply} \leq 0.1$	**316**	-0.60	-0.43	-0.29	-0.17	-0.02	-0.61	-0.36	-0.19	-0.09	-0.02
			[-0.76]	[-0.47]	[-0.24]			[-0.61]	[-0.30]	[-0.10]	
	613	-0.58	-0.43	-0.28	-0.17	-0.03	-0.59	-0.38	-0.18	-0.09	-0.02
$\eta^{Supply} \leq 0.5$	**316**	-0.45	-0.29	-0.20	-0.15	-0.06	-0.49	-0.31	-0.19	-0.11	-0.05
	613	-0.43	-0.29	-0.20	-0.14	-0.06	-0.47	-0.30	-0.19	-0.11	-0.05

Note: $\eta^{O,Production}$ refers to the impact price elasticity of oil demand in production, and $\eta^{O,Use}$ to the average impact price elasticity of oil demand in use, where the latter definition accounts for the role of inventories in smoothing oil consumption. Meanwhile, η^{Supply} indicates the impact price elasticity of oil supply. 16^{th}, 50^{th}, and 84^{th} are the respective percentiles of the posterior. Seed refers to the seed state set in MATLAB for replicating the results. The numbers in square brackets indicate the percentiles reported in Table II of Kilian & Murphy (2014, p. 474).

Table 4.2 illustrates different quantiles (located in the columns) of posterior distributions for $\eta^{O,Production}$ and $\eta^{O,Use}$ under different restrictions (located in the rows). Comparing seeds 316 and 613, the estimates are not the maximum percentage difference, and can produce the same economic conclusions. C-BSVARs for the world crude oil market assign substantial probability mass to the values of $\eta^{O,Production}$ between -0.40 and -0.15, and a lower than 2.5% probability of mass to values close to zero, whose magnitudes are far smaller than the estimates in Kilian & Murphy (2014). Considering inventories, the range [-0.40, -0.15] for $\eta^{O,Production}$ decreases to the range [-0.30, -0.06] for $\eta^{O,Use}$ with a probability of 68% under $\eta^{Supply} \leq 0.0258$. Kilian & Murphy's (2014) estimates of the 16^{th} quantile for $\eta^{O,Use}$ are close to

the 2.5^{th} quantile utilising C-BSVARs and the uncertainty restriction. C-BSVARs modify the posterior distributions reported in Kilian & Murphy (2014), for both $\eta^{O,Production}$ and $\eta^{O,Use}$ under all three supply elasticities' upper bounds of 0.0258, 0.05 and 0.1 towards zero.

Historical decomposition of the real oil price

In this subsection, I illustrate the median cumulative effects of flow supply shock (grey), flow demand shock (yellow) and speculative demand shock (blue) on the real oil price in the upper panel of Figure 4.7. Over the 1978:06–2009:08 period, the flow demand shock drives the trend of the real oil price, while the short-run fluctuation is jointly driven by the flow supply and speculative demand shocks. For example, the log-level real oil price will fluctuate above zero during the majority of the period when the cumulative effects of the flow demand shock are positive. These are consequent to the fact that the absolute values of the oil demand elasticities are larger than those of the oil supply elasticities.

The lower panel of Figure 4.7 reviews the five events mentioned by Kilian & Murphy (2014) and detailed in Appendix C.4. The box plots are the cumulative effects' median and the corresponding 68% posterior confidence sets, with black, orange and blue markers representing the cumulative effects of flow supply, flow demand, and speculative demand shocks, respectively.

During the Iranian revolution, an upward trend of the real oil price was driven by the persistent recovery of the flow oil demand, where the upward hump-shaped oil price from 1979:04 to 1980:01 was caused by the speculative demand shock. Meanwhile, there were no observed supply disruptions, which is consistent with the observation that oil output from outside Iran filled the gaps. In contrast, a high oil price prevailed during the Iran–Iraq War, jointly caused by three shocks: persistently high positive flow demand shocks, speculation demand shocks, and supply disruption shocks. Both episodes resulted in a high oil price accompanied with positive flow demand shocks.

Meanwhile, the Persian Gulf War and Iraq War resulted in weak oil demand. Consequently, the increase of the log real oil price, driven by high volumes of speculation, presented within the short windows of 1990:07–1990:10 and 2002:11–2003:02, respectively. During the 1990:07–1990:10 period, Saudi oil fields were under threat

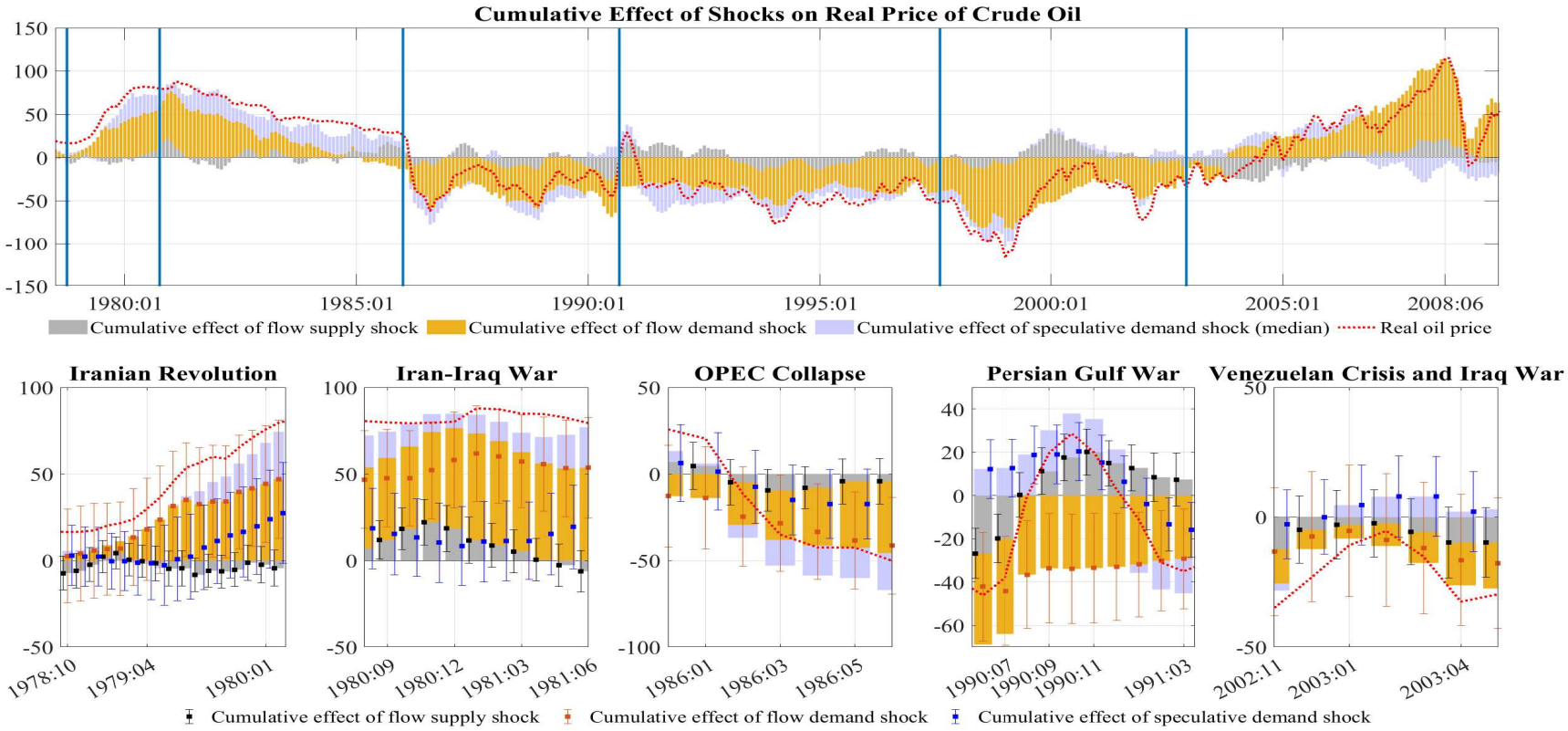

Figure 4.7: The historical decomposition for the real crude oil price for the 1978:06–2009:08 period

Note: *The upper panel* presents the historical decompositions of the real oil price for the 1978:06–2009:08 period. *The stacked bar charts* are the median cumulative effects of the flow supply shock (*grey*), flow demand shock (*yellow*) and speculative demand shock (*blue*). *The lower panels* illustrate the Iranian Revolution, the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, and the Venezuelan crisis, followed by the Iraq War, respectively, which are presented as vertical blue lines in the upper panel (including the Asian Financial Crisis). *The red dashed lines* show the observed oil price. *The box plots* are the cumulative effects' median and the corresponding 68% posterior credible sets. *Black*, *orange* and *blue markers* indicate the cumulative effects of flow supply, flow demand, and speculative demand shocks, respectively.

from the Iraqi military, which induced that investors were bullish on crude oil prices. The threat was removed by the presence of the U.S. troops from 1990:11, which resulted in a drop in the real oil price associated with the weak demand. In October 2002, the U.S. President George W. Bush's proposal to use military strikes against Iraq was authorised by Congress. Coupled with the ongoing crisis in Venezuela, investors became concerned about the drop in world crude oil production, resulting in a subsequent increase in speculative demand that raised oil prices during the 2002:11–2003:02 period. In 2003:03, the Iraq War broke out and the U.S. military did not encounter significant resistance. Declining speculative sentiment and oversupply in crude oil market during that period caused oil prices to fall again.

Moreover, during the first half of the 1980s, OPEC countries decreased production on several occasions with the intention of maintaining high oil prices. Consequently, OPEC's share of the global market fell from about half in the 1970s to less than one-third in 1985. During the same period, despite OPEC members being required to meet their production quotas, many of them exaggerated their foreign exchange reserves in order to achieve higher quotas, engaged in deception or directly refused to comply with the quotas. This phenomenon presents as high positive speculation demand shocks over the 1980:09–1985:12 period in the upper panel of Figure 4.7. By the end of 1985, Saudi Arabia had tired of this behaviour and decided to punish those OPEC countries that lacked discipline by producing at full capacity. Over 1986, high-cost oil production facilities were becoming less or even unprofitable, and the nominal oil price fell to \$7 a barrel.

Further, a comparative analysis between the estimates under $\eta^{Supply} \leq 0.0258$ and $\eta^{Supply} \leq 0.5$ for the impulse responses and historical decompositions is presented in Appendix C.6. Controlling for all other restrictions, the inflation of the oil supply elasticity's upper bound does not change the relative importance of the cumulative effect of supply and demand shocks in terms of the real oil price. Despite, it does lead to a slight divergence of the relative importance of the cumulative effect of flow supply regarding speculation demand shocks (see Figure C.7). This is also confirmed by a comparison of the impulse responses for the two restrictions in Figure C.8. Increasing the upper bound for the supply elasticity will not change any endogenous variable's response to the flow demand shock. These estimates rule out the argument that identifying the restrictions on oil supply elasticity is the key factor for determining the relative importance between the effects of flow oil supply and demand shocks, and the subsequent impact on the real oil price.

4.5 Conclusion

This chapter has developed a novel method to identify and compute the Bayesian SVARs, which is both more efficient and general than the methods used in the existing literature. C-BSVARs use a parallel MH sampler, which simultaneously draws MCMC chains within numerical solutions constrained by a set identification for parameters of interest, to compute posterior inferences. The C-BSVAR, to my knowledge, is the first method for replicating a sufficient posterior sample size using Kilian & Murphy's (2014) world crude oil market structural model. Under their identification, C-BSVARs can replicate all the economic findings described in Kilian & Murphy (2014) but narrow the distribution of their impulse response functions. Moreover, the use of C-BSVARs improved the researcher's efficiency in obtaining for models that satisfied all the restrictions in Kilian & Murphy (2014) from 16/5,000,000 to 100,000/3,580,000, including half of the draws for burnin. However, the posterior densities of the oil price demand elasticities are too diffuse to conclude oil price endogeneity, and their density estimates were sensitive to the choice of Matlab seeds. The convergence diagnostics provide evidence suggesting the cause for the sensitivity is highly likely to be non-unimodal posterior distributions for the elasticities.

Given the importance of the oil demand elasticities and the concern of non-unimodal property, this chapter also proposed an additional restriction for Kilian & Murphy's (2014) set identification — the short-run lower-bound uncertainty of oil demand elasticity for use — in order to shrink the estimates of demand elasticities. Imposing the additional restriction, none of diagnostics considered in this exercise showed evidence of unconvergence, and the sensitivity to the choice of seeds disappeared. C-BSVARs confirmed the economic findings in Kilian & Murphy (2014), and modified their estimations of 68% confidence sets for both impulse response functions and oil demand elasticities. There is evidence that during the 1973:02–2009:08 period, the flow demand played an important role in determining the long-run trend of the real oil price, while the speculation and oil supply shocks jointly drove the short-run oil price fluctuations. This is consequent on the fact that the absolute value of oil demand elasticities are higher than those of oil supply elasticities. Moreover, the uncertainty for a lower-bound oil demand elasticity provides a novel means of giving a prior to the uncertainty of key structural parameters in identifying non-linear restrictions.

Chapter 5

Conclusion

As one of the leading world economic indicators, the crude oil price has long been monitored by central banks, policy makers, and market participants. Moreover, it has been recognised that forward-looking expectations are of importance in terms of agents' price setting behaviour. Therefore, an accurate forecast of the crude oil price is desirable, with Chapters 2 and 3 concerned with improving the point and density forecasting accuracy of real crude oil prices.

Chapter 2 has extended Baumeister & Kilian's (2015) evaluation period to the end of 2016 using the Brent oil price measure, which confirms that their forecast combination is robust in point forecasts. Further, I provided alternative revision and nowcast assumptions, aiming to improve Baumeister & Kilian's (2015) equal-weight combination's forecast accuracy across different forecast horizons. In Chapter 3, I examined the density forecasts. In order to improve the predictability of time-varying parameter models, I have developed an out-of-sample forecasting process for improving highly parameterised models' density forecast accuracy through minimising one-step-ahead in-sample posterior estimates' Kullback–Leibler 'distance', with a process incorporated to eliminate the extreme forecasts. Moreover, this chapter has extended the standard statistical evaluations of point and density forecasts to (1) the probability forecast that are more valuable in terms of excess returns in futures market, and (2) forecasting the likelihood of extreme high and low real crude oil prices.

The computational inefficiency of the traditional SVARs approaches under a set identification used in the oil literature motivates a novel method C-BSVARs,

which is developed in Chapter 4. C-BSVARs use parallel MH samplers to dramatically improve the computation efficiency, and enable a far boarder identification scheme than sign restrictions in Baumeister & Hamilton (2015). The use of C-BSVARs are, to my knowledge, the first method which allows for replicating a sufficient posterior sample size using Kilian & Murphy's (2014) world crude oil market structural model. Nevertheless, a large sample size of posterior is not enough to guarantee the precise estimates of oil demand elasticities, which are diagnosed by unconvergence tests as its underline posterior distribution is highly likely non-unimodal. This will induce the standard weakness of MH sampler, documented in Waggoner et al. (2016). Considering the non-unimodal posterior distribution and the importance of implications for oil demand elasticities, this chapter also proposed an additional restriction for the Kilian & Murphy (2014) model — a prior for the short-run lower-bound uncertainty of oil demand elasticity for use, which is utilised for shrinking the estimates of demand elasticities. The prior is virtually impossible to impose using the methods that currently exists in the econometric literature, but it serves statistically consistent estimates of the elasticities. C-BSVARs provide a novel way of acquiring a prior on the uncertainty of key structural parameters in identifying non-linear restrictions.

Further research

Given the promising results in Chapter 3, the *long–short* investment stratagem utilised for evaluation purposes can be applied in practice. Firstly, this is a more flexible platform for hedgers or speculators to construct their portfolio. Secondly, policy makers can utilise the process as a monitor to track the price of crude oil in order to facilitate policy reactions.

The improvement in terms of stochastic volatility for density forecasting in Chapter 3 has implications for structural analysis. Moreover, the observed smooth structural change for parameters in the global oil market implies that a time-varying analysis is desirable. Therefore, an extension of the C-BSVARs in Chapter 4 considering time-varying parameters and stochastic volatility may be of interest.

The C-BSVARs enable researchers to use complex set identifications. For example, there is a considerable literature contributing to the transmission analysis of oil shocks to stock returns using stock indices. However, none of them allows

firms-level stock returns. The relative magnitude of firms-level stock returns' responses should be ordered as a measure of the firms' exposure to oil. This kind of ranking restrictions could be considered in the further research. C-BSVARs may be a method for adding it into a structural analysis.

As emphasised in Chapter 3, parameter time variation is of importance in real crude oil price forecasting, and there also be a key factor for oil market structural analysis as well. For example, the role of fossil fuels in the energy mix is diminishing, and the U.S. is becoming an oil exporter from the largest importer around the world due to the boom of shale oil. Both of these developments may lead elasticities of oil demand and supply to change over time. Both identifications and computations are the difficulty of applying time-varying SVARs with stochastic volatility in world crude oil market analysis. However, this is a crucial topic in the future analysis.

Appendix A

Process for Constructing CPI, World Economic Activity Index and Backcasts

A.1 The U.S. consumer price index for all urban consumers (CPI)

Real-time data for the monthly seasonally adjusted the U.S. CPI for all urban consumers are obtained from the *Economic Indicators* published by the Council of Economic Advisers (CEA). Data in the most resent month of a vintage is made available by the Federal Reserve Archival System for Economic Research (FRASER) database of the Federal Reserve Bank of St. Louis.[1] Additional real-time CPI data is obtained from the macroeconomic real-time database of the Federal Reserve Bank of Philadelphia, which is only available after the vintage 1998:11.[2] In constructing real-time CPI data before the vintage 1998:11, I encounter a challenge in that the CEA only reports the 12 most recent months of each vintage. The missing data back to 1973:01 are populated according to the quarterly vintages of monthly real-time CPI data from the Federal Reserve Bank of Philadelphia, exploiting the fact that the observations shown in each vintage represent the data available in the middle of the quarter. I have updated the data set on a monthly basis since 2016:12 according to the monthly publication of the *Economic Indicators* from CEA.

[1] https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ECONI&browsePath=1995&isCollapsed=true&leafLevel%5CBrowse=false&ycord=208

[2] https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/pcpi

A.2 An index of the bulk dry cargo ocean shipping freight rates

This index is proposed in Kilian (2009), known as the global real economic activity (rea) index. The rea is one of the regressor in an unrestricted global oil market vector autoregression (VAR) model, detailed as Equation (2.2) in Section 2.3. Resources for constructing rea in this chapter include the single voyage freight rates collected by Drewry Shipping Consultants Ltd., historical exchange rates available in Bloomberg, quoted as GBPUSD, the U.S. Consumer Price Index (CPI) introduced above, and the Baltic Dry Cargo Index available in Bloomberg quoted as BDIY:INDEX.

Kilian's (2009) rea index is continually updated on Kilian's web page (https://drive.google.com/file/d/1GsjrZtJG7k4Z1PCs45q6IPF7tqb2rLOc/view), which is constructed as follows: (1) access the single voyage freight rates collected by Drewry Shipping Consultants Ltd. for various bulk dry cargoes comprising grain, oilseeds, coal, iron ore, fertilizer and scrap metal, which are provided for different commodities, routes and ship sizes; (2) take simple averages of the freight rates; (3) compute the period-to-period growth rates for each series; (4) take the equal-weighted average of these growth rates, and cumulate the average growth rate, having normalised 1968:01 to unity; (5) deflate this series with the U.S. CPI; and (6) linearly detrend the real index.

However, CPI is revised and the deterministic trend is recursively calculated through historical vintages. Therefore, it is impossible to construct a real-time Kilian's (2009) rea index without the raw data of freight rates mentioned above. I seek the raw data published in the Shipping Statistics and Economics journal (ISSN: 0306-1817, No.1–282, 1970–1994).

Therefore, I break the time series into two components: (1) the average monthly single voyage freight rates for 1970:05 to 1985:11; and (2) the nominal shipping rate raw data (the Baltic Dry Cargo Index), available after 1985:01. The Baltic Dry Cargo Index is utilised to construct Kilian's real economic index for analysing the oil market models (see Baumeister & Kilian, 2012; Kilian & Murphy, 2014; Baumeister & Kilian, 2015). The detailed process to construct real-time rea is as follows:

1 Single voyage freight rates from 1970:05 to 1985:115 are collected by Drewry Shipping Consultants Ltd. (only available in hard copy).[3] Various bulk dry cargoes are considered including grain, coal, iron ore, fertiliser and chemicals, oilseeds, scrap metal, iron, steel, miscellaneous products, sugar, other agricultural products,[4] manufactured products,[5] forest products, and other ores and minerals,[6] which are provided for different commodities, routes and ship sizes. During the period considered here, since some of these items are priced in pound sterling (GBP), the GBP value is converted into USD through the monthly historical exchange rates calculated as the average of the daily last observations (available in Bloomberg, quoted as GBPUSD).

2 Take the average of the single voyage freight rates by bulk dry cargoes, shown in the first (top) panel in Figure A.1.

3 Take the average of the average rates categorised by bulk dry cargoes, and then magnify the size of the average rate by normalised 1985:01 to 1,000, as the Baltic Dry Cargo Index begins with 1,000 in 1985:01. This is shown in the second panel in Figure A.1.

4 The Baltic Dry Cargo Index is shown in the third panel in Figure A.1.

5 Combine the equal-weighted dry cargo index and the Baltic Dry Cargo Index, where the over-lapping portions (1985:01 to 1985:11) are the average between the two indexes. Then, the combined index normalises 1970:05 to unity, as shown in the fourth panel in Figure A.1.

6 The real-time world real economic activity index is recursively constructed by the combined index deflated with the U.S. CPI and linearly detrended, and scaled by times 100. The final vintage of 2017:06 is shown in the last panel in Figure A.1.

I also highlight events in oil market after 1973, including the Yom Kippur War/OPEC oil embargo (1973–1974), the Iranian Revolution (1978–1979), the Iran–Iraq War (1980–1988), the Gulf War (1990–1991), Venezuela's civil unrest (2002),

[3] Kilian (2009) track rates back to 1968; however, the data published in the Shipping Statistics and Economics journal are only available back to 1970:05.

[4] Inclusive of citrus pellets, lentils, logs, flour, soya, and tallow.

[5] Inclusive of agricultural products, rice, cement, and petroleum coke.

[6] Inclusive of alumina, barytes, concentrates, bauxite, chrome ore, tincal borate, magnesite and manganese ore.

the Iraq War (2003–2011), the financial crisis (2008), the Gaza War (2009), and the European sovereign debt crisis (2010–2014). These events profoundly influenced the balance of the world oil markets. Consistent with Kilian's (2009) rea, the combined index react to the global real economic activity shown in Figure A.1. Notedly, the real-time rea changes over vintages, due to the revisions of the U.S. CPI, linearly detrending, and rescale processes.

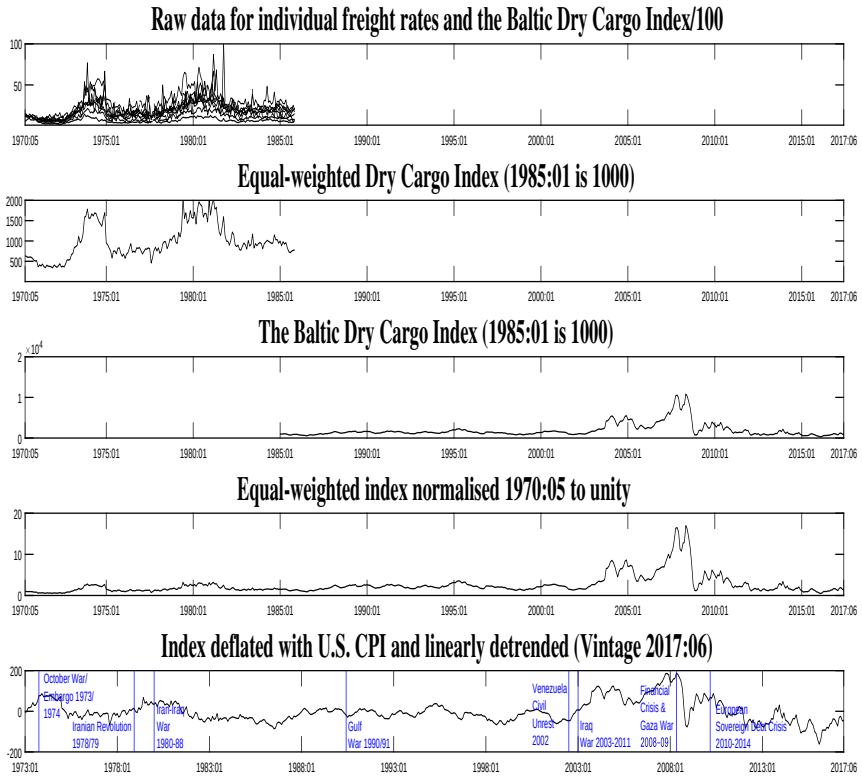

Figure A.1: Monthly index of the global real economic activity based on dry cargo bulk freight rates and the Baltic Dry Cargo Index (1973:01-2017:06)

A.3 Backcasts

This appendix introduces the backcasts for RAC, OECD petroleum stocks, and the futures of WTI and Brent measures. The backcasts of the first two variables are following Baumeister & Kilian (2012), while the backcast for futures are original.

A.3.1 The backcasts for RAC and OECD petroleum stocks

Echoing Baumeister & Kilian (2012), the real-time data set contains vintages from 1991:12 to 2017:06, each covering data extending back to 1973:01. However, RAC and OECD petroleum stocks are available after 1974:01 and 1988:01, respectively. In this appendix, I introduce the method for extrapolating these missed observations.

In constructing the monthly RAC from 1973:01 to 1974:01, I follow Baumeister & Kilian (2012). This procedure involves scaling the monthly percentage rate of change in the U.S. crude oil producer price index (PPI for oil), provided by the U.S. Bureau of Labor Statistics (available at: https://data.bls.gov/timeseries/PCU333132333132), for 1973:01 to 1974:01 by the ratio of the growth rate in the annual refiners' acquisition cost over the growth rate in the annual U.S. PPI for crude oil. The process is:

1. Estimate the approximate monthly RAC growth rate, $\frac{\widehat{RAC_{1973:i+1} - RAC_{1973:i}}}{RAC_{1973:i}}$, where $i = [1, \ldots, 12]$ and if $i = 12$, 1973:12+1 represents 1974:01, through:

$$\frac{\widehat{RAC_{1973:i+1} - RAC_{1973:i}}}{RAC_{1973:i}} = \frac{PPI^{oil}_{1973:i+1} - PPI^{oil}_{1973:i}}{PPI^{oil}_{1973:i}} \frac{\frac{RAC_{1974} - RAC_{1973}}{RAC_{1973}}}{\frac{PPI^{oil}_{1974} - PPI^{oil}_{1973}}{PPI^{oil}_{1973}}},$$

   where, $RAC_{1973:i}$ is the RAC observation in the month 1973:i, while $PPI^{oil}_{1973:i}$ is the monthly PPI for oil observed in the month 1973:i. RAC_{1973} and RAC_{1974} are the annually observation of RAC in 1973 and 1974, which is 4.08 and 12.52 dollars per barrel respectively (Available at https://www.eia.gov/opendata/qb.php?category=293676&sdid=PET.R1300____3.A). PPI^{oil}_{1974} and PPI^{oil}_{1973} are annual PPI for oil, calculated as the average of the monthly observations, in the years 1973 and 1974.

2. The backcasts of RAC then utilise the backward induction from 1974:01 to 1973:01. Initially, $\widehat{RAC_{1973:12}}$ is backcast through:

$$\widehat{RAC_{1973:12}} = \frac{RAC_{1974:01}}{1 + \frac{\widehat{RAC_{1974:01} - RAC_{1973:12}}}{RAC_{1973:12}}}.$$

   Then conversely, $\widehat{RAC_{1973:i}}$ can be estimated where i=[11, . . . , 1] through:

$$\widehat{RAC_{1973:i}} = \frac{\widehat{RAC_{1973:i+1}}}{1 + \frac{\widehat{RAC_{1973:i+1} - RAC_{1973:i}}}{RAC_{1973:i}}}.$$

For OECD petroleum stocks prior to 1988:01, I extrapolate the percentage change in OECD inventories, $OECDpetro_t$, backwards at the growth rate of U.S. petroleum inventories ($USpetro_t$), following Kilian & Murphy (2014):

$$\widehat{OECDpetro}_{1987:12} = \frac{OECDpetro_{1988:01}}{1 + \frac{USpetro_{1988:01} - USpetro_{1987:12}}{USpetro_{1987:12}}}.$$

Then, it can be estimated $\widehat{OECDpetro}_t$ where t is from 1987:11 backwards to 1973:01 through:

$$\widehat{OECDpetro}_t = \frac{\widehat{OECDpetro}_{t+1}}{1 + \frac{USpetro_t - USpetro_{t-1}}{USpetro_{t-1}}}.$$

Additionally, I adjust the real-time OECD petroleum inventory data to account for changes in the set of OECD members reporting inventories in 2001:12. EIA (2001) reported here are four routine revisions: (1) South Korea is added to the table; (2) data for the Czech Republic, Hungary, and Poland are added to "OECD Europe"; (3) data for Mexico are added to "Other OECD"; and (4) OECD is recalculated to reflect the changes in other columns. It is necessary to preserve the consistency of the real-time and the ex-post revised inventory data by adding them into previous vintages provided by Baumeister & Kilian (2012). Consistent with Baumeister & Kilian (2012) ex-post revised data from the EIA supplement the petroleum inventories of these countries to construct real-time equivalents of the OECD petroleum inventory data in vintages prior to 2001:12.

A.3.2 The backcast for oil futures with maturities 1–24 months

WTI futures are traded on the New York Mercantile Exchange (NYMEX), while the Brent oil price futures are traded on the trading floor of Intercontinental Exchange

Futures Europe (ICE). I collect these monthly WTI and Brent futures prices, with maturities from 1 to 24 months, from Bloomberg, calculated as the average of the daily last prices (quoted as cl'n':com and co'n':com for WTI and Brent, respectively, where 'n' is the maturity).

The missing values from the online source in WTI futures are with the 18- to 24-month maturities from 1991:12 to 1995:08, while the missing values in Brent futures are with the 9- to 24-month maturities between 1991:12 and 1994:04, and with the 12- to 24-month maturities in the 1994:04–2005:02 period, respectively. Values for monthly missing data are recursively added through the following process, based on Wiki crude oil futures continuous contracts from CL1 to Cl24 (available at https://www.quandl.com/data/CHRIS-Wiki-Continuous-Futures), where CL is the code for the Light Sweet Crude Oil futures contracts:

1. Calculate the spread between WTI (Brent) futures and different Wiki futures prices with the h maturity.

$$r^h_{ij,t} = \frac{f^h_{i,t}}{f^h_{Wiki,j,t}},$$

where i denotes WTI and Brent crude oil, and t = [1991:12, ..., 2017:06]; $r^h_{ij,t}$ is the rate with the $h-$month maturity, $h = 1, 2, ..., 24$; $f^h_{i,t}$ is the futures of WTI or BRENT with the maturity h; and, $f^h_{Wiki,j,t}$ is the Wiki crude oil futures with the same maturity h, where j indicating 5 Wiki prices at time t: the open, high, low, last and settle prices. If the WTI (or Brent) futures and the 5 Wiki futures with 1- to 24-month maturities at time t are available, there will be the maximum 120 (5×24) spreads calculated.

2. At time t, where there is a missing value of the WTI (or Brent) futures with the maturity h, I replace the missing value, denoted as $\widehat{f}^h_{i,t}$, using:

$$\widehat{f}^h_{i,t} = \frac{1}{J \times v} \sum_{\widetilde{h}=1}^{v} \sum_{j=1}^{J} (r^{\widetilde{h}}_{ij,t} \times f^h_{Wiki,j,t}),$$

where v is the number of available observations with the 1- to 24-month maturities at t for the measure WTI (or Brent); and, $\widetilde{h}$ is the maturity not equal

to (any) of h maturities, where there are missing values. In the event of one missing maturity h over the 24 this will be the mean of $J \times v$ approximations of the missing data, where J is the number of available Wiki prices indicated by j with the maturity h. Hence, the number of observations over which I take the mean depends on the number of maturities' observations that are available. For example, I have missing WIT futures with maturities 18, 21, and 24 months in 1991:12, while all five Wiki prices are available. Hence, the number of values averaged over to fill in the value at $h = 18$ in 1991:12 is $5 \times 21 = 105$, where 21 is calculated as 24 maturities considered in this data set minus the three missing maturities in 1991:12.

Appendix B

Technical Details and Additional Results in Chapter 3

B.1 Prior and the posterior computation for ω_j^* in SMSS

Following Eisenstat et al. (2016), I define $m \times 1$ vectors $\omega^* = (\omega_1^*, ..., \omega_m^*)'$, $\omega = (\omega_1, ..., \omega_m)'$, and $\tau = (\tau_1, ..., \tau_m)'$, to complete the model specification (3.4) and save

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \quad W = \begin{pmatrix} W_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & W_T \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix}, \quad \gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_T \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix},$$

$W_t = Z_t \Omega^{\frac{1}{2}} \Phi$, where $\Omega^{\frac{1}{2}} = diag(\omega_1, ..., \omega_m)$; and, W is stored as a sparse matrix. Moreover, I store $h = (h_1, \dots, h_T)'$. Then, specification (3.4) becomes:

$$y = Z\alpha + W\gamma + \varepsilon. \tag{B.1}$$

And then, $\varepsilon \sim N(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \Sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Sigma_T \end{pmatrix}.$$

Therefore, the posterior draws are obtained by sequentially sampling (MCMC) from:

1. $p(\alpha|y,\gamma,h,\omega^*,\tau,\lambda)$;
2. $p(\gamma|y,\alpha,h,\omega^*,\tau,\lambda)$;
3. $p(h|y,\alpha,\gamma,\omega^*,\tau,\lambda)$;
4. $p(\omega^*|y,\alpha,\gamma,h,\tau,\lambda)$;
5. $p(\tau|y,\alpha,\gamma,\omega^*,h,\lambda)$;
6. $p(\lambda|y,\alpha,\gamma,\omega^*,h,\tau)$.

Steps 1–3 are standard (see details in Eisenstat et al., 2016). As per the Tobit prior explained in Section 3.2.1, ω_j is fully determined by ω_j^*, which is sampled from its full conditional distribution. To that end, define $G_t = Z_t diag(\gamma_t)$, and let

$$G = \begin{pmatrix} G_1 \\ \vdots \\ G_T \end{pmatrix}.$$

Then the measurement Equation (B.1) becomes

$$y = Z\alpha + G\omega + \epsilon. \tag{B.2}$$

And, g_j denotes the j-th column of G, while $G_{\setminus j}$ is the rest of the columns of G stacked in a $Tn \times (m-1)$ matrix. Similarly, $\omega_{\setminus j}$ is ω eliminating ω_j. And,

$$\upsilon_j \equiv y - Z\alpha - G_{\setminus j}\omega_{\setminus j}.$$

Calculate the further posterior quantities:

$$\begin{aligned}
\widehat{\tau}_j^2 &= (\tau_j^{-2} + g_j^{'}\Sigma^{-1}g_j)^{-1}; \\
\widehat{\mu}_j &= \widehat{\tau}_j^2(\mu_j/\tau_j^2 + g_j^{'}\Sigma^{-1}\upsilon_j); \\
\widehat{\psi}_j &= \frac{\Phi(\widehat{\mu}_j/\widehat{\tau}_j)}{\Phi(-\mu_j/\tau_j)}\frac{\widehat{\tau}_j}{\tau_j}exp\{\frac{1}{2}(\frac{\widehat{\mu}_j^2}{\widehat{\tau}_j^2} - \frac{\mu_j^2}{\tau_j^2})\}; \\
\widehat{\pi}_j &= (1-\widehat{\psi}_j)^{-1}.
\end{aligned}$$

The μ_j is the expected mean of a normal distribution, and it is restricted to zero in Chapter 3 to reduce the computation burden. Then, the conditional density of ω_j^* in step 4 is a 2-component mixture of truncated normals:

$$p(\omega_j^*|y,\alpha,\gamma,h,\tau,\lambda,\omega_{\setminus j}^*) = \widehat{\pi}_j \phi_{-\infty,0}(\omega_j^*|\mu_j,\tau_j^2) + (1-\widehat{\pi}_j)\phi_{0,\infty}(\omega_j^*|\widehat{\mu}_j,\widehat{\tau}_j^2), \quad \text{(B.3)}$$

(see a more detailed discussion in Eisenstat et al., 2016, p. 1646). Steps 5 and 6 are standard as below:

$$(\tau_j^{-2}|\lambda,\omega_j^*) \sim \boldsymbol{IG}(\sqrt{\frac{\lambda^2}{(\omega_j^*)^2}},\lambda^2),$$

$$(\lambda^2|\tau) \sim \boldsymbol{\Gamma}(\lambda_{01}+m, \lambda_{02}+\frac{1}{2}\sum_{j=1}^{m}\tau_j^2),$$

where $\boldsymbol{IG}$ and $\boldsymbol{\Gamma}$ denote the inverse Gaussian and Gamma distributions respectively. The starting value of τ_j^2 is based on exponential ($\mathbf{E}$) distribution $\tau_j^2 \sim \boldsymbol{E}(1)$.

B.2 Technical details for specifications

The models without stochastic volatility, including BVAR, TVP, and TVPsmss, use a constant variance of ε, Σ so that:

$$\Sigma = \begin{pmatrix} \Sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Sigma_1 \end{pmatrix},$$

Σ_1 is sampled following a standard independent prior from an inverse Gamma distribution, $\boldsymbol{I\Gamma}$:

$$\Sigma_1 \sim \boldsymbol{I\Gamma}(v_0 + T/2, \frac{1}{S_0 + \varepsilon'\varepsilon/2}).$$

The hyper parameters v_0 and S_0 are $n \times 1$ vector with zeros and -1 respectively. This follows Chan & Eisenstat (2018).

Moreover, the specifications with TVP but without SMSS include TVP and TVPSV. I sample ω_j^* using the positive truncated distribution

$$\omega_j^* \sim \phi_{0,\infty}(\omega_j^*|\widehat{\mu}_j,\widehat{\tau}_j^2)$$

and ignoring $\widehat{\pi}_j$ in Equation (B.3). This utilises a straightforward implementation of hierarchical shrinkage, namely LASSO.

Further, the specifications without TVP include BVAR and SV, where the ω_j is restricted to zeros for all $j = 1, \ldots, m$.

B.3 The extension of Baumeister & Kilian's (2015) combination into density forecasts

In this appendix, I extend the models introduced in Section 2.3 with approximate density forecasts. The majority of them are sampled from a normal distribution with the mean of the point forecast at a forecasting horizon and the variance based on the square of the model's recursive historical forecasting error.

B.3.1 An unrestricted globe oil market vector autoregression (VAR)

The density forecast from VAR is based on the point forecast of $\widehat{R}^1_{t+h}$ in Equation (2.2). The density is simulated through 1,000 simple Monte Carlo iterations, conditional on a normal distribution with the mean of the point forecast at the horizon h, and the variance calculated by the VAR.[1] Then, the density forecast of RAC is denoted as $g(\widehat{R}^1_{t+h})$. WTI and Brent measures are again extrapolated from the RAC forecasts, as mentioned in Section 3.2.3.

B.3.2 A commodity price based model

The density forecast based on the price of non-oil industrial raw materials is calculated as:

$$\widehat{R}^2_{t+h} = \widehat{R}^{oil}_t(1 + \pi^{h,\text{industrial raw materials}}_t - E_t(\pi^h_{t+h})) + \varepsilon_{t+h},$$

where $\varepsilon_{t+h} \sim NID(0, \sigma^2_{h|C})$, and $\sigma^2_{h|C}$ is the observed average forecast squared error at horizon h conditional on the commodity price based model; $E_t(\pi^h_{t+h})$ is the inflation expectation formulated in Equation (2.4). Hence, the forecast density $g(\widehat{R}^2_{t+h})$

[1] The choice of density size is detailed in the end of Section 3.2.1.

at the horizon h is the Monte Carlo integration simulated by 1,000 simple Monte Carlo iterations conditional on the most recent data.

B.3.3 An oil futures spread based model

The oil futures-based density forecast, $g(\widehat{R}^3_{t+h})$, is based on the addition of a mean-zero error:

$$\widehat{R}^3_{t+h} = \widehat{R}^{oil}_t(1 + f^h_t - s_t - E_t(\pi^h_{t+h})) + \varepsilon_{t+h},$$

where $\varepsilon_{t+h} \sim NID(0, \sigma^2_{h|f})$ and $\sigma^2_{h|f}$ is the observed average forecast squared error at horizon h conditional on the oil futures spread based model; and again, $E_t(\pi^h_{t+h})$ is the inflation expectation calculated in Equation (2.4). Then, the density forecasts from the oil futures spread based model for real oil prices are the Monte Carlo integration based on the simulation of 1,000 simple Monte Carlo iterations.

B.3.4 A gasoline spread based model

The forecast based on the spread between the spot prices of gasoline and crude oil is specified as:

$$\widehat{R}^4_{t+h} = \widehat{R}^{oil}_t exp\{\widehat{\beta}[s^{gas}_t - s_t] + \varepsilon_{t+h} - E_t(\pi^h_{t+h})\},$$

where s^{gas}_t is the log of the nominal U.S. spot price of gasoline, s_t is the log nominal WTI spot price of crude oil, and $E_t(\pi^h_{t+h})$, similarly, is defined as the expected inflation rate over the next h periods.[2] $\widehat{\beta}$ is correspondingly the recursively least-square estimation based on:

$$\Delta s_{t+h|t} = \beta[s^{gas}_t - s_t] + \varepsilon_{t+h},$$

where $\varepsilon_{t+h} \sim NID(0, \sigma^2_{h|g})$; and $\sigma^2_{h|g}$ is the constant variance of the error term at the forecast horizon h. Therefore, the density forecasts of real oil prices based on this model, $g(\widehat{R}^4_{t+h})$, are calculated by the Monte Carlo integration simulated with 1,000 simple Monte Carlo iterations.

[2]Whereas crude oil prices are reported in U.S. cents per barrel, gasoline prices are reported in cents per gallon, which are then transformed to dollars per barrel. This involves multiplying by 42 (gallons/barrel) and dividing by 100 (cents/dollar).

B.3.5 A time varying parameter product spread model

The time-varying parameter model is a two state–space model, which recursively estimates the time-varying regression with its uncertainty. According to the estimates, the TVP model forecast is constructed directly through:

$$\widehat{R}^5_{t+h} = \widehat{R}^{oil}_t exp\{\widehat{\beta_{1t}}[s^{gas}_t - s_t] + \widehat{\beta_{2t}}[s^{heat}_t - s_t] - E_t(\pi^h_{t+h})\},$$

where the time-varying parameters $\widehat{\beta_{1t}}$ and $\widehat{\beta_{2t}}$, are estimated from:

$$\Delta s_{t+h} = \beta_{1t}[s^{gas}_t - s_t] + \beta_{2t}[s^{heat}_t - s_t] + e_{t+h},$$

where, $e_{t+h} \sim NID(0, \sigma^2_{h|TVP})$, $\sigma^2_{h|TVP}$ is the constant variance of the error term. This time-varying parameter model is motivated by Reeve & Vigfusson (2011) and developed by Baumeister et al. (2013), which employs independent Normal-Wishart prior and the Gibbs sampling algorithm for the forecasts. Moreover, $E_t(\pi^h_{t+h})$ is the inflation expectation introduced in Equation (2.4). The density forecasts use the 1,000 Gibbs posterior iterations of the time-varying parameters, excepting burn-in draws (5,000). Therefore, the size of the forecast density $g(\widehat{R}^5_{t+h})$ is 1,000.

B.3.6 No-change or random walk model

The no-change density forecast adds $\varepsilon_{t+h} \sim NID(0, \sigma^2_{h|N})$, where $\sigma^2_{h|N}$ is the observed average forecast squared error at horizon h conditional on the no-change forecasts, into the model as $\widehat{R}^6_{t+h} = R^{oil}_t + \varepsilon_{t+h}$ with the aim of generating the forecast densities $g(\widehat{R}^6_{t+h})$. The Monte Carlo integration is the forecasts simulated by 1,000 simple Monte Carlo iterations, conditional on the most recent data.

The density forecasts of BK are the average of the six densities forecasted at the 1- to 24-month horizons with size 1,000. More specifically, I have six random numerical densities at size 1,000, forecasted by the six models above respectively. Then I randomly choose a forecast from each density and calculate the mean of the six forecasts to replicate Baumeister & Kilian's (2015) equal-weight point forecast. Repeating this, I randomly calculate 1,000 means, which is defined as a density forecast of BK.

B.4 Convergence diagnostics

The results from MCMC runs with 1000 and 10,000 draws are not statistically different, where the convergence diagnostics seem satisfactory considering the very high dimensionality of the specifications. In this exercise, I diagnose the unconvergence using Geweke's (1992) IF and $Z-$score, as well as, Gelman & Rubin's (1992) PSRF, which are widely used in MCMC convergence diagnosing literature (see reviews in Cowles & Carlin (1996) and Roy (2019)).

IFs are the inverse of the relative numerical efficiency measure of Geweke (1992), for the posterior estimates of the parameters, and is calculated as $(1 + 2\sum_{k=1}^{\infty}\rho_k)$, where ρ_k is the k^{th} autocorrelation of the chain. In this application the estimate is performed using a 4 percent tapered window for the estimation of the spectral density at frequency zero. Values of the IFs below or around 20 are regarded as satisfactory, see for example in Primiceri (2005). Geweke's (1992) $Z-$score uses methods from spectral analysis to compare the mean difference between the first 10% and the last 50% iterations of a retained MCMC chain. Hence, if the $Z-$score were close to zero, there would be no evidence for unconvergence. Since the $Z-$score follows the normal distribution, its 5%-significance interval is [-1.96, 1.96]. Gelman & Rubin's (1992) PSRF is calculated using the two parallel chains, where I use the adjacent vintages $T-1$ and T in this exercise for all estimates, except the last time-varying coefficients at time T for the T-vintage, such as h_T and γ_T. PSRF is calculated as the square root of the ratio between an overestimated variance considering the variance between parallel chains and an underestimated variance which is calculated within the chains respectively. The cutoff point, suggested by Gelman & Rubin (1992), is 1.1. In other words, the parallel chains are diagnosed as 'no evidence of unconvergence' when PSRF$<$ 1.1.

In this exercise, I also test the null hypothesis that one specification's estimates across vintages from 1992:12 to 2016:12 are converged (or consistent). Therefore, one specification's convergence diagnostic is conditional on a very high dimensional estimates. For example, the parameters for TVPSVsmss(12) includes α, $\Omega^{\frac{1}{2}}$, γ_t, h_t, and R, where there are 105,888 estimates for vintage 2016:12 only. Considering all vintages from 1991:12 to 2016:12, there are 22,527,300 estimates in total, while the VAR contains the smallest size of estimates, which is 61,800 for all vintages.

Table B.1 presents the mean, min, max and the percentage of unconvergened estimates for the three diagnostics respectively. Under Geweke's (1992) $Z-$score, the percentage of unconverged estimates for all specifications are below 10%, and the mean of the absolute values for all $Z-$scores are strictly smaller than 0.1. From the view of IFs, all the mean of all estimates of IFs are less than 20 for all specifications, and the probabilities that IF $>$ 20 are smaller than 10%, excepting for model TVPsmss. However, when considering a larger threshold of 25, the probability of unconverged estimates is below 10% including the TVPsmss method. (See a discussion for the convergence of large dimensionality estimates in Primiceri (2005).)

Table B.1: Convergence diagnostics for all specifications

	IF				Z-score				PSRF			
Specifications	Mean	Min	Max	$P(IF > 20)$	Mean	Min	Max	$P(Z < -1.96) + P(Z > 1.96)$	Mean	Min	Max	$P(PSRF > 1.1)$
VAR	1.2312	0.3936	4.0317	0.00%	0.0013	-3.6837	3.3661	0.80%	1.0147	0.9995	4.4781	0.62%
SV	12.6123	0.3030	37.6800	2.77%	-0.0819	-11.1718	4.3027	0.23%	1.0827	0.9995	5.2672	9.91%
TVP	12.5502	0.2759	39.0466	2.54%	0.0183	-9.3226	5.3054	7.39%	1.3458	0.9995	14.8160	11.03%
TVPsmss	14.1749	0.3520	39.2386	21.60%	0.0677	-13.5797	8.3048	8.70%	1.4969	0.9995	90.8418	27.21%
TVPSV	15.1013	0.3469	37.2953	0.34%	-0.0474	-6.5837	5.7793	2.48%	1.0612	0.9995	3.2818	0.73%
TVPSVsmss(1)	13.5622	0.3377	38.4233	6.46%	-0.0595	-26.2229	24.5103	2.75%	1.1121	0.9995	6.4201	13.83%
TVPSVsmss(2)	13.4108	0.3987	38.0386	2.27%	-0.0396	-28.4348	33.4840	2.02%	1.1082	0.9995	6.0292	9.81%
TVPSVsmss(3)	13.3589	0.3575	36.7675	1.68%	-0.0585	-16.9409	7.6970	2.14%	1.0949	0.9995	5.2638	6.65%
TVPSVsmss(4)	13.5151	0.3642	36.9858	1.40%	-0.0688	-21.2937	18.8466	2.47%	1.0952	0.9995	5.6086	5.66%
TVPSVsmss(5)	13.6738	0.3693	37.4659	1.30%	-0.0735	-7.1556	9.0330	2.67%	1.0882	0.9995	5.6457	5.23%
TVPSVsmss(6)	13.8491	0.3610	37.1207	1.34%	-0.0860	-10.3858	12.7081	3.17%	1.0895	0.9995	5.0822	5.80%
TVPSVsmss(7)	14.0589	0.3560	37.6535	1.09%	-0.0786	-6.2233	10.6698	3.39%	1.0972	0.9995	5.0337	5.55%
TVPSVsmss(8)	14.0187	0.3086	37.2777	0.93%	-0.0944	-9.9592	7.4143	3.28%	1.0908	0.9995	4.8400	4.52%
TVPSVsmss(9)	14.0796	0.3637	37.3003	0.84%	-0.0721	-6.7994	9.0406	3.21%	1.0834	0.9995	4.8073	3.64%
TVPSVsmss(10)	14.3398	0.3228	37.7328	0.80%	-0.0557	-9.9488	10.9417	3.12%	1.0980	0.9995	6.1269	4.04%
TVPSVsmss(11)	14.6952	0.3728	37.1757	0.78%	-0.0487	-5.5527	10.5308	3.20%	1.0767	0.9995	6.8567	3.12%
TVPSVsmss(12)	14.6159	0.3503	37.9918	0.68%	-0.0606	-9.4254	17.5765	3.24%	1.0795	0.9995	8.0405	3.18%

Note: IF presents the inefficiency factor, which is the inverse of the relative numerical efficiency measure of Geweke (1992). Values of the IFs below or around 20 are regarded as satisfactory, see for example in Primiceri (2005). Geweke's (1992) $Z-$score using methods from spectral analysis to compare the mean difference between the first 10% and the last 50% iterations of a retained MCMC chain. Hence, if the $Z-$score closes to zero, there would be no evidence for unconvergence, and its 5%-significance interval is [-1.96, 1.96]. Gelman & Rubin's (1992) PSRF is calculated through the two parallel chains, where I use the adjacent vintages $T-1$ and T in this exercise for all estimates excepting the last time-varying coefficients at time T for the T-vintage, such as h_T and γ_T. The parallel chains are diagnosed as 'no evidence of unconvergence' when PSRF$<$ 1.1.

The Gelman & Rubin's (1992) PSRF statistic considers not only the convergence of the estimates, but their consistency across vintages. I utilise the corresponding estimates of vintages t and $t+1$, where $t \subseteq$ [1991:12, ..., 2016:11], excepting the estimates of time varying parameters on month $t+1$ of the vintage $t+1$, as parallel MCMC chains for the calculation of PSRF. In Table B.1, all specifications' mean PSRFs, except TVP, TVPsmss, TVPSVsmss(1), and TVPSVsmss(2), are smaller than 1.1, and the $P(PSRF > 1.1)$ are smaller than 10%. However, when considering a higher cutoff point of 1.2, which is often used in the literature (Vats & Knudson, 2018), more than 90% PSRFs for TVPSVsmss(1) and TVPSVsmss(2) are under the threshold. The TVP's and TVPsmss's PSRFs indicate some evidence of unconvergence, but the majority of the PSRFs estimates are smaller than 1.1, 88.79% and 72.79% for the two respectively. It is also necessary to mention that, there is no diagnostic that can provide precise tests for convergence, where all of the

diagnostics are an approximate guide on the evidence of the unconvergence (Cowles & Carlin, 1996; Roy, 2019).

All in all statistics, there is weak evidence for unconvergence, given a large sample of estimates.

B.5 A justification for MKLD accompanied with the extreme-forecasts-elimination process

MKLD accompanied with the extreme-forecasts-elimination process is used as an additional shrinkage method applied directly on the forecasts. It is necessary to justify if the resulting forecasts are from the model being estimated. Bearing in mind this consideration, comparing the out-of-sample (from the 1- to 24-month forecasting horizons) autocorrelation coefficients at one lag (without intercept) — AR(1) coefficients — I found that the method used in this exercise is equivalent to placing zero prior weight on a specific region of the out-of-sample random walks of time-varying-parameter estimates, whose standard error of the AR(1) coefficients are around three times the size of the accepted estimates. It is clear that the process is closely follows the estimates of specifications.

In this exercise I use TVPSVsmss(12) based on the 2016:12 vintage as an example. I collect all accepted and rejected draws of $B_{0,T+h}$, $\alpha^{\backslash B_0} + \Omega^{\frac{1}{2},\backslash B_0}\gamma_{T+h}^{\backslash B_0}$, and h_{T+h}, for h=1, ..., 24, respectively. (Note: I discard all time-invariant estimates, as they are the same for accepted and rejected draws.) Since the out-of-sample time-varying coefficients are drawn through the random walk process, I calculate recursive out-of-sample AR(1) coefficients, estimated using OLS. For example, the estimate b_h, drawn from $vec(B_{0,T+h}) = b_h vec(B_{0,T+h-1}) + e$ and $e \sim N(0,1)$, is based on samples for accepted and rejected forecasts respectively. The $vec(B_{0,T+h})$ is a vector $[[b0_{1,T+h|1}, \ldots, b0_{6,T+h|1}]', \ldots, [b0_{1,T+h|\iota}, \ldots, b0_{6,T+h|\iota}]', \ldots]'$, where there are 6 estimates in a small block for the elements in a lower triangular matrix $B_{0,T+h|\iota}$ for draw ι in the accepted and rejected samples respectively. The AR(1) coefficients are calculated for horizons $h = 1, \ldots, 24$ recursively.

Moreover, I also calculate the pooled OLS AR(1) estimates, based on all $h = 1, \ldots, 24$ together. For example, the pooled OLS AR(1) for $B_{0,T}$ is estimated

based on the independent and dependent variables x and y. The x vector contains $[vec(B_{0,T}), \ldots, vec(B_{0,T+23})]'$ and the corresponding y is $[vec(B_{0,T+1}), \ldots, vec(B_{0,T+24})]'$ for accepted and rejected samples respectively. Table B.2 illustrates the relationship between the time-varying coefficients from the accepted and rejected samples. The columns under recursive AR(1) estimates are the mean, min, max, standard deviation across h, as well as a ratio between the standard deviations for the draws rejected and accepted. The columns under pooled OLS AR(1) estimates are the estimates of the coefficients, standard errors and a ratio for the standard errors of the draws rejected relative to accepted.

Table B.2: Out-of-sample AR(1) coefficients for time-varying parameters

		Recursive AR(1) estimates					Pooled OLS AR(1) estimates		
TVPs	Sample	Mean	Min	Max	*Sd.*	$\frac{sd_R}{sd_A}$	AR(1) Coefficient	*Se.*	$\frac{se_R}{se_A}$
$B_{0,T}$	Accept	1.0002	0.9983	1.0025	1.2×10^{-3}	3.5082	1.0002	2.1×10^{-4}	2.7563
	Reject	1.0000	0.9922	1.0078	4.3×10^{-3}		1.0000	5.7×10^{-4}	
$\alpha^{\setminus B_0} + \Omega^{\frac{1}{2},\setminus B_0}\gamma_T^{\setminus B_0}$	Accept	1.0000	0.9996	1.0002	1.8×10^{-4}	3.4420	1.0000	2.6×10^{-5}	2.8337
	Reject	1.0006	0.9992	1.0016	6.0×10^{-4}		1.0006	7.2×10^{-5}	
h_T	Accept	0.9989	0.9982	1.0004	4.3×10^{-4}	1.8131	0.9989	5.0×10^{-5}	2.9407
	Reject	0.9989	0.9976	1.0008	7.8×10^{-4}		0.9989	1.5×10^{-4}	

Note: The recursive out-of-sample AR(1) coefficients are estimated through OLS, which are, for example, the estimate b_h from $vec(B_{0,T+h}) = b_h vec(B_{0,T+h-1}) + e$ and $e \sim N(0,1)$, based on samples stocked for accepted and rejected forecasts respectively. The $vec(B_{0,T+h})$ is a vector $[[b0_{1,T+h|1}, \ldots, b0_{6,T+h|1}]', \ldots, [b0_{1,T+h|\iota}, \ldots, b0_{6,T+h|\iota}]', \ldots]'$, where there are 6 estimates in a small block for the elements in a lower triangular matrix $B_{0,T+h|\iota}$ for draw ι in accepted and rejected samples respectively. The AR(1) coefficients are calculated for horizons $h = 1, \ldots, 24$ recursively. Moreover, I also calculate the pooled OLS AR(1) estimates, based on all $h = 1, \ldots, 24$ together. For example, the pooled OLS AR(1) for $B_{0,T}$ is estimated based on the independent and dependent variables x and y. The x is a vector contains $[vec(B_{0,T}), \ldots, vec(B_{0,T+23})]'$ and the corresponding y is $[vec(B_{0,T+1}), \ldots, vec(B_{0,T+24})]'$ for accepted and rejected samples respectively. The columns under recursive AR(1) estimates are the mean, min, max, standard deviation across h, as well as a ratio between the standard deviations for the draws rejected and accepted. The columns under pooled OLS AR(1) estimates are the estimate of the coefficient, standard error and a ratio for the standard errors of the draws rejected relative to accepted.

From Table B.2, there is no significant differences between the accepted and rejected draws at the mean for the contemporary correlations, time-varying parameters for lagged observations and stochastic volatilities, and the AR(1) coefficients are close to 1. This is acceptable, as their out-of-sample innovation follows a random walk. However, there is a significant difference for the standard divinations (*Sd.*) across the 1- to 24-month forecasting horizons. The *Sd.* for $B_{0,T}$ and $\alpha^{\setminus B_0} + \Omega^{\frac{1}{2},\setminus B_0}\gamma_T^{\setminus B_0}$, using rejected draws, is around 3.5 times of it using accepted draws. For h_T, the $\frac{sd_R}{sd_A}$ is 1.8131, which is still large. This is also observed when using the pooled OLS AR(1) estimation. The estimates of AR(1) coefficients are around 1 for all the three blocks of time-varying parameters using accepted and rejected samples respectively, while the estimated standard errors, *Se.*, between the two samples are different. The estimates of $\frac{se_R}{se_A}$ for the three blocks are about 2.8,

which is close to the definition of an extreme forecast that a value is more than three scaled median absolute deviations away from the median forecast.

Figure B.1 presents all recursive AR(1) coefficients for $B_{0,T}$, $\alpha^{\backslash B_0}+\Omega^{\frac{1}{2},\backslash B_0}\gamma_T^{\backslash B_0}$, and h_T respectively. It is confirmed that the AR(1) coefficients for rejected forecasts are more volatile than for accepted draws. The recursive AR(1) coefficients of the time-varying parameters for lagged variables using rejected draws are slightly above 1 over the 5- to 18-month horizons. These observations motivate an assumption of placing a stationary restriction on the VAR or TVP-VAR (as is commonly done) and discarding all non-stationary MCMC draws. Considering the out-of-sample forecasts, the selection of stationary VAR with constant coefficients is not difficult, while it is difficult to achieve where keeping time-varying parameters stable when allowing their out-of-sample innovation.

Two ways were tried: (1) randomly sampling time-varying parameters in out-of-sample, following a random-walk assumption, and accepting the stable VAR system at $T+h$. Unfortunately, due to the large number of estimates, the acceptance rate is strictly below 1% for 1-step ahead forecasts. (2) I allow time-varying parameters to innovate in out-of-sample, then only choose the draws' AR(1) coefficients strictly below 1. Nevertheless, at forecasting steps beyond 5, some of parameters are close to 0. Setting a lower bound in the selection of AR(1) coefficients is practically impossible. On the one hand, if the lower bound is close to 1, the acceptance rate will be too low to apply the innovation rule. On the other hand, if the bound is low, the problem of sustainability over a long forecasting horizons would be raised again. To set a constant AR(1) coefficient for the out-of-sample time-variation is not a good idea, as we do not know the AR(1) coefficient.

To summarise, the method used in this exercise for out-of-sample forecasts is equivalent to rejecting a specific region for the time-varying parameters, whose AR(1) coefficient standard deviation in the out-of-sample period (across the 1- to 24-month horizons) is around three times of those for the accepted draws.

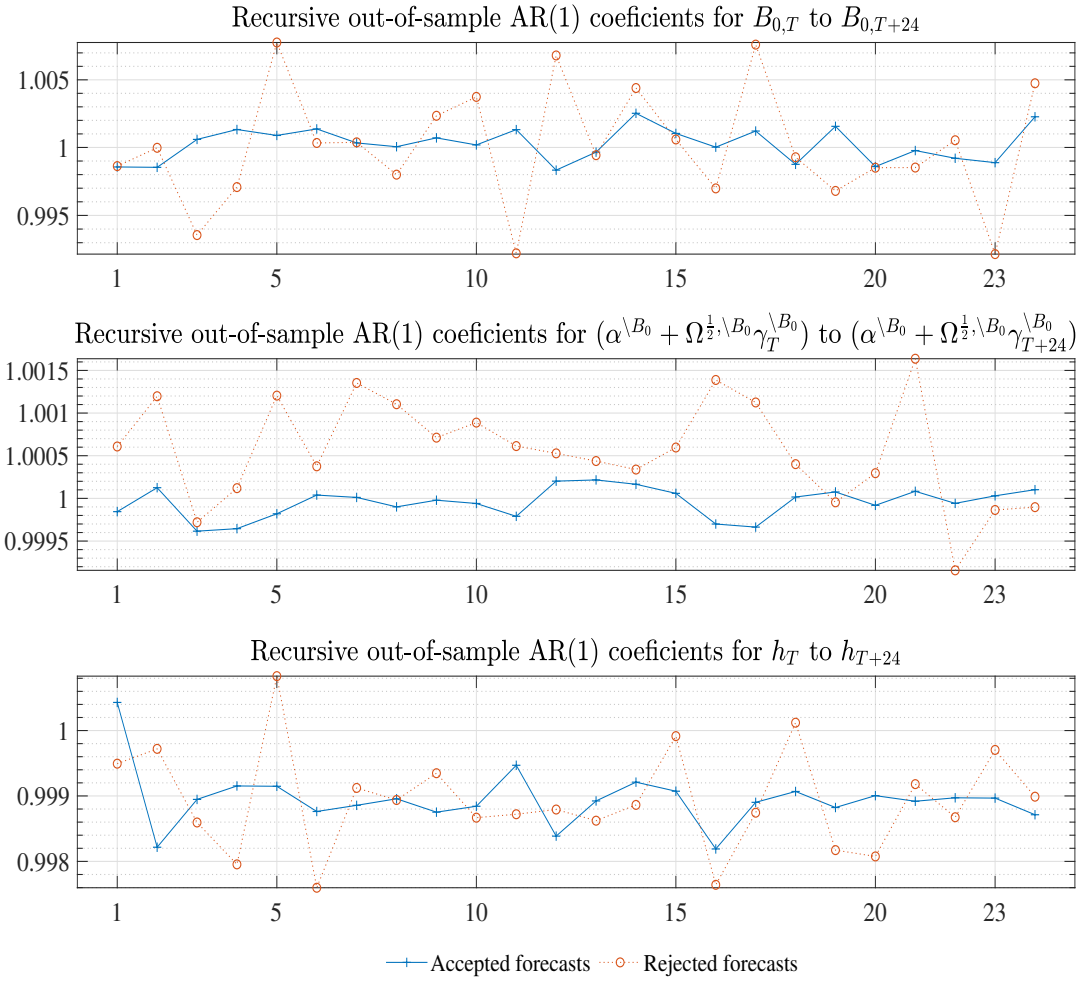

Figure B.1: Recursive out-of-sample AR(1) coefficients for time-varying parameters

Note: The recursive out-of-sample AR(1) coefficients are estimated through OLS, which are, for example, the estimate b_h from $vec(B_{0,T+h}) = b_h vec(B_{0,T+h-1}) + e$ and $e \sim N(0,1)$, based on samples stocked for accepted and rejected forecasts respectively. The $vec(B_{0,T+h})$ is a vector $[[b0_{1,T+h|1}, \ldots, b0_{6,T+h|1}]', \ldots, [b0_{1,T+h|\iota}, \ldots, b0_{6,T+h|\iota}]', \ldots]'$, where there are 6 estimates in a small block for the elements in a lower triangular matrix $B_{0,T+h|\iota}$ for draw ι in accepted and rejected samples respectively. The AR(1) coefficients are calculated for horizons $h = 1, \ldots, 24$ recursively.

B.6 A realistic signalling strategy for daily crude oil futures, conditional on real crude oil price density forecasts

Four possible scenarios in realistic world for WTI and Brent futures, when calculating the indicator of *long* or *short* positions, denoted as $\zeta^{i}_{d,h,M,\pi^c}|\cdot$, are shown in Figure B.2. During the two EIA reporting dates T and $T+1$, or vintages T and $T+1$, there could be zero, one or two LTDs. When there is one LTD and it is located in the "second half" (the right side of the vertical red-dashed line), shown as the top timeline in Figure B.2, the daily signal before the LTD for *long* or *short* WTI (Brent) futures is conditional on both the implied log real RAC in $T+h$, which is denoted as $E_d(r^{oil|i}_{T+h})$ and based on daily WTI (Brent) futures with maturation of h observed on day d, as well as, the h-month ahead density forecasts from specifications in vintage T, $g(r^{oil}_{T+h|M})$. I denote it as $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h|M})$. After the LTD, the most recent futures contract (CL1 or CO1) becomes the contract maturation in one month later. Therefore, I use the density forecasts in $h+1$ for calculating the *long–short* signal, which is $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h+1|M})$.

When the LTD is located in the "first half", shown in the second timeline, the calculation of *long–short* signals will use the futures maturation in $h+1$ months later, where $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h+1})\&g(r^{oil}_{T+h|M})$. This is because the most recent futures (CL1 or CO1) is too soon to maturate and the density forecast at the 1-month horizon is far later than the LTD. After the LTD, I use $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h|M})$ again. The same logic can be applied in the scenario that there are two LTDs shown in the third timeline. If the day locates before the LTD and within the range of the "first half", the *long–short* signal will be calculated as $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h+1})\&g(r^{oil}_{T+h|M})$. Meanwhile, if the day is after the LTD and in the "second half", the signal will be $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h+1|M})$. Between the two LTDs, the calculation returns to $\zeta^{i}_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h|M})$, which is also used for the situation when there is no LTD between the two EIA reporting dates (vintages T and $T+1$), as shown in the last timeline. For the purpose of an evaluation, I calculate all the excess returns on the days between the two EIA reporting dates.

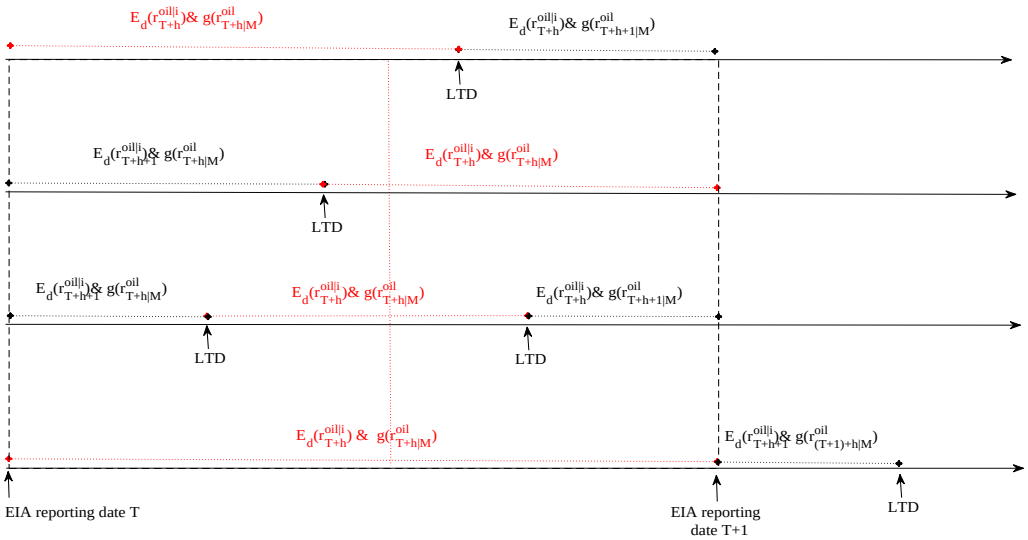

Figure B.2: Four scenarios for the calculation of $\zeta_{T,h}$

The calculation of $\zeta^i_{d,h,M,\pi^c}|\cdot$ depends on the density forecasts and daily implied log real RAC prices. For example, $\zeta^i_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h|M})$ is a strategic indicator (signal) at horizon h on day d based on the forecasts in the vintage T:

$$\zeta^i_{d,h,M,\pi^c}|E_d(r^{oil|i}_{T+h})\&g(r^{oil}_{T+h|M}) = \begin{cases} 1\ (Long) & \text{if } P[g(r^{oil}_{T+h|M}) < E_d(r^{oil|i}_{T+h})] >= \pi^c) \\ \text{-1}\ (Short) & \text{if } P[g(r^{oil}_{T+h|M}) > E_d(r^{oil|i}_{T+h})] >= \pi^c) \\ 0\ (No\ Actions) & Otherwise, \end{cases} \tag{B.4}$$

where $g(r^{oil}_{T+h|M})$ is the density forecasts of real log RAC using the specification M. $E_d(r^{oil|i}_{T+h})$ is the implied log-level real RAC spot price on day d, h months later, conditional on the WTI (or Brent) measure, and their nominal daily futures price on day d with maturity in h months is denoted as $f^i_{h|d}$, where:

$$E_d(r^{oil|i}_{T+h}) = log(\frac{R^i_T(1 + f^i_{h|d} - s^i_d - E_T(\pi^h_{T+h}))}{spread^{\text{i, RAC}}_T}). \tag{B.5}$$

where $spread^{\text{i, RAC}}_T$ is the spread between the last observation of the nominal price WTI (Brent) measure in vintage T relative to the last observation of the nominal price of RAC, such as:

$$spread_T^{\mathrm{i,\ RAC}} = \frac{p_T^{\mathrm{i}}}{p_T^{RAC}};$$

$R_T^i(1 + f_{h|d}^i - s_d^i - E_T(\pi_{T+h}^h))$ is widely used for real crude oil price forecasting literature (see Baumeister et al., 2018; Baumeister & Kilian, 2014; and Baumeister & Kilian, 2015), namely futures-spread forecasts; the R_T^i is the last observed real WTI (Brent) in vintage T; and s_d is the corresponding log-level daily WTI (Brent) spot price. Echoing Baumeister & Kilian's (2015) definition of $E_T(\pi_{T+h}^h)$, the expected U.S. inflation, as the average U.S. CPI inflation available at time T, where the averaging begins in 1986:07, is hence:

$$E_T(\pi_{T+h}^h) = [1 + \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}}^{\overline{\tau}} (ln(CPI_{t+1}) - ln(CPI_t))]^h - 1,$$

where $t = [\underline{\tau}, ..., \overline{\tau}]$, $\underline{\tau}$=1986:07 and $\overline{\tau}$ is the last observation in time T.

Moreover, the probability of that the forecast $g(r_{T+h|M}^{oil})$ is smaller than the log level transferred real futures price of RAC, $E_d(r_{T+h}^{oil|i})$, from WTI or Brent measures is notated as:

$$P[g(r_{T+h|M}^{oil}) < E_d(r_{T+h}^{oil|i})] = \int_{-\infty}^{E_d(r_{T+h}^{oil|i})} p(u)du,$$

where $p(.)$ is the probability based on density forecasts $g(r_{T+h|M}^{oil})$. Thereby,

$$P[g(r_{T+h|M}^{oil}) > E_d(r_{T+h}^{oil|i})] = 1 - P[g(r_{T+h|M}^{oil}) < E_d(r_{T+h}^{oil|i})].$$

The π^c is a threshold, reflecting the risk-aversion level, and decided by the relevant decision maker who participates in the financial market for the purpose of hedging or speculation. From the point forecast view, the market participants can only set $\pi^c = 50\%$, which is risk-nature. Thereafter, the *long–short* strategy based on density forecasts would be more flexible, and may help investors to recover from difficulties when using the point forecasts for signalling.

B.7 The calculation of excess returns

The calculation of futures returns in this study is broadly consistent with the financial literature (see Gorton et al., 2012; Hong & Yogo, 2012; and Bakshi et al.. 2017). However, the procedure for constructing futures returns in this exercise accounts for the last trading day (LTD), but not the first notice day (FND) in Bakshi et al. (2017), in order to avoid a physical delivery from the counterparty. This is because after the LTD, the most recent crude oil futures contract (CL1) will be stopped being able to trade in the market. Moreover, the FND is approximately 10 days after the LTD for WTI, and which has been the same day as the LTD for Brent futures since the contract from 1992:11 (COX92). In this chapter, the return of *long–short* futures daily position, which is taken on day d for the contract with maturity h, until the LTD is defined as:

$$r^i_{h|d} = \begin{cases} \dfrac{f^i_{h|LTD} - f^i_{h|d}}{f^i_{h|d}} + r^f_d & \text{if take } \textit{long} \text{ position} \\ -\dfrac{f^i_{h|LTD} - f^i_{h|d}}{f^i_{h|d}} + r^f_d & \text{if take } \textit{short} \text{ position,} \end{cases} \tag{B.6}$$

where i represents the WTI and Brent measures, respectively: $i \subseteq$ [WTI, Brent]; $f^i_{h|d}$ and $f^i_{h|LTD}$ are the nominal oil futures with the maturity h observed on day d and the corresponding LTD, respectively; and r^f_d indicates the interest earned on the fully collateralised futures position (see for example Bakshi et al.'s (2017) Eq'uation (1) or Hong & Yogo's (2012) Equation (14)). Hence, the excess return with the maturity h, denoted as $er^i_{h|d}$, is calculated as:

$$er^i_{h|d} = r^i_{h|d} - r^f_d.$$

Aiming to make the $er^i_{h|d}$ at different horizons are comparable, I standardise the $er^i_{h|d}$ to be a monthly return, denoted as $er^i_{\overline{m}|d,h}$, via:

$$er^i_{\overline{m}|d,h} = \sqrt[h]{1 + er^i_{h|d}} - 1. \tag{B.7}$$

This is because the h-month compound excess return should satisfy:

$$(1 + er^i_{\overline{m}|d,h})^h = 1 + er^i_{h|d}.$$

Following this exercise, I respectively calculate daily returns at the h-month

horizon, where h=1, ..., 24, between the two EIA data release dates, which distinguishes the vintages T and $T+1$, of the *Monthly Energy Review*, thus updating the oil supply, demand and inventories data used for density forecasts. Hence, there are three assumptions: (1) the investor can afford any additional deposit (initial margin) charge for holding the futures position until the LTD; (2) the transition fee is negligible, as we do not have data for real-time transition fees; and (3) the duration is the same for daily returns calculated with the $h-$month maturity. For example, when daily returns calculated for $h = 12$, each of them will be treated as a 12-month cumulative return. Hence, I can cumulate the $er^i_{\overline{m}|d,h}$ between T and $T+1$ for the 1- to 24-month horizons independently, conditional on forecasts at the corresponding horizons.

B.8 The cumulative excess returns for WTI and Brent measures in futures market

WTI and Brent measures' cumulative excess returns at horizon h with risk-aversion level π^c for specification M, denoted as cer^i_{h,M,π^c} and calculated via

$$cer^i_{h,M,\pi^c} = \sum_d er^{i,long}_{\overline{m}|d,h} \times \zeta^i_{d,h,M,\pi^c}|\cdot,$$

where, $er^{i,long}_{\overline{m}|d,h}$ is the excess return of a *long* position. According to Equation (B.6), the excess return for a *short* position is $-er^{i,long}_{\overline{m}|d,h}$. This is a good measure for checking the model's ability to generate profit. However, cer^i_{h,M,π^c} is sensitive to an extreme excess return $er^i_{\overline{m}|d,h}$. For example, the positive cer^i_{h,M,π^c} may be caused by a large gain at one time. And the extreme $er^i_{\overline{m}|d,h}$ offers less opportunity to learn from a forecasting perspective. Consequently, if there is one extreme positive $er^i_{\overline{m}|d,h}$ that can drive the cer^i_{h,M,π^c} to be positive, then the cer^i_{h,M,π^c} will basically measure if the density forecast can catch the extreme observations. In other words, the narrower the density, the higher the probability of identifying the extreme, and the higher cer^i_{h,M,π^c}. The illustration of cer^i_{h,M,π^c} with $\pi^c = 50\%$, $\pi^c = 68\%$, and $\pi^c = 85\%$ during the 1992:01–2016:12 period at the 1-, 3-, 6-, 12-, 18-, and 24-month horizons, for WTI and Brent measures indicate five observations in Figures B.3 and B.4:

- VAR's, BVAR's and BK's cer^i_{h,M,π^c} are not sensitive to π^c, because their density forecasts are narrow, please (see Section 3.4.3 for a detailed discussion) and the cer^i_{h,M,π^c} is generally conditional on their main forecasts. Therefore, their abnormal returns relative to other models using $\pi^c = 68\%$ and $\pi^c = 85\%$

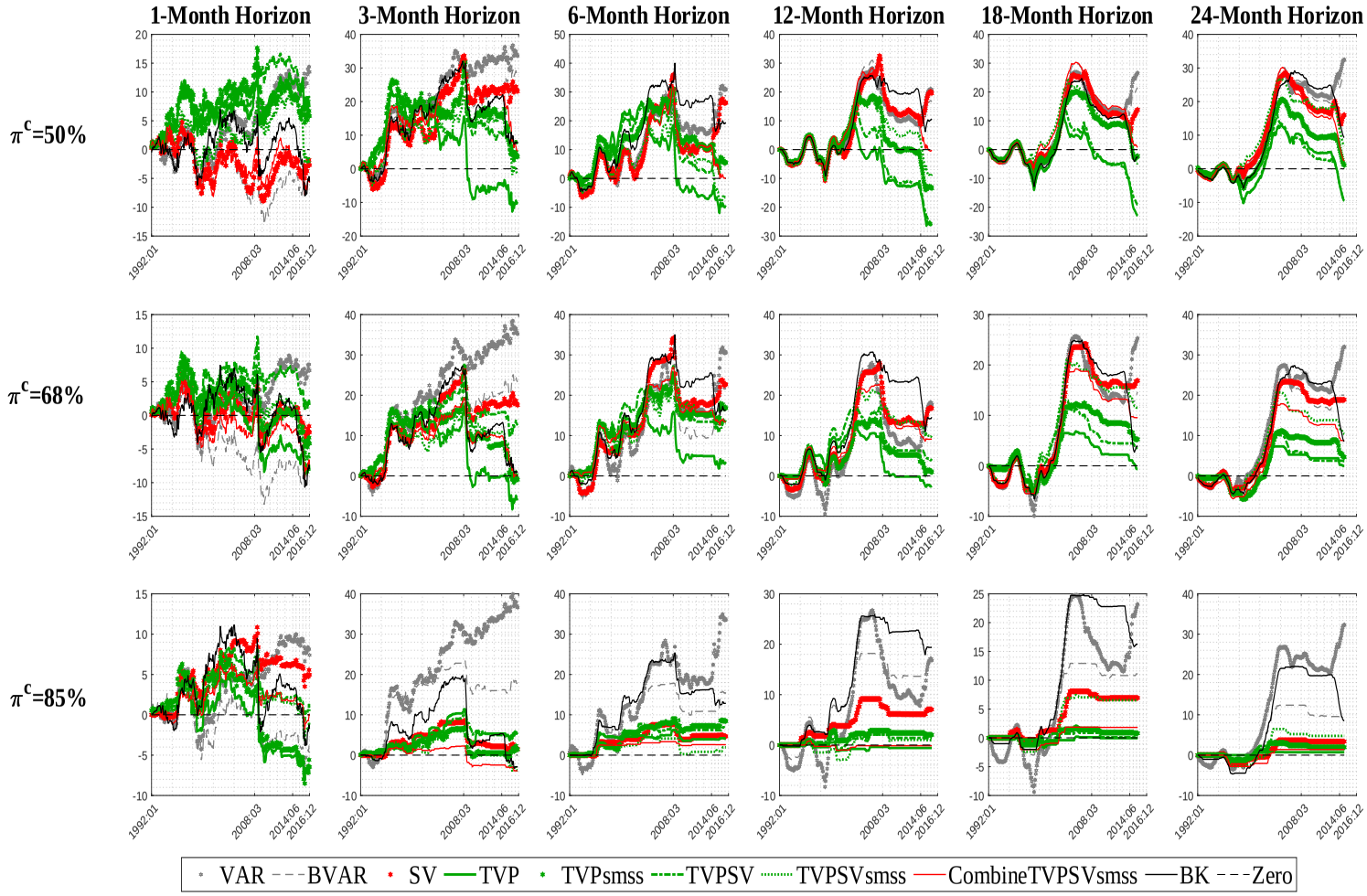

Figure B.3: Financial market cumulative excess returns for the WTI measure in the period 1992:01–2016:12

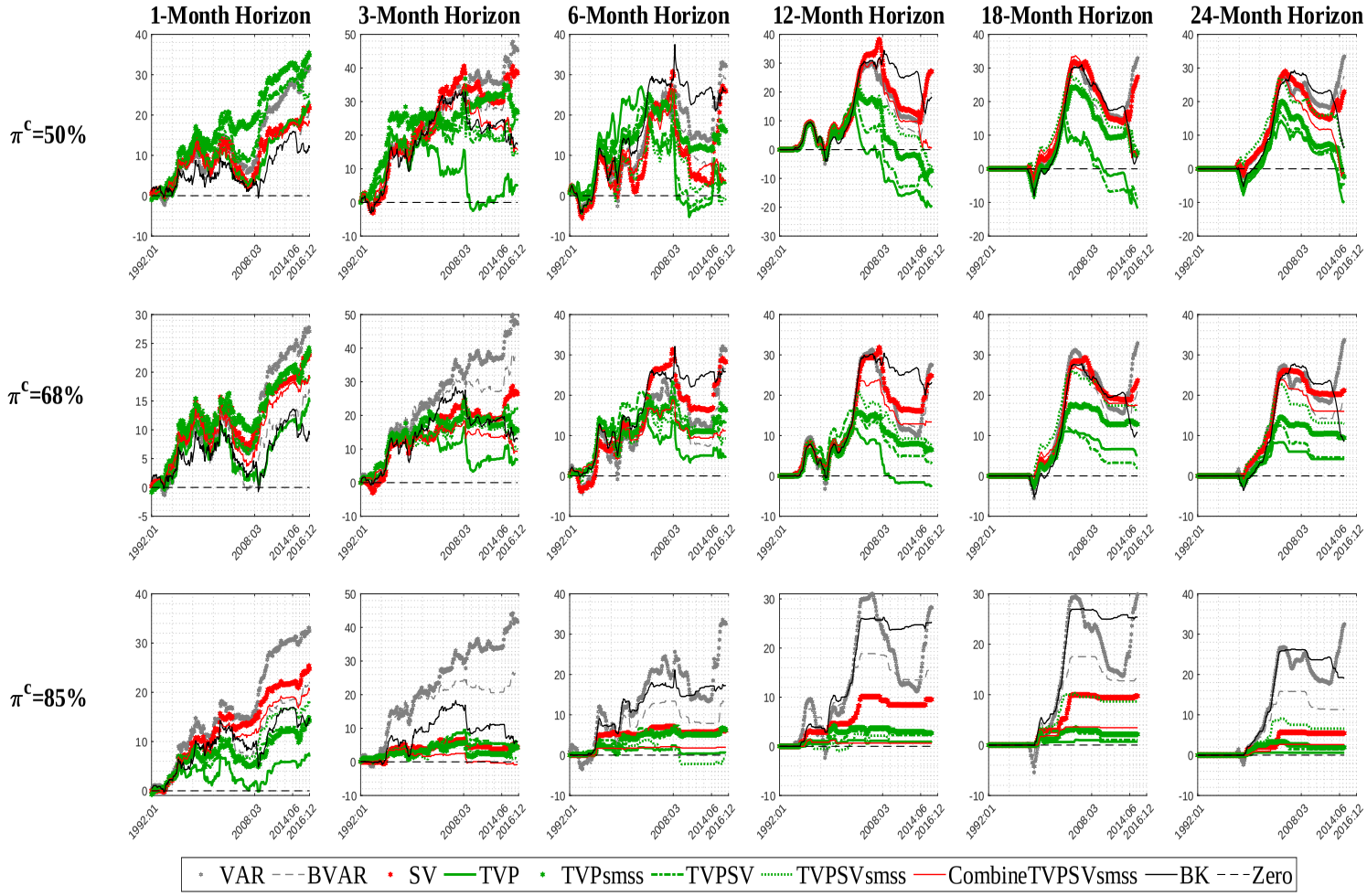

Figure B.4: Financial market cumulative excess returns for the Brent measure in the period 1992:01–2016:12

cannot imply that their forecasts are superior.

- The increase of π^c can improve the probability of positive $er^i_{\overline{m}|d,h}$ for the SV, TVP, TVPsmss, TVPSV, TVPSVsmss, and CombineTVPSVsmss models. For example, the rise of π^c from 50% to 85% improves SV's cer^i_{h,M,π^c} from negative to positive over the 1992:01–2016:12 period at horizon 1 for the WTI measure.

- Although the cer^i_{h,M,π^c} generated by SV at horizon 1 is negative for WTI measure using $\pi^c = 50\%$ and $\pi^c = 68\%$, SV's cer^i_{h,M,π^c} can dominate BK's at all horizons for both WTI and Brent measures using the same π^c.

- The cer^i_{h,M,π^c} based on VAR, TVP, and TVPSV, whose density forecasts are relatively more diffuse than other models' (see Section 3.4.3 for a detailed illustration), is positive and higher than other models' at the 1-month horizon for both WTI and Brent measures. However, their cer^i_{h,M,π^c} are dominated by SV's and CombineTVPSVsmss's at the majority of horizons lager than 1, using $\pi^c = 50\%$, $\pi^c = 68\%$, and $\pi^c = 85\%$ for the two measures.

- The cumulative excess returns of all models for the Brent measure are higher than for WTI at 1-month horizon using $\pi^c = 50\%$, $\pi^c = 68\%$, and $\pi^c = 85\%$.

B.9 Probabilities for negative excess returns

The cumulative excess return is sensitive to extreme observations. Instead, this appendix will employ the probability of negative $er^i_{\overline{m}|d,h}$ at the 1-, 3-, 6-, 12-, 18-, and 24-month horizons over the 1992:01–2016:12 period for evaluating the specifications. Tables B.3 and B.4 present the probability of negative returns for all models using $\pi^c = 50\%$, $\pi^c = 68\%$, $\pi^c = 75\%$, $\pi^c = 85\%$, $\pi^c = 90\%$, and $\pi^c = 95\%$ for the WTI and Brent measures, respectively. There are six observations below.

- Under $\pi^c = 68\%$, $\pi^c = 75\%$, and $\pi^c = 85\%$, SV performs better than all other specifications. For example, SV generates the lowest negative $er^i_{\overline{m}|d,h}$ probabilities among all models at the 6-, 12-, and 18-month horizons using $\pi^c = 68\%$ and $\pi^c = 75\%$ for both WTI and Brent measures.

- The probabilities of BVAR, BK and particularly VAR are not sensitive to π^c, due to their narrow density forecasts, please (see Section 3.4.3 for a detailed discussion of the densities' calibration), while the probability of negative excess returns generated by the other methods presents a decreasing trend when

increasing π^c. Therefore, in practice the narrow density forecasts are not preferred to the density forecasts of other models for risk control through π^c.

- Under $\pi^c = 50\%$, the specifications SV, TVP, TVPsmss, TVPSV, TVPSVsmss, and CombineTVPSVsmss result in higher negative $er^i_{\overline{m}|d,h}$ probabilities than the narrow density forecasts generated by VAR, BVAR, and BK. For instance, the probabilities of negative $er^i_{\overline{m}|d,h}$ generated by all specifications excepting VAR, BVAR, and BK are even above 50% at the 18-month horizon for the WTI measure.

- Under $\pi^c = 90\%$, SV generates the lowest probabilities of negative $er^i_{\overline{m}|d,h}$ among all specifications at all horizons excepting the 1-month for Brent. For the WTI measure, SV can generate the lowest probabilities at 1- and 3-month horizons, but it performs worse at the 18- and 24-month horizons, which are larger than 50%. Additionally, the wide density forecasts for the WTI measure generated through TVP can generate the lowest probabilities of negative $er^i_{\overline{m}|d,h}$.

- Under $\pi^c = 95\%$, there are limited *long–short* positions have been taken at 6-, 12-, 18- and 24-month horizons. Hence there are not available (NaN), 0 and 1 observations. Moreover, CombineTVPSVsmss performs better than all other models at the 1-month horizon for WTI, while TVPSV performs better than other specifications at both the 1- and 3-month horizons for the Brent measure.

- The probability of negative $er^i_{\overline{m}|d,h}$ using the Brent measure is smaller than using WTI. As an example, the average negative $er^i_{\overline{m}|d,h}$ probability using $\pi^c = 68\%$ at the 1-month horizon for WTI for all specifications is approximately 0.46, while the probability for the Brent measure is approximately 0.39.

Table B.3: Probability of negative financial market excess returns for the WTI measure, 1992:01–2016:12

	Forecast Horizons					
Specifications/MH	1-Month	3-Month	6-Month	12-Month	18-Month	24-Month
			$\pi^c = 50\%$			
VAR	0.47	0.44	0.44	**0.44**	**0.48**	**0.40**
BVAR	0.47	**0.43**	0.43	0.45	**0.48**	0.42
SV	0.49	**0.43**	0.43	0.45	0.50	0.46
TVP	0.46	0.45	0.52	0.55	0.57	0.55
TVPsmss	0.49	0.46	0.48	0.52	0.50	0.51
TVPSV	0.46	0.46	0.48	0.55	0.59	0.57
TVPSVsmss	0.47	0.46	0.51	0.49	0.53	0.49
CombineTVPSVsmss	0.50	0.46	0.50	0.47	0.53	0.52
BK	**0.44**	0.45	**0.42**	0.46	**0.48**	0.48
			$\pi^c = 68\%$			
VAR	0.46	0.44	0.43	0.45	0.47	**0.40**
BVAR	0.48	0.46	0.43	0.45	0.50	0.46
SV	0.47	0.42	**0.41**	**0.42**	**0.42**	0.43
TVP	**0.43**	0.48	0.49	0.46	0.53	0.51
TVPsmss	0.46	**0.41**	0.43	0.49	0.50	0.53
TVPSV	0.44	0.47	0.48	0.46	0.51	0.53
TVPSVsmss	0.47	0.46	0.45	0.45	0.50	0.50
CombineTVPSVsmss	0.46	0.47	0.45	0.46	0.50	0.53
BK	0.46	0.45	**0.41**	0.43	0.48	0.47
			$\pi^c = 75\%$			
VAR	0.44	0.44	0.43	0.45	0.49	**0.40**
BVAR	0.46	0.44	0.41	0.44	0.48	0.45
SV	0.43	**0.41**	**0.34**	**0.37**	**0.41**	0.41
TVP	0.41	0.48	0.44	0.46	0.45	**0.40**
TVPsmss	0.41	0.44	0.44	0.43	0.43	0.49
TVPSV	**0.38**	0.46	0.45	0.39	0.48	0.49
TVPSVsmss	0.42	0.48	0.48	0.41	0.47	0.47
CombineTVPSVsmss	0.42	0.46	0.42	0.41	0.46	0.45
BK	0.45	0.46	0.41	0.41	0.49	0.46
			$\pi^c = 85\%$			
VAR	0.41	0.43	0.44	0.46	0.48	0.41
BVAR	0.46	0.44	0.43	0.42	0.43	0.45
SV	0.41	0.51	**0.33**	**0.32**	**0.29**	0.43
TVP	0.38	**0.37**	0.54	0.38	0.40	**0.10**
TVPsmss	0.39	0.47	0.43	0.44	0.50	0.41
TVPSV	**0.37**	0.45	0.48	0.39	0.51	0.41
TVPSVsmss	0.40	0.52	0.50	0.37	0.39	0.48
CombineTVPSVsmss	0.38	0.52	0.50	0.35	0.33	0.44
BK	0.42	0.45	0.37	0.36	0.39	0.44
			$\pi^c = 90\%$			
VAR	0.38	0.43	0.42	0.46	0.49	0.41
BVAR	0.43	0.44	0.44	0.34	0.44	0.48
SV	**0.34**	**0.41**	0.44	0.42	0.53	0.64
TVP	0.38	**0.41**	0.48	0.57	**0.33**	**0.00**
TVPsmss	0.40	0.47	**0.34**	0.52	0.56	0.25
TVPSV	0.36	0.49	0.50	0.44	0.57	**0.13**
TVPSVsmss	0.38	0.51	0.61	0.48	0.42	0.45
CombineTVPSVsmss	0.35	0.53	0.40	0.45	0.50	0.63
BK	0.39	0.47	0.35	**0.33**	0.39	0.44
			$\pi^c = 95\%$			
VAR	0.36	0.43	0.44	0.46	0.50	0.41
BVAR	0.42	**0.40**	0.37	0.33	0.41	0.55
SV	0.35	0.43	**0.20**	0.33	**0.00**	NaN
TVP	0.38	0.47	**0.20**	1.00	0.50	NaN
TVPsmss	0.37	0.52	0.36	**0.31**	0.54	0.11
TVPSV	0.40	0.38	0.54	0.67	0.33	**0.00**
TVPSVsmss	0.37	0.52	0.50	0.46	0.40	**0.08**
CombineTVPSVsmss	**0.34**	0.56	**0.00**	NaN	1.00	1.00
BK	0.37	0.42	0.38	0.32	**0.31**	0.35

Note: MH represents monthly forecast horizons. π^c denotes the choice of probability to conduct a *long* or *short* stratagem, representing the financial markets participants' risk-aversion level. Boldface indicates the lowest diagnostic statistics, including 0. 'NaN' indicates that there is no *long–short* positions during the 1992:01–2016:12 period.

Table B.4: Probability of negative financial market excess returns for the Brent measure, 1992:01–2016:12

	Forecast Horizons					
Specifications/MH	1-Month	3-Month	6-Month	12-Month	18-Month	24-Month
			$\pi^c = 50\%$			
VAR	0.41	0.40	**0.43**	0.45	**0.45**	**0.33**
BVAR	0.43	0.40	0.47	0.43	0.47	0.38
SV	0.42	**0.37**	0.45	0.43	0.48	0.40
TVP	0.43	0.47	0.47	0.50	0.53	0.52
TVPsmss	0.40	0.42	0.49	0.49	0.52	0.52
TVPSV	**0.39**	0.47	0.47	0.50	0.54	0.51
TVPSVsmss	0.41	0.45	0.49	0.51	0.54	0.45
CombineTVPSVsmss	0.42	0.43	0.48	0.48	0.52	0.49
BK	0.45	0.40	0.44	**0.40**	0.48	0.44
			$\pi^c = 68\%$			
VAR	0.39	0.40	0.45	0.45	0.44	0.33
BVAR	0.39	0.39	0.47	0.44	0.51	0.37
SV	0.37	**0.36**	**0.42**	**0.36**	**0.41**	**0.30**
TVP	0.39	0.46	0.49	0.46	0.51	0.38
TVPsmss	**0.36**	0.41	0.45	0.46	0.45	0.37
TVPSV	0.40	0.42	0.49	0.49	0.55	0.45
TVPSVsmss	0.37	0.42	0.48	0.45	0.46	0.38
CombineTVPSVsmss	0.39	0.43	0.45	0.46	0.47	0.37
BK	0.44	0.38	0.43	0.38	0.45	0.41
			$\pi^c = 75\%$			
VAR	**0.36**	0.41	0.45	0.44	0.45	0.33
BVAR	0.38	0.41	0.45	0.41	0.50	0.38
SV	0.37	0.40	**0.38**	**0.34**	**0.38**	**0.23**
TVP	**0.36**	0.45	0.49	0.45	0.51	0.27
TVPsmss	**0.36**	0.41	0.46	0.44	0.44	0.28
TVPSV	0.37	0.42	0.49	0.42	0.50	0.40
TVPSVsmss	0.37	0.42	0.48	0.45	0.46	0.34
CombineTVPSVsmss	0.38	**0.39**	0.42	0.35	0.39	0.24
BK	0.44	**0.39**	0.40	0.38	0.44	0.41
			$\pi^c = 85\%$			
VAR	**0.34**	**0.38**	0.45	0.44	0.46	0.34
BVAR	0.38	0.40	0.44	0.34	0.41	0.30
SV	**0.34**	0.44	0.34	0.28	0.24	0.13
TVP	0.38	**0.38**	0.56	0.30	0.53	0.10
TVPsmss	0.38	0.48	0.47	0.44	0.40	0.24
TVPSV	0.39	**0.38**	0.45	0.39	0.59	0.29
TVPSVsmss	0.39	0.44	0.47	0.40	0.40	0.33
CombineTVPSVsmss	0.37	0.45	**0.33**	0.35	**0.19**	**0.05**
BK	0.43	0.42	0.37	**0.26**	0.27	0.31
			$\pi^c = 90\%$			
VAR	**0.29**	0.40	0.45	0.43	0.47	0.34
BVAR	0.35	0.39	0.47	0.31	0.45	0.20
SV	0.34	**0.38**	**0.25**	**0.29**	**0.22**	**0.00**
TVP	0.39	**0.38**	0.50	0.50	0.50	**0.00**
TVPsmss	0.35	0.42	0.36	0.41	0.59	**0.17**
TVPSV	0.35	**0.38**	0.45	0.42	0.52	0.19
TVPSVsmss	0.39	0.43	0.56	0.52	0.39	0.21
CombineTVPSVsmss	0.36	0.46	**0.00**	0.50	**0.00**	**0.00**
BK	0.38	0.40	0.39	0.30	0.27	0.23
			$\pi^c = 95\%$			
VAR	0.31	0.40	0.43	0.42	0.48	0.34
BVAR	0.35	0.38	0.43	**0.31**	0.37	0.27
SV	0.27	0.33	0.67	**0.00**	**0.00**	**0.00**
TVP	0.35	0.24	**0.00**	1.00	NaN	NaN
TVPsmss	0.33	0.45	**0.20**	0.33	0.60	**0.11**
TVPSV	**0.24**	**0.22**	0.50	0.40	0.60	**0.00**
TVPSVsmss	0.34	0.48	0.53	0.58	0.33	**0.00**
CombineTVPSVsmss	0.28	0.50	NaN	NaN	NaN	NaN
BK	0.35	0.43	0.39	**0.31**	**0.30**	0.14

Note: MH represents monthly forecast horizons. π^c denotes the choice of probability to conduct a *long* or *short* stratagem, representing the financial markets participants' risk-aversion level. Boldface indicates the lowest diagnostic statistics, including 0. 'NaN' indicates that there is no *long–short* positions during the 1992:01–2016:12 period.

Appendix C

Technical Details and Additional Results in Chapter 4

C.1 Reference priors for Σ_u and B

Prior for $\Sigma_u|A$**.** According to Equation (4.9), prior information of $\Sigma_u|A$ uses the gamma distribution $\Gamma(\kappa_i, \tau_i)$. Hence, the expected mean $E(\sigma_{ii})$ and variance $V(\sigma_{ii})$ before see the data, where σ_{ii} is the ii element in Σ_u, are:

$$\begin{cases} E(\sigma_{ii}) = \frac{\kappa_i}{\tau_i} \\ V(\sigma_{ii}) = \frac{\kappa_i}{\tau_i^2}. \end{cases} \tag{C.1}$$

It is clear that the increase of τ_i leads to a more confident belief on prior (or a tighter prior). Following Baumeister & Hamilton (2019), I set $\kappa_i = 2$, and τ_i is conditional on A, denoted as $\tau_i(A)$. Recalling Equation (4.2), let $\hat{\epsilon}_{it}$ denote the p^{th}-order univariate autoregression estimation residual of the series i, and $\hat{\Sigma}_\epsilon$ stands for the sample variance matrix of these univariate residuals, whose element $s_{ij} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_{it}\hat{\epsilon}_{jt}$. As $\Sigma_\epsilon = A^{-1}\Sigma_u(A^{-1})'$,

$$E(\Sigma_u) = A\hat{\Sigma}_\epsilon A'.$$

In other words, the ii element in $E(\Sigma_u)$, $E(\sigma_{ii}) = a_i\hat{\Sigma}_\epsilon a_i'$, where a_i' is the i^{th} row of A. Recall Equation (C.1),

$$\tau_i(A) = a_i'\hat{\Sigma}_\epsilon a_i.$$

Prior for $B|A, \Sigma_u$**.** Echoing Equation (4.10), the expectation and variance of B are:

$$\begin{cases} E(b_i) = m_i \\ V(b_i) = \sigma_{ii} M_i, \end{cases} \tag{C.2}$$

where b_i is the i^{th} row of B. Following Baumeister & Hamilton (2015, 2018, 2019), m_i depends on A denoted as $m_i(A)$, while M_i does not. The prior is Minnesota type, where the individual series are assumed to be random walk (Doan et al., 1984). This exercise follows Sims & Zha (1998) to present the random-walk assumption.

As shown in Equation (4.2), $\Phi = A^{-1}B$, assuming $E(\Phi) = \varphi$, where

$$\underset{[n\times(n*p+1)]}{\varphi} = \left[\underset{(n\times n)}{I_n} \quad \underset{[n\times(n*p+1-n)]}{0} \right].$$

From $B = A\Phi$, $E(B|A) = E(A\Phi) = A\varphi$. Hence, its mean of the i^{th} row is:

$$E(b_i|A) = m_i(A) = \varphi^{'} a_i.$$

Meanwhile, as proposed in Doan et al. (1984) modified by Sims & Zha (1998), the shrinkage of B is conditional on the assumption that parameters of higher lags are more confident to be zero than lower lags.

$$\underset{(1\times p)}{v_1^{'}} = (1/1^{2\lambda_1} \ \ldots, \ 1/(p^{2\lambda_1})),$$

$$\underset{(1\times n)}{v_2^{'}} = (s_{11}^{-1}, \ \ldots, \ s_{nn}^{-1}),$$

$$v_3 = \lambda_0^2 \begin{bmatrix} v_1 \bigotimes v_2 \\ \lambda_3^2 \end{bmatrix}.$$

where $\bigotimes$ indicates Kronecker product. Then M_i is taken to be a diagonal matrix whose (r, r) element is the r^{th} element of v_3, where

$$M_{i,rr} = v_{3r}.$$

It is clear that the Shrinkage is jointly controlled by λ_0, λ_1 and λ_3, which are sum-

Table C.1: Hyperparameter for Shrinkage

Hyperparameter	Meaning	Example	Setting
λ_0	Over all confidence in the prior	$\lambda_0 \downarrow \Longrightarrow$ weights to prior $\uparrow$	0.5
λ_1	Confidence in higher-lag coefficients equal to 0	$\lambda_1 = 0 \Longrightarrow$ equal weights for all lags	1
λ_3	Tightness of the constant term	$--$	100

marised in Table C.1.

C.2 Sampling from the posterior distribution

Since the Minnesota prior is natural conjugate, this appendix will introduce the posterior sampling, denoted as $P(A, B, \Sigma_u|Y_T) = P(A|Y_T)P(\Sigma_u|A, Y_T)P(B|A, \Sigma_u, Y_T)$, which starts with the prior $P(A, B, \Sigma_u) = P(A)P(\Sigma_u|A)P(B|A, \Sigma_u)$ then is revised by data $Y_T = (y_1^{'}, \ \ldots, \ y_T^{'})^{'}$.

First, recalling $\sigma_{ii}^{-1}|A \sim \Gamma(\kappa_i, \tau_i(A))$ – the ii element of prior $P(\Sigma_u|A)$, then the corresponding element in posterior $P(\Sigma_u|A, Y_T)$ is:

$$\sigma_{ii}^{-1}|A, Y_T \sim \Gamma(\kappa_i^*, \tau_i^*(A)),$$

where

$$\kappa_i^* = \kappa_i + T/2;$$

$$\tau_i^*(A) = \tau_i(A) + \zeta_i^*(A);$$

and $\zeta_i^*(A)$ denotes the sum of squared residuals of regression of $\tilde{Y}_i(A)$ on $\tilde{X}_i$, detailed as below.

$$\zeta_i^*(A) = (\tilde{Y}_i^{'}(A)\tilde{Y}_i(A)) - (\tilde{Y}_i^{'}(A)\tilde{X}_i)(\tilde{X}_i^{'}\tilde{X}_i)(\tilde{X}_i^{'}\tilde{Y}_i(A)),$$

$$\underset{[(T+n*p+1)\times 1]}{\tilde{Y}_i(A)} = \left[a_i^{'}Y_1 \ \ldots \ a_i^{'}Y_T \ m_i(A)^{'}P_i\right]^{'},$$

$$\underset{[(T+n*p+1)\times(n*p+1)]}{\tilde{X}_i} = \left[X_0 \ \ldots \ X_{T-1} \ P_i\right],$$

where P_i is the Cholesky factor of $M_i^{'} = P_i P_i^{'}$.

Second, as the i^{th} row of prior $P(B|A, \Sigma_u)$ is $b_i|A, \Sigma_u \sim N(m_i(A), \sigma_{ii} M_i)$, then the i^{th} row of posterior $P(B|A, \Sigma_u, Y_T)$ becomes:

$$b_i|A, \Sigma_u, Y_T \sim N(m_i^*(A), \sigma_{ii} M_i^*),$$

where

$$m_i^*(A) = (\tilde{X}_i^{'} \tilde{X}_i)^{-1} (\tilde{X}_i^{'} \tilde{Y}_i(A)),$$

and

$$M_i^* = (\tilde{X}_i^{'} \tilde{X}_i)^{-1}.$$

Third, the posterior of A is derived by Baumeister & Hamilton (2015) as:

$$P(A|Y_T) = \frac{k_t P(A)[det(A\hat{\Omega}_t A^{'})]^{T/2}}{\Pi_{i=1}^{n}[(2/T)\tau_i^*(A)]^{\kappa^*}} \prod_{i=1}^{n} \tau_i(A)^{\kappa_i},$$

where $P(A)$ is the prior for A; $\hat{\Omega}_t$ is the sample variance matrix for the reduced form VAR residuals,

$$\hat{\Omega}_t = T^{-1}\{\sum_{t=1}^{T} y_t y_t^{'} - (\sum_{t=1}^{T} y_t x_{t-1}^{'})(\sum_{t=1}^{T} x_{t-1} x_{t-1}^{'})^{-1}(\sum_{t=1}^{T} x_{t-1} x_{t-1}^{'})\};$$

and, k_t is the function of data and prior parameters, but not related to A, Σ_u, and B, which is used to integrate the probability to unity over the set of allowable values for A. Hence, there is no necessity to calculate k_t for forming the posterior.

C.3 Instruction for A^{-1} based on the identification restrictions in Kilian & Murphy (2014)

This appendix guides the application of the identifications in Kilian & Murphy (2014) on A^{-1} as constraints.

Recall the impulse response function defined in (4.7): $H_s = \Psi_s A^{-1} chol(\Sigma_u)'$.

Since $chol(\Sigma_u)'$ is positive definite and Ψ_s takes the form of the identity matrix when $s = 0$, the sign restrictions proposed in Table 4.1 are defined by the non-diagonal elements in A^{-1}. I utilise the truncated Student t distributions prior to present the sign of response restrictions, as per Baumeister & Hamilton (2015, 2018, 2019). Meanwhile, the prior of diagonal elements, A_{ii}^{-1} where i=1, ..., 4 in this application, is presented through symmetric Student $t(\mu_2, \sigma_2, \nu_2)$ with $\mu_2 = 0.8$, $\sigma_2 = 0.2$, and $\nu_2 = 3$, implying a 98.6% prior probability that $A_{ii}^{-1} > 0$. The prior reflects a modest probability that the endogenous variables' response to the shocks they have caused is negative.

Table C.2 summarises the prior distributions for all the identified parameters, where the (asymmetric or truncated) Student t distributions are utilised. The location parameter refers to the prior distribution mode, while the scale and degree of freedom parameters jointly indicate the uncertainty of the prior. In this exercise, all the degrees of freedoms are kept to 3 to ensure a fatter tail relative to the normal distribution. Additionally, the majority of the scales are set at 0.2 in order to address the relative uncertainty. Since Bayesian models are sensitive to the tightness of the prior in the macro-econometric literature, as per Bańbura et al. (2010) and Chan, Jacobi, & Zhu (2018), I leave one of the scale parameters with the highest uncertainty in the parameter for the arg-minimisation process, $\boldsymbol{\phi}$.

As shown in Table C.2, A_{21}^{-1} is the economic activity response to a "negative flow supply shock", namely an unexpected boom of oil production such as "flow supply shock", defined in the oil literature as unexpected oil supply disruptions. Hence, the response is positive defined and presented as Student $t(0.1, 0.2, 3)$, truncated to be positive, where the mode, scale, and degree of freedom are at 0.1, 0.2, and 3, respectively. With the same logic, the prior distribution of the economic activity response to speculative demand shock is Student $t(-0.1, 0.2, 3)$, truncated to be negative.

Moreover, Equation (4.19), the demand elasticity for oil in use, $\eta^{O,Use}$, is non-linear and conditional on A_{31}^{-1} and A_{41}^{-1}, for the responses of the oil price and inventories to a negative flow supply shock. I represent them by Student $t(-1.5, 0.2, 3)$, truncated to be negative and symmetric Student $t(-1, 0.2, 3)$, respectively. Under the prior distribution, the probability of $-0.8 \leqslant \eta^{O,Use} \leqslant 0$ is 87.8% based on 50,000

random prior samples of A_{31}^{-1} and A_{41}^{-1}.[1]

Table C.2: Prior distributions for model parameters

Parameter	Meaning	Location	Scale	DoF	Skew	Truncate Restrictions
	Prior affecting contemporaneous coefficients A^{-1}					
		Student t **distribution**				
A_{21}^{-1}	Economic activity response to negative flow supply shock	0.1	0.2	3	–	$0 < A_{21}^{-1}$
A_{31}^{-1}	Oil price response to negative flow supply shock	-1.5	0.2	3	–	$A_{31}^{-1} < 0$
A_{41}^{-1}	Oil inventory response to negative flow supply shock	-1	0.2	3	–	none
A_{12}^{-1}	Oil supply response to flow demand shock	0.01	0.2	3	–	$0 < A_{12}^{-1}$
A_{32}^{-1}	Oil price response to flow demand shock	0.8	0.2	3	–	$0 < A_{32}^{-1}$
A_{42}^{-1}	Oil inventory response to flow demand shock	0	0.2	3	–	none
A_{13}^{-1}	Oil supply response to speculative demand shock	0.01	0.2	3	–	$0 < A_{13}^{-1}$
A_{23}^{-1}	Economic activity response to speculative demand shock	-0.1	0.2	3	–	$A_{23}^{-1} < 0$
A_{43}^{-1}	Oil inventory response to speculative demand shock	2	$\sigma^{A_{43}^{-1}}$	3	–	$0 < A_{43}^{-1}$
A_{14}^{-1}	Oil supply response to inventory shock	0	0.2	3	–	none
A_{24}^{-1}	Economic activity response to inventory shock	0	0.2	3	–	none
A_{34}^{-1}	Oil price response to inventory shock	0	0.2	3	–	none
A_{ii}^{-1}	Effect of endogenous variables' shocks on themself (diagonal elements of A^{-1})	0.8	0.2	3	–	none
$\underline{\eta}^{O,Use}$	Lower bound of oil demand elasticity for use	-0.1	0.2	3	–	$-0.8 \leq \underline{\eta}^{O,Use} \leq -0.1$
		Asymmetric t **distribution**				
h_1	Determinant of A	0.6	1.6	3	2	none

Note: For the (Asymmetric) Student t distribution, the location parameter refers to the mode; scale and degree of freedom (DoF) parameters that jointly determine the uncertainty of the prior. This chapter sets all the DoF to 3 for a fat tail prior relative to a normal distribution.

Short-run oil supply elasticities, restricted as $0 < \eta^{Supply} < 0.0258$, are cal-

[1] From a simulation practice, the smaller A_{31}^{-1}'s location (e.g. -5), the higher the probability of $-0.8 \leqslant \eta^{O,Use} \leqslant 0$. However, a large negative value of A_{31}^{-1}'s location will lead the model to violate a dynamic sign restriction — the responses of the global real activity to an unanticipated flow supply disruption are negative for the first 12 months, which is $H_{s,21} < 0$, $\forall s \leqslant 12$.

culated through the oil supply response to the flow demand shock divided by the oil price response to the same shock, $\frac{A_{12}^{-1}}{A_{32}^{-1}}$, as well as the oil supply response to speculative demand shock divided by the oil price response to the same shock, $\frac{A_{13}^{-1}}{A_{33}^{-1}}$. The median of A_{33}^{-1} is approximately 0.8, as its prior is symmetric Student $t(0.8, 0.2, 3)$. Therefore, to locate the prior model in the restricted range, the prior distributions for A_{12}^{-1}, A_{32}^{-1} and A_{13}^{-1} are Student $t(0.01, 0.2, 3)$, Student $t(0.8, 0.2, 3)$ and Student $t(0.01, 0.2, 3)$, which are positively truncated, respectively.

The oil inventory response to a flow demand shock (A_{42}^{-1}) and the responses of the oil price, the short-run production and the economic activity to an inventory shock (A_{14}^{-1}, A_{24}^{-1}, and A_{34}^{-1}) are not restricted by sign, but should not be restricted to exactly zero. Therefore, their prior distribution is symmetric Student $t(0, 0.2, 3)$. The contemporary economic activity response to a speculative demand shock is restricted to be negative, and hence is presented as Student $t(-0.1, 0.2, 3)$, truncated to be negative.

C.4 The economic findings in Specification I

In this appendix I illustrate the economic findings in Kilian & Murphy (2014) through the historical decompositions.

Figure C.1 plots the historical decomposition of the real oil price expressed in log-level (times 100) for the 1978:06–2009:08 period. The solid black lines are the Bayesian posterior median from C-BSVARs, the shaded regions are the corresponding 68% posterior confidence sets, the dotted red lines are the real price of oil, while the dotted blue lines are the benchmark estimates in Kilian & Murphy (2014, p. 466). It is clear that the 68% confidence sets from C-BSVARs cover the preferred measure of the cumulative effects by Kilian & Murphy (2014) over the periods of interest. During the 2003–2008 period, the estimates of C-BSVARs confirm the observation in Kilian & Murphy (2014) that the effects of the flow demand shock drove the real crude oil price. There was no significant up-ward trend for the flow supply and speculative shocks' effects, which fluctuated around their mean during the period.

The C-SBVARs, following Kilian & Murphy (2014), illustrated five episodes

where the speculative demand shock played an important role, namely the oil price shocks in 1979 and 1980, the collapse of OPEC in 1986, the inventory conundrum of 1990, the Venezuelan crisis and the Iraq War of 2002/03. Echoing Kilian & Murphy (2014), Figure C.2 illustrates the cumulative effects of the flow supply (in grey) and speculative demand (in blue) shocks on the real oil price (upper panel) and on the change of oil inventories (lower panel) during the five periods, respectively.

In 1979 and 1980, Kilian & Murphy (2014) argue that the oil price increase was caused by the cumulation of positive speculative demand, but not by the flow supply disruptions associated with the Iranian Revolution of late 1978 and early 1979. The first column of Figure C.2 shows that not only was there a dramatic and persistent increase in the real oil price driven by the positive flow demand shocks in 1979 and 1980, but also that the increase was reinforced after 1979:04 by a cumulation in speculative demand. This differs from the persistent price increase after 2003. Persistently positive effects of the speculative shock were also reflected in the change of oil inventories. However, there was little evidence of flow supply shocks being responsible for the oil price surge of 1979, consistent with the fact that overall the global oil production increased in 1979, reflecting additional oil productions outside of Iran. Only in late 1980 and early 1981 was there a moderate spike in the real price of oil driven by flow supply shocks, in part associated with the outbreak of the Iran–Iraq War presented in the second column of Figure C.2.

Prior to the collapse of OPEC (in 1985:12), the real oil price was driven by the cumulative effect of speculative demand shocks. Following the collapse, shown in the third column of Figure C.2, a drop in the real oil price was reinforced by a decline in speculative demand from 1986:01. Additionally, unexpected increases in oil supply before 1986:03 reduced the real price of oil, while the oil inventories rose in response. These observations are consistent with Kilian & Murphy (2014). Moreover, the last column of Figure C.2 focuses on the flow supply shock of 2002/03 when, within the space of several months Venezuelan oil productions slowed considerably (at the end of 2002) and then Iraqi oil productions ceased altogether in early 2003. These events reflect a combination of negative flow supply and positive speculative demand shocks.

Furthermore, an unexpected flow supply disruption and increase of speculative demand jointly caused the oil price spike during the Persian Gulf War. Given one of the largest unexpected oil supply disruptions in history in August of 1990, the

change of inventory was modest. Kilian & Murphy (2014) found there was a definite negative effect of the oil-supply drop in 1990:08 on the change in inventories, and the influence of this oil-supply drop on inventories was absorbed by a positive effect caused by the increase of speculative demand. C-BSVARs also found evidence for the joint influence of speculative and flow supply shock in 1990:08, illustrated in the fourth column of Figure C.2. When the threat of Saudi oil fields being captured by Iraq had been removed by the presence of the U.S. troops in 1990:11, a drop in the speculative demand drove the real price of oil down, while the flow demand did not show any change from its average level during this episode.

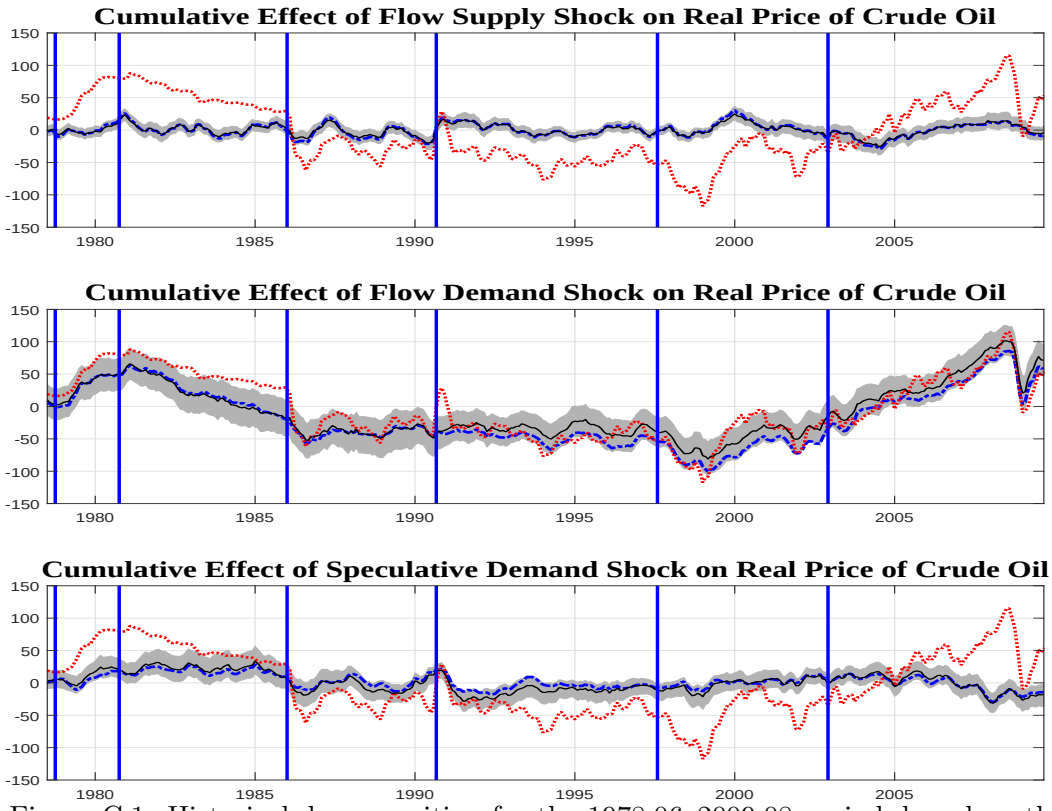

Figure C.1: Historical decomposition for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$

Note: *Solid black lines*: Bayesian posterior median; *shaded region*: 68% posterior credible sets; *dotted red lines*: the real price of oil expressed in log-level (times 100); *dotted blue lines*: benchmark estimates in Kilian & Murphy (2014, p. 466); *vertical bars* indicate the major exogenous events in the oil markets, notably the outbreak of the Iranian Revolution and the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, the Asian Financial Crisis, and the Venezuelan crisis, followed by the Iraq War. Following Kilian & Murphy (2014), I discard the first five years of data in an effort to remove the transition dynamics in constructing the historical decomposition.

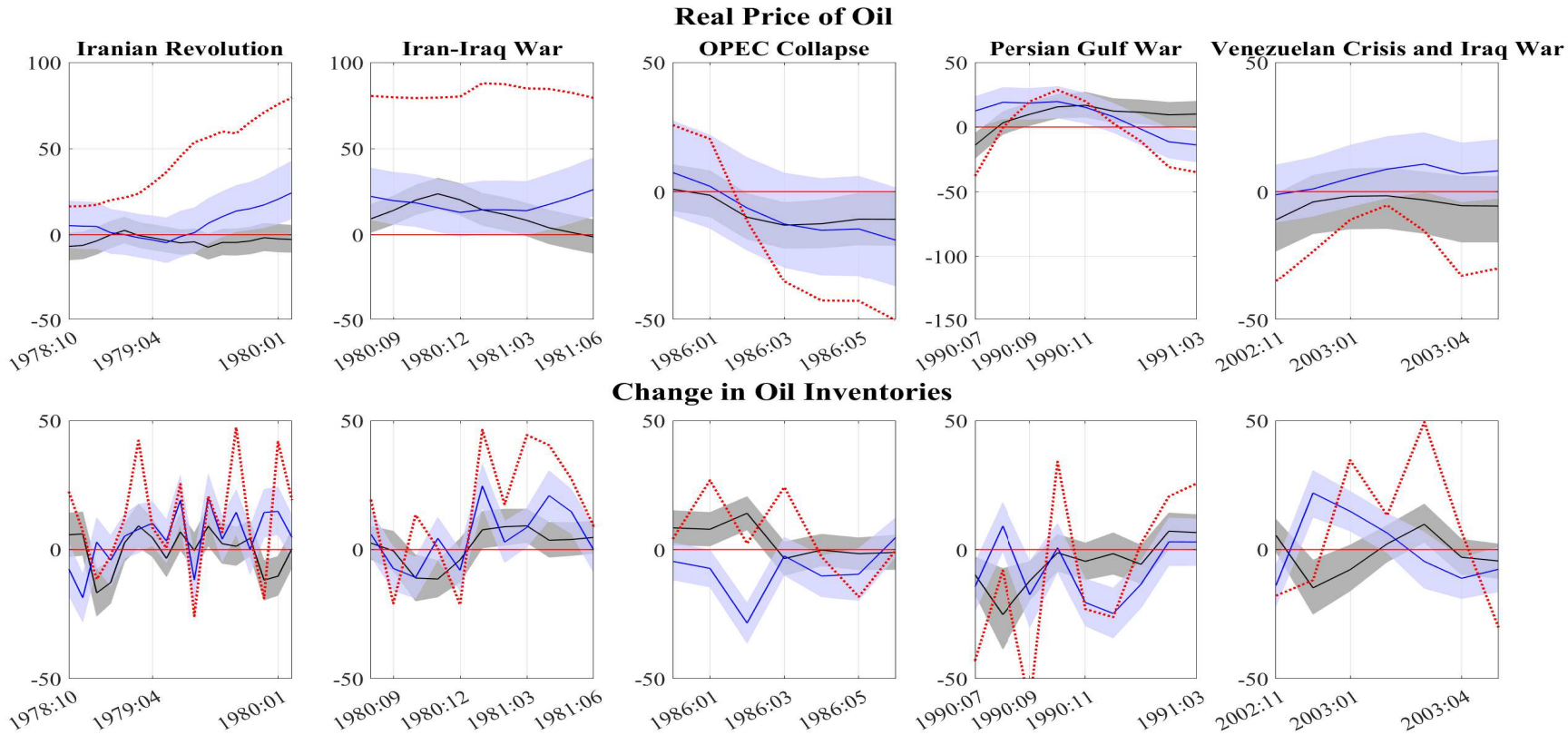

Figure C.2: Historical decompositions for specific events (in columns) of the real price of oil (the upper panels) and the change in oil inventories (the lower panels)

Note: The events are the Iranian Revolution and the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, the Asian Financial Crisis, and the Venezuelan crisis, followed by the Iraq War, respectively. *Red dashed lines*: the observed oil price (*upper panel*) and change in oil inventories (*lower panel*), respectively; *solid black lines*: the Bayesian posterior median of the cumulative effect of flow supply shocks; and the *grey shaded region*: the corresponding 68% posterior credible sets; *solid blue lines*: Bayesian posterior median of the cumulative effect of speculative demand shocks; *blue shaded region*: the corresponding 68% posterior credible sets.

C.5 Additional results for Specification II

Prior and posterior distributions for A^{-1}

Figure C.3 compares the prior distribution (solid red lines) with posterior distributions shown as histograms.

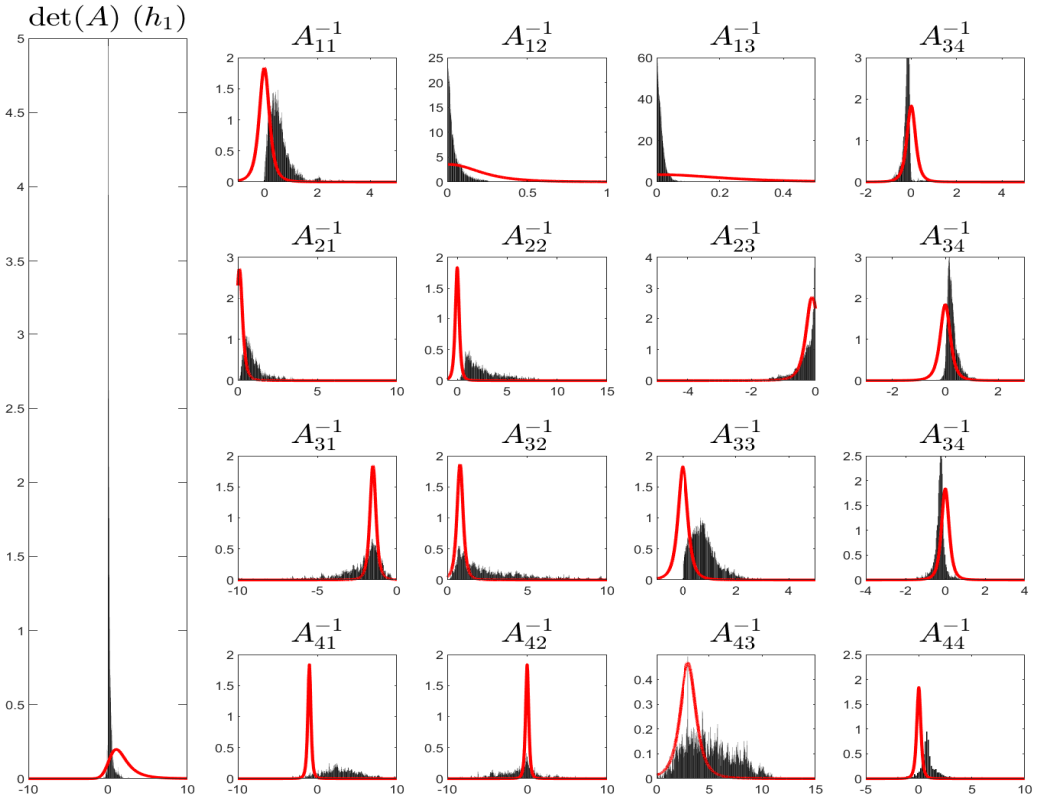

Figure C.3: Prior (*red lines*) and posterior (*black histograms*) distributions for the unknown elements in A^{-1} in the Bayesian implementation of the 4-variable baseline model ($\eta^{Supply} \leqslant 0.0258$)

Impose response functions and historical decompositions

The structural impulse responses to the oil supply and demand shocks are illustrated in Figure C.4, produced by C-BSVARs with the lower-bound uncertainty of oil demand elasticity for use. Additionally, I draw the 95% posterior credible sets as dotted black lines.

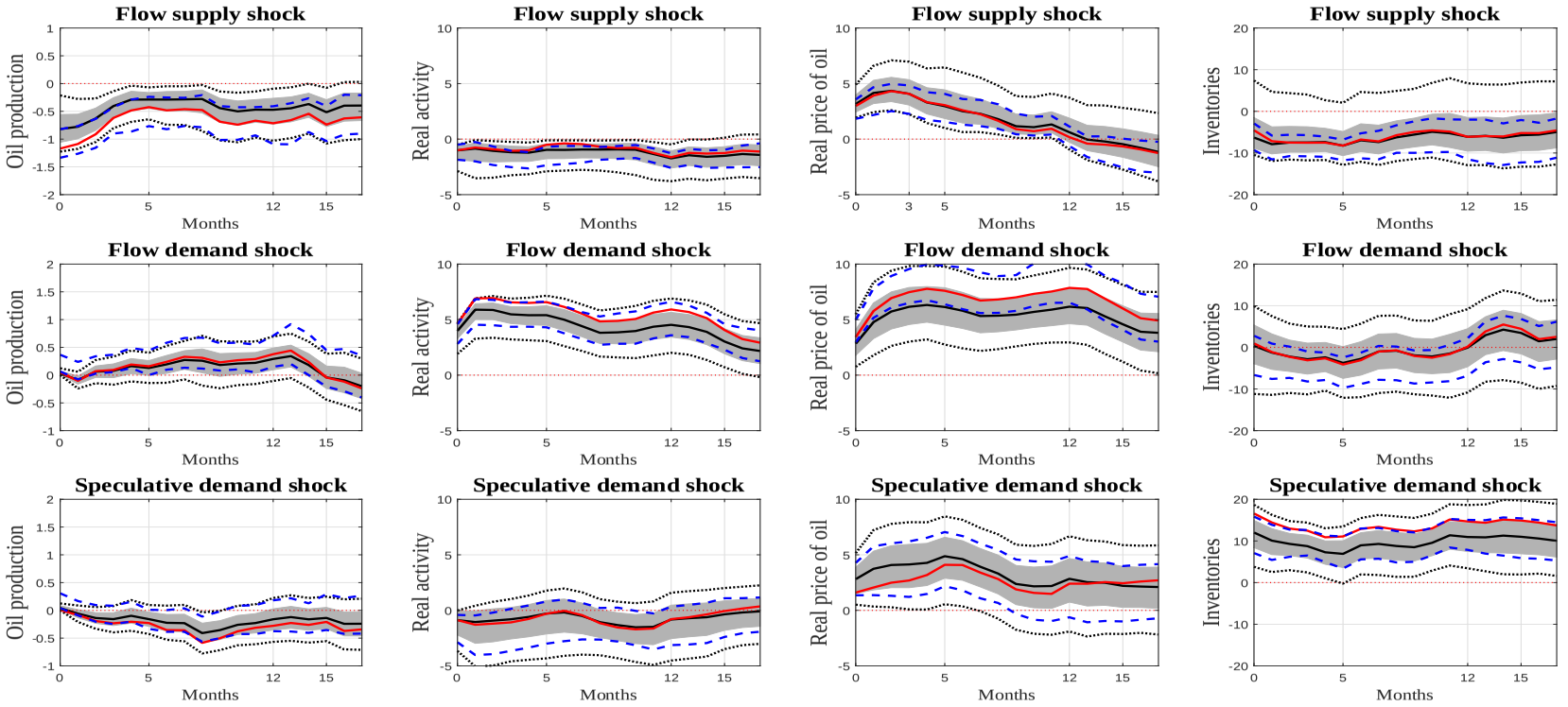

Figure C.4: Structural impulse responses under the uncertainty of the oil demand elasticity in use for the 1973:02–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$

Note: *Solid black lines*: Bayesian posterior median; *shaded region*: 68% posterior credible sets; *dotted black lines*: 95% posterior credible sets; *solid red lines*: the impulse response estimates for the model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity amongst the admissible structural models obtained, conditional on the least-squares estimate of the reduced-form VAR model replicated with Kilian & Murphy's (2014) code available on the journal's website; *dotted blue lines*: the corresponding point-wise 68% posterior error bands replicated with the aforementioned code.

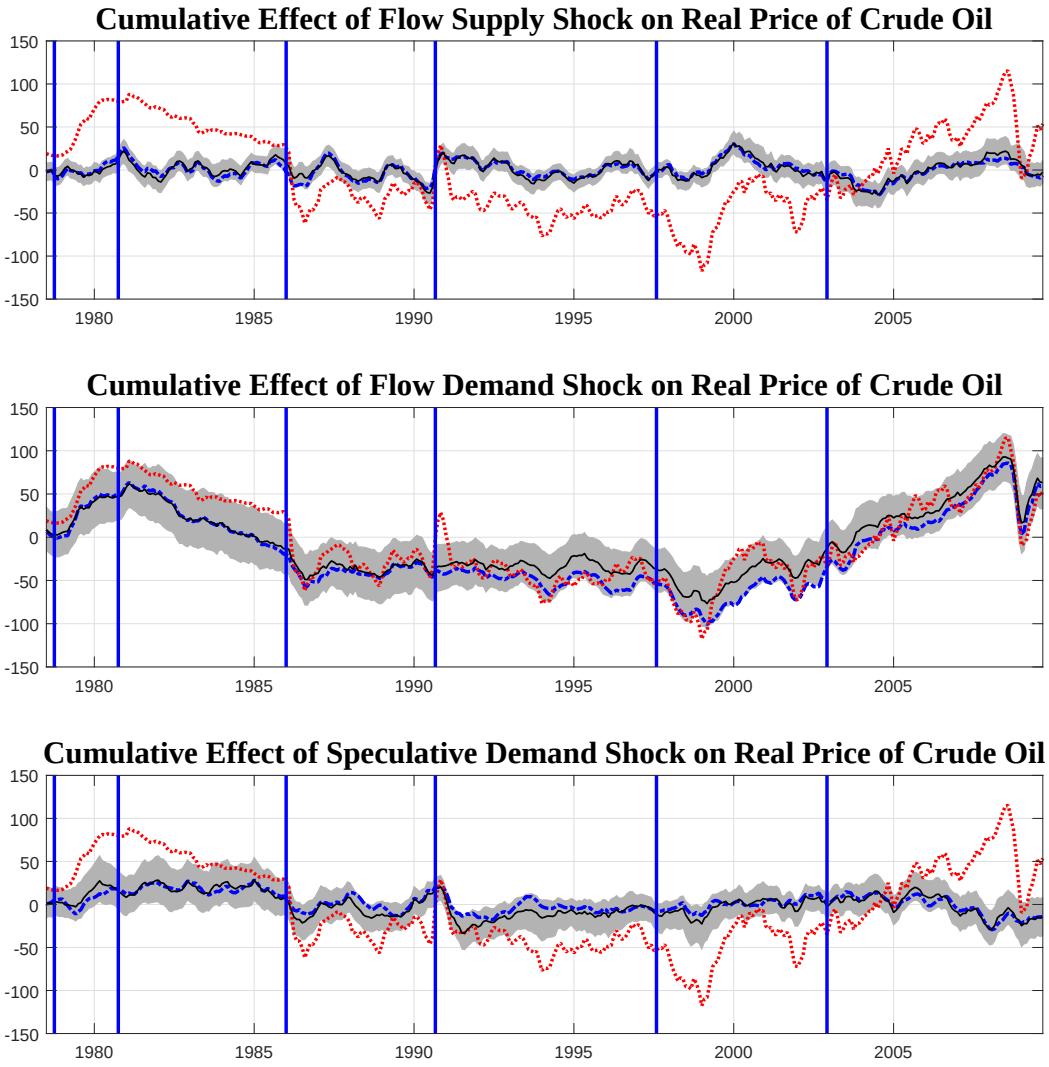

Figure C.5: Historical decomposition under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$

Note: *Solid black lines*: Bayesian posterior median; *shaded region*: 68% posterior credible sets; *dotted red lines*: the real price of oil expressed in log-level (times 100); *dotted blue line*: benchmark estimates in Kilian & Murphy (2014, p. 466); *vertical bars* indicate major exogenous events in oil markets, notably the outbreak of the Iranian Revolution and the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, the Asian Financial Crisis, and the Venezuelan crisis, followed by the Iraq War. Following Kilian & Murphy (2014), I discard the first five years of data in an effort to remove the transition dynamics in constructing the historical decomposition.

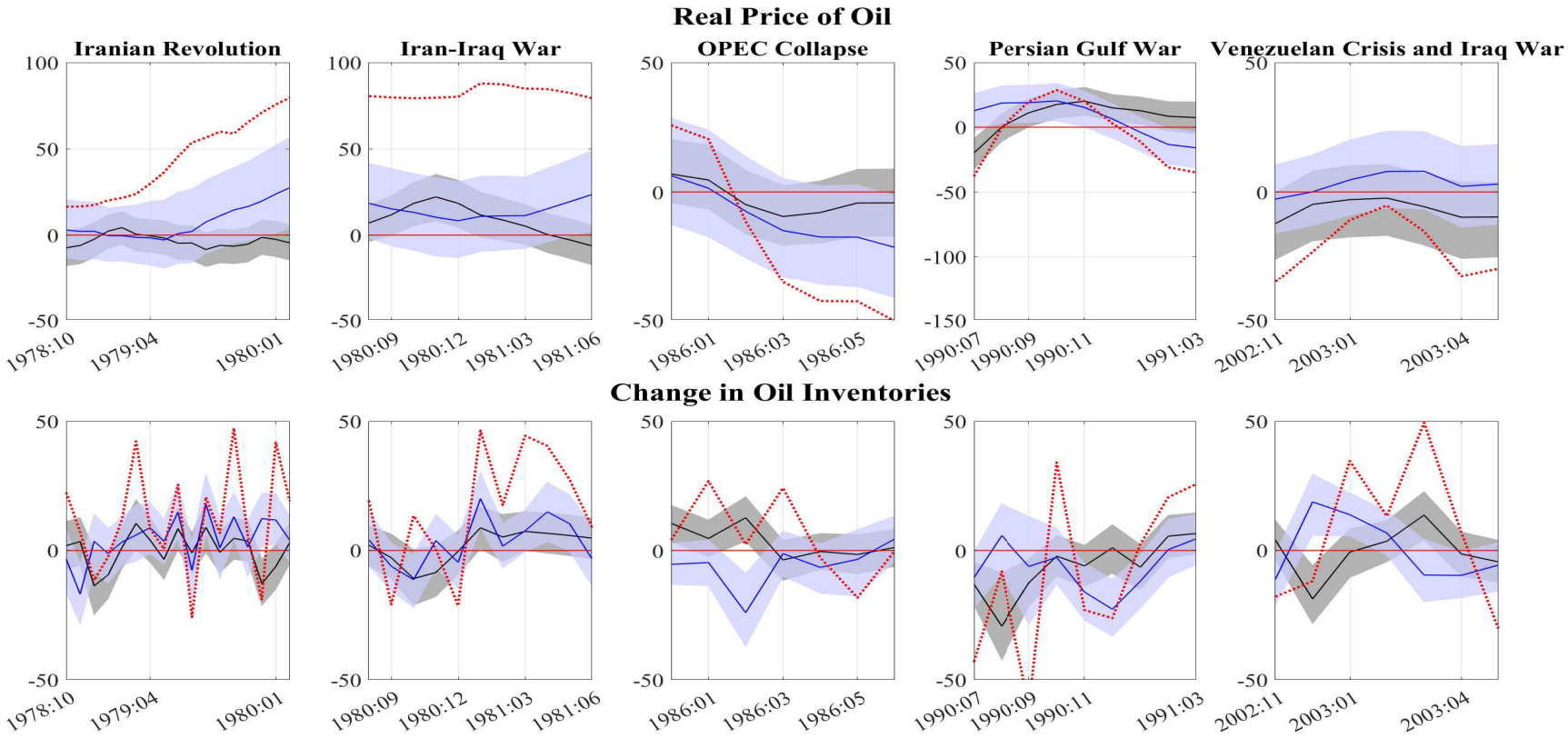

Figure C.6: Historical decompositions for specific events (in columns) of the real price of oil (the upper panels) and the change in oil inventories (the lower panels) under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, based on the model with $\eta^{Supply} \leq 0.0258$

Note: The events are the Iranian Revolution and the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, the Asian Financial Crisis, and the Venezuelan crisis, followed by the Iraq War, respectively. *Red dashed lines*: the observed oil price (*upper panel*) and change in oil inventories (*lower panel*), respectively; *solid black lines*: the Bayesian posterior median of the cumulative effect of flow supply shocks; and the *grey shaded region*: the corresponding 68% posterior credible sets; *solid blue lines*: the Bayesian posterior median of the cumulative effect of speculative demand shocks; and the *blue shaded region*: the corresponding 68% posterior credible sets.

C.6 $\eta^{Supply} \leq 0.0258$ VS. $\eta^{Supply} \leq 0.5$

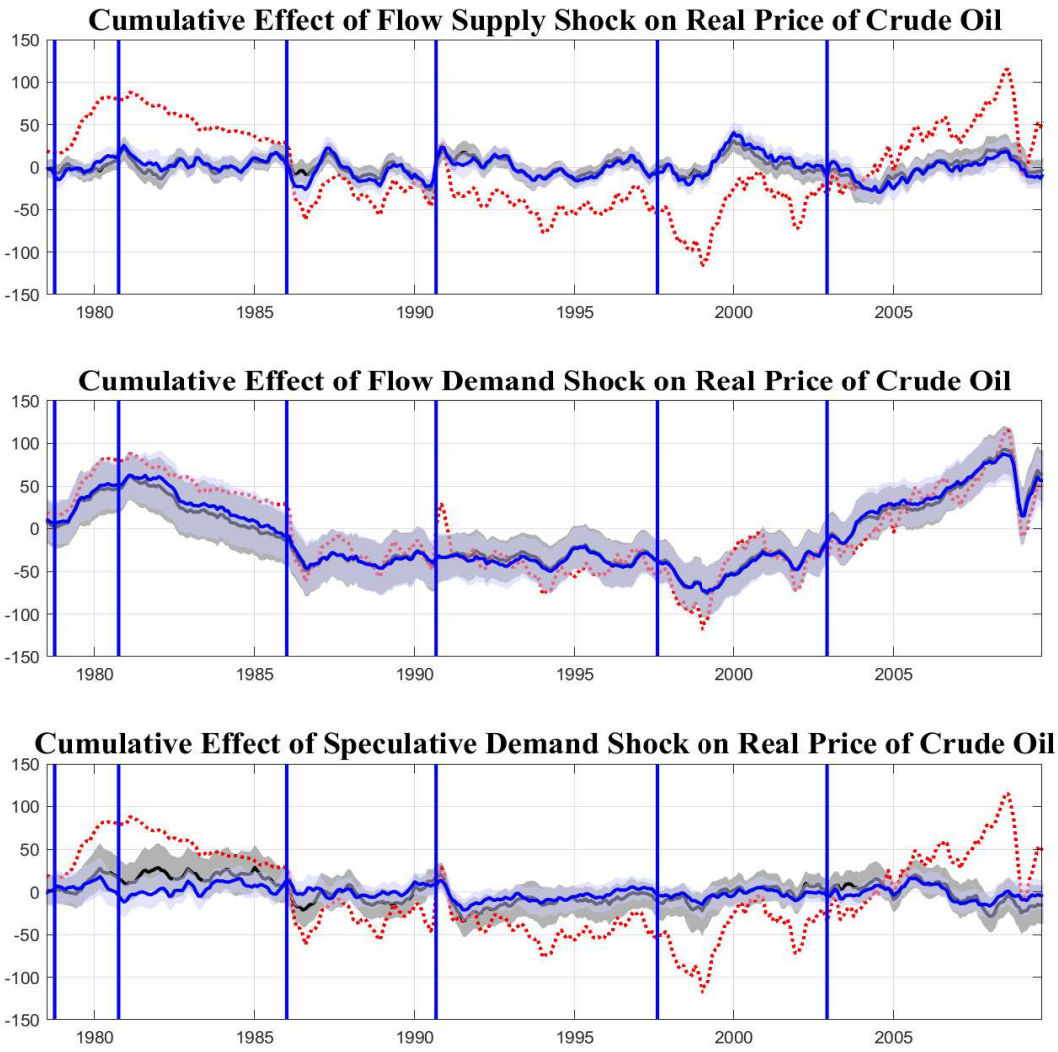

Figure C.7: Historical decomposition under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, comparing the models with $\eta^{Supply} \leq 0.0258$ and $\eta^{Supply} \leq 0.5$

Note: The model with $\eta^{Supply} \leq 0.0258$ is shown as grey, while the model with $\eta^{Supply} \leq 0.5$ is blue. *Solid lines*: the Bayesian posterior median; *shaded region*: 68% posterior credible sets; *dotted red lines*: the real price of oil expressed in log-level (times 100); *vertical bars* indicate the major exogenous events in the oil markets, notably the outbreak of the Iranian Revolution and the Iran–Iraq War, the collapse of OPEC, the outbreak of the Persian Gulf War, the Asian Financial Crisis, and the Venezuelan crisis, followed by the Iraq War. Following Kilian & Murphy (2014), I discard the first five years of data in an effort to remove the transition dynamics in constructing the historical decomposition.

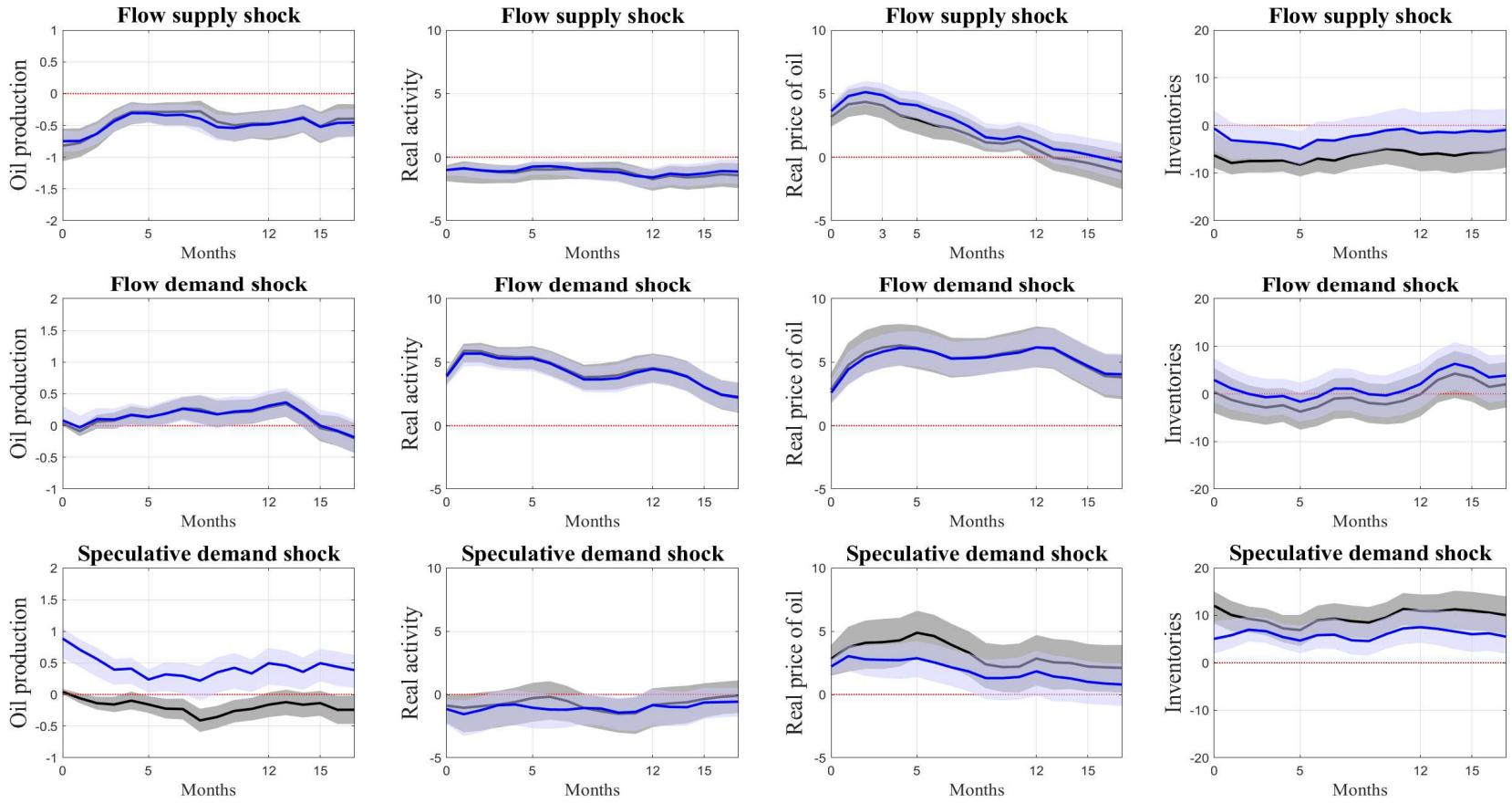

Figure C.8: Structural impulse responses under the uncertainty of the oil demand elasticity in use for the 1978:06–2009:08 period, comparing the models with $\eta^{Supply} \leq 0.0258$ and $\eta^{Supply} \leq 0.5$

Note: *Solid lines*: the Bayesian posterior median; *shaded region*: 68% posterior credible sets. *Colour grey* is for $\eta^{Supply} \leq 0.0258$, while *colour blue* is for $\eta^{Supply} \leq 0.5$.

References

Alberola, E., Chevallier, J., & Chèze, B. (2008). Price drivers and structural breaks in european carbon prices 2005–2007. *Energy Policy*, *36*(2), 787–797.

Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, *25*(4), 539–573.

Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. In *Handbook of economic forecasting* (Vol. 2, pp. 427–507). Elsevier.

Amir-Ahmadi, P., & Drautzburg, T. (2019). Identification and inference with ranking restrictions. Available at: https://sites.google.com/site/tdrautzburg/research.

Amisano, G., & Giannini, C. (2012). *Topics in structural var econometrics*. Springer Science & Business Media.

Antolín-Díaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for svars. *American Economic Review*, *108*(10), 2802–29.

Arias, J. E., Caldara, D., & Rubio-Ramirez, J. F. (2019). The systematic component of monetary policy in svars: An agnostic identification procedure. *Journal of Monetary Economics*, *101*, 1–13.

Arias, J. E., Rubio-Ramírez, J. F., & Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica*, *86*(2), 685–720.

Bakshi, G., Gao, X., & Rossi, A. G. (2017). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, *65*(2), 619–641.

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

Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, *83*(5), 1963–1999.

Baumeister, C., & Hamilton, J. D. (2017). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *NBER Working Paper*(w24167). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095128.

Baumeister, C., & Hamilton, J. D. (2018). Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations. *Journal of Monetary Economics*, *100*, 48–65.

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, *109*(5), 1873–1910.

Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, *30*(2), 326–336.

Baumeister, C., & Kilian, L. (2014). Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review*, *62*(1), 119–145.

Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: a forecast combination approach. *Journal of Business & Economic Statistics*, *33*(3), 338–351.

Baumeister, C., & Kilian, L. (2016a). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *The Journal of Economic Perspectives*, *30*(1), 139–160.

Baumeister, C., & Kilian, L. (2016b). Lower oil prices and the us economy: Is this time different? *Brookings Papers on Economic Activity*, *2016*(2), 287–357.

Baumeister, C., & Kilian, L. (2016c). Understanding the decline in the price of oil since june 2014. *Journal of the Association of Environmental and Resource Economists*, *3*(1), 131–158.

Baumeister, C., Kilian, L., & Zhou, X. (2013). Are product spreads useful for forecasting? An empirical evaluation of the verleger hypothesis. *Mimeo: University of Michigan*, 338–341.

Baumeister, C., Kilian, L., & Zhou, X. (2018). Are product spreads useful for forecasting oil prices? An empirical evaluation of the verleger hypothesis. *Macroeconomic Dynamics*, *22*(3), 562–580.

Bauwens, L., Korobilis, D., Koop, G., & Rombouts, J. V. (2011). A comparison of forecasting procedures for macroeconomic series: the contribution of structural break models.

Belmonte, M. A., Koop, G., & Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, *33*(1), 80–94.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, *7*(4), 434–455.

Bruns, M., & Piffer, M. (2019). Bayesian structural var models: a new approach for prior beliefs on impulse responses. Available at SSRN: https://ssrn.com/abstract=3366913.

Caldara, D., Cavallo, M., & Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, *103*, 1–20.

Chan, J., & Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of applied econometrics*, *33*(4), 509–532.

Chan, J., Eisenstat, E., Hou, C., & Koop, G. (2018). Composite likelihood methods for large Bayesian vars with stochastic volatility. Available at CAMA: https://cama.crawford.anu.edu.au/sites/default/files/publication/cama_crawford_anu_edu_au/2018-05/26_2018_chan_eisenstat_hou_koop.pdf.

Chan, J., Jacobi, L., & Zhu, D. (2018). How sensitive are var forecasts to prior hyperparameters? An automated sensitivity analysis. Available at: http://dx.doi.org/10.2139/ssrn.3185915.

Chen, Z., & Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics*, *59*(4), 762–769.

Cogley, T., & Sbordone, A. M. (2008). Trend inflation, indexation, and inflation persistence in the new keynesian phillips curve. *American Economic Review*, *98*(5), 2101–26.

Coglianese, J., Davis, L. W., Kilian, L., & Stock, J. H. (2017). Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics*, *32*(1), 1–15.

Cooper, J. C. (2003). Price elasticity of demand for crude oil: estimates for 23 countries. *OPEC review*, *27*(1), 1–8.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, *91*(434), 883–904.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101.

Dahl, C. (1993). A survey of oil demand elasticities for developing countries. *OPEC review*, *17*(4), 399–410.

Dahl, C., & Sterner, T. (1991). Analysing gasoline demand elasticities: a survey. *Energy economics*, *13*(3), 203–210.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, 278–292.

Diebold, F. X. (1991). A note on bayesian forecast combination procedures. In *Economic structural change* (pp. 225–232). Springer.

Diebold, F. X., Hahn, J., & Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics*, *81*(4), 661–673.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3).

Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, *6*(1), 21–40.

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

Duffie, D., Gray, S., & Hoang, P. (1999). Volatility in energy prices. In *Managing energy price risk. 2. ed.*

ECB. (2014). Understanding the recent decline in oil prices. *Monthly Bulletin — ECB*, *November*, 14–15.

EIA, U. (2001). Monthly energy review. *Energy Information Administration, US*.

Eisenstat, E., Chan, J. C., & Strachan, R. W. (2016). Stochastic model specification search for time-varying parameter vars. *Econometric Reviews*, *35*(8-10), 1638–1665.

Espey, M. (1998). Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics*, *20*(3), 273–295.

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

Galbraith, J. W., & Norden, S. v. (2012). Assessing gross domestic product and inflation probability forecasts derived from bank of england fan charts. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(3), 713–727.

Garratt, A., Henckel, T., & Vahey, S. P. (2019). Empirically-transformed linear opinion pools. Available at: https://shaunvahey.files.wordpress.com/2019/06/ghv-24june2019.pdf.

Garratt, A., Lee, K., & Shields, K. (2016). Forecasting global recessions in a gvar model of actual and expected output. *International Journal of Forecasting*, *32*(2), 374–390.

Garratt, A., & Mise, E. (2014). Forecasting exchange rates using panel model and model averaging. *Economic Modelling*, *37*, 32–40.

Garratt, A., Mitchell, J., & Vahey, S. P. (2014). Measuring output gap nowcast uncertainty. *International Journal of Forecasting*, *30*(2), 268–279.

Garratt, A., Mitchell, J., Vahey, S. P., & Wakerly, E. C. (2011). Real-time inflation forecast densities from ensemble phillips curves. *The North American Journal of Economics and Finance*, *22*(1), 77–87.

Garratt, A., Vahey, S. P., & Zhang, Y. (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, *34*(3), 456–462.

Gel, Y., Raftery, A. E., & Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method. *Journal of the American Statistical Association*, *99*(467), 575–583.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–511.

Gelman, M., Gorodnichenko, Y., Kariv, S., Koustas, D., Shapiro, M. D., Silverman, D., & Tadelis, S. (2016). The response of consumer spending to changes in gasoline prices. (No. w22969). Available at NBER: https://www.nber.org/papers/w22969.pdf.

Geweke, J. (1992). Evaluating the accurating of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, *4*, 169–193.

Geweke, J. (2010). *Complete and incomplete econometric models*. Princeton University Press.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422.

Gneiting, T., Ranjan, R., et al. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, *7*, 1747–1782.

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. *Review of Finance*, *17*(1), 35–105.

Grimit, E. P., Gneiting, T., Berrocal, V., & Johnson, N. A. (2006). The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quarterly Journal of the Royal Meteorological Society*, *132*(621C), 2925–2942.

Groen, J. J., Paap, R., & Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, *31*(1), 29–44.

Hamilton, G., & Herrera, A. (2004). Oil shocks and aggregate macroeconomic behavior: The role of monetary policy. *Journal of Money, Credit, and Banking*, *36(2)*, 265–286.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton University Press, Princeton, NJ.

Hamilton, J. D. (2008). Understanding crude oil prices. (No. w14492). Available at NBER: https://www.nber.org/papers/w14492.pdf.

Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08. (No. w15002). Available at NBER: https://www.nber.org/papers/w15002.pdf.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(3), 205–228.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, *13*(2), 281–291.

Hausman, J. A., & Newey, W. K. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica: Journal of the Econometric Society*, 1445–1476.

Herrera, A. M., & Rangaraju, S. K. (2018). The effect of oil supply shocks on us economic activity: What have we learned? Available at SSRN: http://gattonweb.uky.edu/faculty/herrera/documents/HRoil.pdf.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382–401.

Hong, H., & Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, *105*(3), 473–490.

Hughes, J. E., Knittel, C. R., & Sperling, D. (2006). Evidence of a shift in the short-run price elasticity of gasoline demand. (No. w12530). Available at NBER: https://www.nber.org/papers/w12530.pdf.

Inoue, A., & Kilian, L. (2013). Inference on impulse response functions in structural var models. *Journal of Econometrics*, *177*(1), 1–13.

Jebabli, I., Arouri, M., & Teulon, F. (2014). On the effects of world stock market and oil price shocks on food prices: An empirical investigation based on TVP-VAR models with stochastic volatility. *Energy Economics*, *45*, 66–98.

Johnstone, D., Jones, S., Jose, V., & Peat, M. (2013). Measures of the economic value of probabilities of bankruptcy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(3), 635–653.

Kang, S. H., Kang, S.-M., & Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, *31*(1), 119–125.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *The American Economic Review*, *99*(3), 1053–1069.

Kilian, L. (2017). How the tight oil boom has changed oil and gasoline markets. Available at CESifo: https://www.econstor.eu/bitstream/10419/155622/1/cesifo1_wp6380.pdf.

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Kilian, L., & Murphy, D. (2010). The role of inventories and speculative trading in the global market for crude oil. Available at: https://www.cftc.gov/sites/default/files/idc/groups/public/@swaps/documents/file/plstudy_28_cepr.pdf.

Kilian, L., & Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market var models. *Journal of the European Economic Association*, *10*(5), 1166–1188.

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, *29*(3), 454–478.

Kilian, L., & Park, C. (2009). The impact of oil price shocks on the us stock market. *International Economic Review*, *50*(4), 1267–1287.

Kilian, L., & Zhou, X. (2018). Structural interpretation of vector autoregressions with incomplete information: Revisiting the role of oil supply and demand shocks:

Comment. Available at CESifo: https://www.econstor.eu/bitstream/10419/185364/1/cesifo1_wp7166.pdf.

Kinney, S. K., & Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics*, *63*(3), 690–698.

Knetsch, T. A. (2007). Forecasting the price of crude oil via convenience yield predictions. *Journal of Forecasting*, *26*(7), 527–549.

Koop, G., & Korobilis, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, *177*(2), 185–198.

Koop, G., Korobilis, D., et al. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends® in Econometrics*, *3*(4), 267–358.

Koop, G., Poirier, D. J., & Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

Korobilis, D. (2013). VAR forecasting using bayesian variable selection. *Journal of Applied Econometrics*, *28*(2), 204–230.

Larsson, K., & Nossman, M. (2011). Jumps and stochastic volatility in oil prices: Time series evidence. *Energy Economics*, *33*(3), 504–514.

Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, 580–590.

Mitchell, J., & Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, *26*(6), 1023–1040.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of arima-garch models. *Energy Economics*, *32*(5), 1001–1008.

Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, *23*(3), 325–338.

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of applied Meteorology*, *12*(4), 595–600.

Noceti, P., Smith, J., & Hodges, S. (2003). An evaluation of tests of distributional forecasts. *Journal of Forecasting*, *22*(6-7), 447–455.

Panagiotelis, A., & Smith, M. (2008). Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting*, *24*(4), 710–727.

Pesaran, M. H., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, *104*(485), 325–337.

Pettenuzzo, D., & Ravazzolo, F. (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics*, *31*(7), 1312–1332.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, *72*(3), 821–852.

Ravazzolo, F., & Vahey, S. P. (2014). Forecast densities for economic aggregates from disaggregate ensembles. *Studies in Nonlinear Dynamics & Econometrics*, *18*(4), 367–381.

Reeve, T. A., & Vigfusson, R. J. (2011). Evaluating the forecasting performance of commodity futures prices. *FRB International Finance Discussion Paper*(1025).

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, *23*(3), 470–472.

Roy, V. (2019). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, *7*. Available at: https://arxiv.org/pdf/1909.11827.pdf.

Rubio-Ramirez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, *77*(2), 665–696.

Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. *Monographs on Statistics and Applied Probability*, *65*, 1–68.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Sims, C. A., & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 949–968.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Stone, M., et al. (1961). The opinion pool. *The Annals of Mathematical Statistics*, *32*(4), 1339–1342.

Trolle, A. B., & Schwartz, E. S. (2009). Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies*, *22*(11), 4423–4461.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, 59–73.

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.

Vats, D., & Knudson, C. (2018). Revisiting the gelman-rubin diagnostic. *arXiv preprint arXiv:1812.09384*. Available at: https://arxiv.org/pdf/1812.09384.pdf.

Vo, M. T. (2009). Regime-switching stochastic volatility: evidence from the crude oil market. *Energy Economics*, *31*(5), 779–788.

Waggoner, D. F., Wu, H., & Zha, T. (2016). Striated Metropolis–Hastings sampler for high-dimensional models. *Journal of Econometrics*, *192*(2), 406–420.

Wallis, K. F. (2005). Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, *67*(s1), 983–994.

Wang, Y., Liu, L., & Wu, C. (2017). Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. *Energy Economics*, *66*, 337–348.

Wang, Y., Wu, C., & Yang, L. (2015). Forecasting the real prices of crude oil: A dynamic model averaging approach. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2590195.

Ye, M., Zyren, J., & Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, *21*(3), 491–501.

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an emd-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), 2623–2635.