THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY • AUSTRALIA

University of New South Wales
School of Economics

Honours Thesis

Improving the Australian Macro-economic Forecasts

Author:
(Jeff) Tsz Ho Dawin Heung
Student ID: z3442867

Supervisor:
A/Prof. Valentyn Panchenko
Prof. Robert Kohn

Bachelor of Economics (Econometrics) (Honours)
and
Bachelor of Economics (Econometrics and Financial Economics)

28th October, 2016

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge, it contains no material which has been published or written by another person(s), except where I have rightly acknowledged. This Honours thesis has not been submitted for the award of any degree or diploma at the University of New South Wales, or at any other educational institution.

(Jeff) Tsz Ho Dawin Heung
28th October, 2016

Acknowledgements

First, I would like to express my deep gratitude to my supervisor, A/Prof. Valentyn Panchenko, for your unconditional support, useful suggestions and writing tips. Your patience and encouragement are invaluable and indispensible to me for completing the Honours thesis. I am very grateful to have you as one my supervisors. Without you, none of this would have been possible.

I would also like to thank Prof. Robert Kohn for spending your weekend with me discussing my thesis even though you are very busy indeed. I also thank A/Prof. Christopher Carter and Dr. Arpita Chatterjee for providing insightful feedback that have contributed greatly to my Honours thesis.

I would also like give my special thanks to the 2016 Honours cohort for your supports. I will never forget this enjoyable but hard experience with all of you.

I would like to thank my girlfriend, Phoebe Cheng, for supporting me to continue my study in Australia. Your understanding is one of my motivations to overcome every challenge in this year. I would also like to thank my dearest parents who have supported me for 22 years, especially during the Honours year. Thank you so much for your unconditional supports and encouragements. Without your enormous sacrifice, there would not be the present me.

Last but not least, praise to our Heavenly Father for his mercy and grace, for giving me the ability, wisdom and health to complete this thesis.

Contents

Abstract **1**

1 Introduction and Literature review **2**

2 Data description **6**

3 Model **9**

3.1 Motivation and model construction 9

3.2 Model in state-space respentation 11

3.3 VARMA Approximation 13

3.4 Further approximation 14

4 Methodology **15**

4.1 LASSO Regularisation 15

4.1.1 LASSO intuition 15

4.1.2 LASSO within the model 17

4.2 Estimation 17

4.2.1 Tunning parameter selection 18

4.3 Forecast 20

5 Results **23**

5.1 Regularisation of VARMA approximations by LASSO 23

5.2 Forecast 28

6 Forecast evaluation **30**

6.1 Scaled RMSE criterion 30

6.1.1 Derivation of scaled RMSE 30

6.1.2 Results of Scaled RMSE 31

6.2 Multivariate Diebold-Mariano tests 34

6.2.1 Derivation of the test 34

6.2.2 Results of Multivariate Diebold-Mariano test 35

6.3 Rolling Window Forecast 38

6.3.1 Introduction of rolling window 38
6.3.2 Results of rolling window 38

7 Extension: Model averaging **40**

8 Conclusion and discussion **43**

A Tables **45**

B Figures **52**

List of Figures

1.1 Graphical introduction 5

4.1 Graphical explanation of LASSO constrainted minimisation problem . 16
4.2 Illustration of rolling window for tunning parameter selection 19

5.1 LASSO shrinkage (LASSO-VAR(10)) by using 100-folds CV 23
5.2 LASSO shrinkage (LASSO-VAR(10)) by using one step ahead RMSE method 24
5.3 LASSO shrinkage (LASSO-VARMA(6,1)-VAR(10)) by using 100-folds CV method 25
5.4 LASSO shrinkage (LASSO-VARMA(6,1)-VAR(10)) by using one step ahead RMSE method 26
5.5 Illustration of estimation and predictive sets 28

6.1 Illustration of rolling window analysis 38

B.1 (CV) 5 years ahead forecast of $\triangle Y_t$ 52
B.2 (CV) 5 years ahead forecast of π_t 53
B.3 (CV) 5 years ahead forecast of $\triangle U_t$ 53
B.4 (CV) 5 years ahead forecast of $\triangle C_t$ 54
B.5 (CV) 5 years ahead forecast of $\triangle EXR_t$ 54
B.6 (CV) 5 years ahead forecast of $\triangle I_t$ 55
B.7 (CV) 5 years ahead forecast of $\triangle E_t$ 55
B.8 (CV) 5 years ahead forecast of $\triangle R_t$ 56
B.9 (RMSE) 5 years ahead forecast of $\triangle Y_t$ 56
B.10 (RMSE) 5 years ahead forecast of π_t 57
B.11 (RMSE) 5 years ahead forecast of $\triangle U_t$ 57
B.12 (RMSE) 5 years ahead forecast of $\triangle C_t$ 58
B.13 (RMSE) 5 years ahead forecast of $\triangle EXR_t$ 58
B.14 (RMSE) 5 years ahead forecast of $\triangle I_t$ 59
B.15 (RMSE) 5 years ahead forecast of $\triangle E_t$ 59
B.16 (RMSE) 5 years ahead forecast of $\triangle I_t$ 60

B.17 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$ 60
B.18 CV- LASSO-VAR(10) 5 years ahead averaged forecast of π_t 61
B.19 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$ 61
B.20 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$ 62
B.21 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$ . . 62
B.22 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ 63
B.23 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$ 63
B.24 CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ 64
B.25 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$ 64
B.26 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of π_t 65
B.27 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$ 65
B.28 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$ 66
B.29 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$ 66
B.30 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ 67
B.31 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$ 67
B.32 CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle R_t$ 68
B.33 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$ . . . 68
B.34 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of π_t 69
B.35 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$ . . 69
B.36 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$ . . 70
B.37 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$ 70
B.38 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ . . . 71
B.39 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$ . . 71
B.40 RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ . . . 72
B.41 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years steps ahead averaged forecast of $\triangle Y_t$ 72
B.42 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of π_t 73
B.43 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$ 73

B.44 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$ 74
B.45 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$ 74
B.46 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$ 75
B.47 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$ 75
B.48 RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle R_t$ 76

List of Tables

2.1 Data description 7
2.2 Notation and Transformation 8

3.1 State variable 12

5.1 Parameter space 26

6.1 sample standard deviation 31
6.2 Multivariate DM-test results 35
6.3 Elements in $\bar{\mathbf{d}}$ 36

7.1 Multivariate Diebold-Mariano test for average predictions 41

A.1 Value of tunning parameter λ of LASSO-VAR(10) by using 100-folds cross-validation 45
A.2 Value of tunning parameter λ of LASSO-VAR(10) by using one step ahead RMSE method 45
A.3 Value of tunning parameter λ of LASSO-VARMA(6,1) -VAR(10) by using 100-folds CV 46
A.4 Value of tunning parameter λ of LASSO-VARMA(6,1) -VAR(10) by using one step ahead RMSE method 46
A.5 Scaled RMSE 47
A.6 Scaled RMSE (cont.) 48
A.7 RMSE of rolling window 49
A.8 RMSE of rolling window cont. 50
A.9 Elements in $\bar{\mathbf{d}}$ for average predictions 51

Abstract

Forecasting has been always a popular but challenging task for econometricians. Although one-step or multi-steps ahead forecasts are readily available from many some statistical software packages, high forecast accuracy is difficult to achieve. From a traditional economics prospective, the most notable reason for this is that agents' decision making in the economy is unobservable and varying. From a statistical point of view, the econometric model might involve too many variables which is called over-parameterisation. First, I motivate an econometric model using agents' expectations. Understanding the agents' expectations and how they influence the realisations of actual outcomes may improve forecasting performance.

Second, I apply the penalised regression technique, pioneered by Tibshirani (1996) - "least absolute shrinkage and selection operator"(LASSO), for reducing the parameter space of the large econometric model. Through the LASSO regularisation, I demonstrate that the forecasting performance is significantly improved for all economic series, especially in long-term prediction. I also use several evaluation tests to further confirm that regularised model is better in terms of predictions. Last, I show that the forecast accuracy can be improved averaging the predictions of LASSO regularised and over-parameterised models.

Chapter 1
Introduction and Literature review

Over the last few decades, forecasting macroeconomic variables gained more importance as it became influential in policy decisions. As a starting point in this thesis, I motivate a statistical model using agents' behaviour and decision making processes. Agents make their decisions based on their own expectations of the future economic outcomes and the aggregate behaviour contributes to the actual economic outcomes. However, agents' expectations about the future economic outcomes are varying because agents will update their expectation based on the inflow of the new information. For example, agents tend to spend more today if they believe the news that the inflation rate is likely to be higher tomorrow.

Several ways of modelling agents' expectations have been put forward before. Examples include inferring person's behaviour intention and expectation using the survey results by Manski (2004); specifying a closed-form solution for VIX (Volatility Index) by Nossman and Wilhemsson (2009); and using dynamic stochastic general equilibrium (DSGE) models to "quantify" the latent agents' expectation by Schorfheide and Del Negro (2012).[1] However, all these model heavily rely on behavioural assumptions. In particular, Manski (2004) asks hypothetical questions in the survey for infering agents' expectations but agent might react differently in a hypothetical environment and in reality. In addition, Nossman and Wilhemsson (2009) uses VIX to infer agents' expectation, but VIX mainly reflects agents' expectations about the financial market.[2] DSGE models provide a good structural description of the economy, but they rely on full rationality assumption. Moreover, they contains some unobservable components such as tastes, preference, etc.

The link between agents' expectations and actual economic outcomes is explicitly modelled in Gibbs and Kulish (2015). In their model, the current economic outcome is a linear function of last period's economic outcome, agents' expectations of future

[1]Schorfheide and Del Negro (2012) suggests that DSGE models summarises latent variables including tastes, preferences and technologies which indirectly alter agents' expectations and thus change their decision rules.

[2]VIX uses options price to infer agents' expectations about the financial market. This provides insight about how agents expect "financially-related" economic variables but provide less information about the expectation of some non-financially related variables (e.g. unemployment rate)

economic outcome and current unexpected events. For simplicity, they assume the linear relationship, and therefore each current economic variable can be written in a linear regression form (e.g. AR process). The first contribution of this paper is to specify the reduced-form of Gibbs and Kulish (2015)'s model with the minimal representation of DSGE models in state-space form suggested in Giacomini et al. (2013). Equivalently, the state-space model can be re-expressed in a Vector-autogression (VAR) representation which is originally suggested by Sims(1980).

However, VAR including multiple macro-variables may suffer over-parameterisation. In order words, large VAR system requires estimating too many parameters for a given data size. There are two main problems of over-parameterisation. First, Nicholson et al.(2014) criticises that over-parameterisation enlarges the parameter space exponentially (aka curse of dimensionality), which drains the available degrees of freedom rapidly. Therefore, using VAR with a large parameter space to compute prediction for low frequency macro-economic data with a relativelly small sample size is statistically infeasible. Second, Robertson et al. (1999) and Trevor and Thorp (1986) suggest that over-parameterised model results in poor and imprecise out-of-sample forecast even though in-sample fit is good. Hence, constructing a large model with many variables do not necessarily generate a prediction with high accuracy.

Addressing the problem of overly large parameter space, Tibshirani (1996) pioneers a regularisation with norm-1 penalty methodology - 'least absolute shrinkage and selection operator '(LASSO). LASSO provides sparse solutions reducing some estimates to zero. To achieve this, LASSO introduces an additional penalty term to the Loss function (e.g. sum of squared residuals).[3] The rationale of regularising some estimates to zero is to decease the parameter space by eliminating less informative variables in order to improve forecast accuracy, which enables researcher to deal with large parameter space models or even systems of equations. In particular LASSO becomes popular in mulitvariate time series modela, especially vector-autoregression (VAR), see, e.g., Hsu et al. (2008), Haufe et al. (2009), Basu and Michailidis (2015) etc. The automatic shrinkage provides variable and lag selections, which are typically practical issues in traditional VAR models. Nicholson et al. (2014) suggests some augnmented LASSO penalty functions, applied to the group of parameters and imposing structural restrictions. This augnmentation provides a greater flexibility to incorprate the true dynamic system in the economy. In addition, Ren and Zhang (2010) uses adaptive LASSO method by adding individual penalty weights to each parameter in attempt to improve the properties of the

[3] A more detail explanation of LASSO and other penalty methodologies will be discussed in Chapter 4.

parameter estimates. However, these approaches require high computational cost and potentially accumulates more estimation errors. For instance, in adaptive LASSO, the weight corresponding to each coefficient is the reciprocal of the ridge regression estimates. Hence, there 2 are stages involved in the optimsation problem and more estimation errors are more likely to be generated. Therefore, I decide to use the simpliest LASSO algorithm to generate prediction for the Australian economy, which has not been explored much in the literature. The second contribution of this paper is to examine which forecasting method provides better short and long- term predictions for the Australian macro-economic variables.

There are several reasons of why LASSO is preferable than other variable selection algorithms. First, it does not require any prior information about the parameters. For instance, in Bayesian inference, any inappropiate selection of the prior densities might result in slow or non-convergence to the true parameter. Nicholson et al.(2014) suggests that LASSO-VAR is preferable than Bayesian-VAR (BVAR) in terms of better performance in multi-steps ahead forecasts and is computationally tractable in large parameter space model. Second, LASSO provides an automatic variable and lag selections with low computational cost. Commonly used variable/lag selection method such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) are very expensive when the parameter space is large because they require numerious model combinations. Alternative approaches popular in applied literature include "general -to-specific" (*Gets*) method, which eliminates variable sequentially based on their statistical signifiance, see Krolzig (2003). However, numerical optimisation is inaccurate in over-parameterised models due to multimodality of the likelihood functions (Chen and Chan (2011)), which makes this procedure unreliable. There are other regularisation methods, such as ridge regression, see Tikhonov (1963). The big difference between LASSO and ridge regression is that LASSO imposes norm-1 constraint, while ridge regression imposes norm-2 constraint.[4] Practically, LASSO regularisation does both shrinkage and variable selection, while ridge regularisation only does shrinkage. Hence, with LASSO we can get better understanding which variable/lags are important for forecasting.

To summarise, the purpose of this thesis is to construct a reduced-form macroeconomic model with unobservable agents' expectations. In addition, this paper also seeks to investigate whether the LASSO methods provide predictions with higher forecast accuracy in terms of long and short-run forecasts. The forecasting results are evaluated in out-of-sample context by using fixed sample and rolling

[4]Graphically, in 2 variable case, the norm-1 constraint can be visualised as a diamond shape and the norm-2 constraint is just a circle.

window sample. As a new forecast evaluation criterion, I use scaled RMSE which incorporates heterokedasticity and provides easy and clear interpretation. The main goal of my paper is to assess forecasting performances of different models on the Australian macroeconomic data and understand which variables and at which lags are most relevent for prediction. Figure 1.1 provides a graphical illustration of the logical flow in my paper.

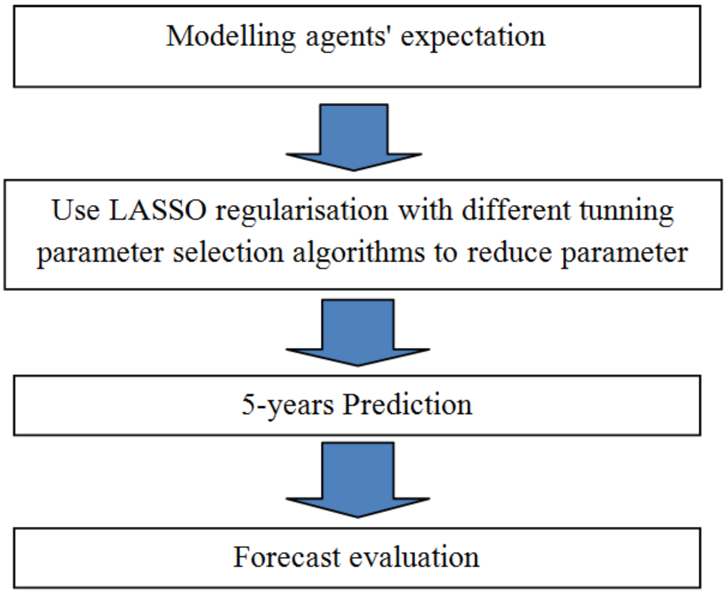

Figure 1.1: Graphical introduction

This thesis is organised as following. Chapter 2 gives data description. Chapter 3 motivates a reduced-form version of Gibbs and Kulish (2016) model and transform the the state-space model to Vector-autoregression moving-average (VARMA). Chapter 4 reviews the LASSO algorithm, provides LASSO-VAR expression with respect to the models and explain the estimation and forecast methodologies. Chapter 5 report the in-sample and out-of-sample results. Chapter 6 compare models by several forecast evaluation methods. Chapter 7 demonstrates the averaged predictions. Chapter 8 gives conclusion and discussion.

Chapter 2
Data description

The dataset I use is a simplified version of the variable list constructed by Trevor and Thorp (1986). I initially decide to include the real GDP as it is a representative indicator of reflecting the macro-economic conditions in Australia. I use real GDP in this paper in order to avoid the confounding effect from inflation on GDP. Putting variables in a system is necessary to ensure that all variables provide useful information to one another, that is, no variable is completely independent to other variables. I justify the inclusion of additional variables based on economic intuition and past literature.

Keynes (1937) explicitly models the relationship between inflation and real GDP relationship by using the aggregate demand and aggregate supply model (AS-AD model). This AS-AD model explains the relationship between the real GDP and inflation. Hence, any new information coming from the inflation rate might provide useful information about the future GDP in the economy.

I further add unemployment rate since it has a strong link with the inflation rate. The Philips curve, initiated by Philips (1958), suggests a negative relationship between the inflation rate and the unemployment rate. A higher unemployment rate implies that agents' wealth or purchasing power is decreasing. Hence, agents tend to reduce their spending, which, in turn, decreases the inflation rate subsequently. Bodman (1998) provides econometric results evidently supporting the solid link between the unemployment growth and GDP in Australia. His analysis shows that job destruction is strongly persistent during recession periods but decreases significantly in a boom. The results confirm the conjecture that unemployment rate is directly correlated with the real GDP.

Next, I include the real interest rate because it is a lead indicator of the business cycle. Neumeyera and Perri (2005) studies the relationship between interest rate and business cycle in the Australian economy.[1] and they show that the real interest rate is fairly procyclical and provides "signal" prior to the future realisation of the real GDP. One justification is that business owners observe the interest rate to decide

[1] Neumeyera and Perri (2005) uses log GDP as a proxy for the Australia business cycle.

how much capital they are investing or how much money they are borrowing from investors. Therefore, the interest rate influences the future business cycle in the economy. Hence, including the real interest rate in my model helps to capture the fluctuations of GDP within the economy business cycle. Furthermore, as the real interest rate affects households' consumption decisions, I also include the households' final private consumption in order to further understand the agents' behaviour.

Lastly, I include the USD/AUS exchange rate, Australian import and export values for summarising the information from foreign investors' expectations and impacts from other economies on the Australian macro-economy. For instance, an appreciation in the Australian dollar implies that foreign investors hold more Australian money. This could potentially reflect that foreign investors have positive expectation about the Australian macro-economy so they are more likely to invest more in the future. The inflow of capital stimulates the Australian economy by creating more jobs and increasing the inflation rate.

All Australian macroeconomic series are in quarterly frequency ranging from 1978Q1 to 2013Q1 with 140 observations in total. Most of my data is obtained from the Federal Reserve Economic Data (FRED) - St. Louis website. The inflation rate and final private consummption are obtained from the Australian Bureau of Statistics and Organisation for Economic Co-operation and Development (OECD) website respectively. All series are summarised in Table 2.1 below.

Table 2.1: Data description

	Unit	**Seasonal Adjustment**
Real GDP	Billions of Chained 07-08 AUS Dollars	yes
Inflation	Percentage	no
Unemployment rate	Percentage	yes
Consumption	Million	yes
Exchange Rate	US Dollar/ AUS Dollar	no
Import	National currency	yes
Export	National currency	yes
Real Interest Rate	Percentage	no

In order to achieve stationary series, all series (except inflation) are transformed into first logarithm difference in order to prevent spurious regression and time-varying parameters. The stationarity property is checked by Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests which are shown in Table 2.2. I also specify the notation of each variable I am using in the following chapters.

Table 2.2: **Notation and Transformation**

	Notation	ADF test	KPSS test
$\triangle$ Real GDP	$\triangle Y_t$	stationary	stationary
Inflation	π_t	stationary	Non-stationary
$\triangle$ Unemployment rate	$\triangle U_t$	stationary	stationary
$\triangle$ Consumption	$\triangle C_t$	stationary	Non-stationary
$\triangle$ Exchange Rate	$\triangle EXR_t$	stationary	stationary
$\triangle$ Import	$\triangle I_t$	stationary	stationary
$\triangle$ Export	$\triangle E_t$	stationary	stationary
$\triangle$ Real Interest Rate	$\triangle R_t$	stationary	stationary

In the ADF test, all null hypothesis (unit root existence) are rejected, confirming that all series are stationary. However, since the ADF test has low power to distinguish between highly persistent stationary and non-stationary series, I use the KPSS test to further assess the stationarity of all series. From the KPSS test results, I reject the null hypothesis, the series is stationary, for π_t and $\triangle C_t$ while others' align with the ADF test results. To summarise, according to the stationarity test results, π_t and $\triangle C_t$ are potential non-stationary but other series are believed to be stationary.

Chapter 3
Model

3.1 Motivation and model construction

I begin the motivation of my model by briefly explaining the most conventional way that agents build their expectations for future economic outcomes. When the current economic outcomes are observed by agents,[1] they form expectations about future economic outcomes based on the updated information. Importantly, any inflowing new information changes agents' decision making about future endowments (e.g saving and consumption) because their beliefs of the future are affected. Therefore, the future economic outcomes are also affected subsequently. Gibbs and Kulish (2015) implements this idea by suggesting that the true economic outcomes are determinded by yesterday's economic outcomes and agents' expectations. The general form for such macro-economic model can be specified as:

$$X_t = f(X_{t-1}, E^*(X_{t+1}), u_t) \tag{3.1}$$

where X_t is the current economic outcomes, X_{t-1} is the past economic outcomes, $E^*(X_{t+1})$ is agents' expectations and u_t is current unexpected event.

Moreover, Gibbs and Kulish (2015) further assumes that agents may have heterogenous expectations. Namely, they use two types of agents in the economy - rational expectation and adaptive learning. Agents using rational expectation are forward-looking and utility-maximising. They use fundamental economic theory to make rational prediction based on the most updated information in the economy, assuming that the structure of the macro-economy is known and every agent uses rational expections. For instance, agents expect that a tightening monetary policy is being adopted soon if the Reserve bank announces the current inflation rate is higher than the normal rate as this would happen in a standard model. However, assuming all agents in the economy use rational expectation to predict future economic outcome is unrealistic, because some agents might not be entirely rational or use simpler predictive rules, e.g., AR-type models. Moreover, agents do not always incorporate government's announcement in computing their rational expectations when there is

[1] In my model, agents are all assumed to be a forward looking and utility-maximising individual

a problem of low credibility. [2]

Therefore, Gibbs and Kulish (2015) specifies that some agents use adaptive learning to form their expectations instead of using pure rational expectation.[3] Evans and Honkapohja (2001), who are the first to parameterise this concept, suggests that adaptive learning agents use linear regression with unknown parameters to do prediction. Incorporating the adaptive learning agents also further assumes that agents are not only forward looking but also backward looking, using past information, which provides a better understanding of the agents' behaviour and fits the empirical data.

Modelling expectation is not an easy task because agents' rational expectations are not explicitly observed in most of the data. Moreover, rational expectation model often include latent variables such as tastes and preferences etc. Broze, Gourieroux and Szafarz (1995) provide feasible methodologies to estimate coefficients with respect to the unobservable agents' expectation directly. However, they have to impose a restriction identifying assumption to do this. Moreover, it is not obvious that quantifying the unobervable expectation directly improves prediction, which is my main concern here. I specify the reduced form of Gibbs and Kulish (2015) model as follows

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \ldots + A_7 Z_{t-7} + \gamma \Gamma_t + \varepsilon_t, \varepsilon_t \sim iid \quad \mathcal{N}(0, \Delta_{\varepsilon_t}) \tag{3.2}$$

where Z_{t-i} $(i = 0, \ldots, 7)$ are vectors of observable variables, Γ_t is vector of unobservable states, A_i $(i = 1, .., 7)$ are coefficient matrices, γ is a coefficient matrix, ε_t is a vector of independently and identically distributed noise and Δ_{ε_t} is a diagonal variance-covariance matrix. The dimensions and choices of observable variables are explained in next section.

In the reduced form representation, I combine the past eonomic outcomes and adpative learning components together because they are essentially the same. Motivated by Giacomini et al. (2015), I include multiple lags in the VAR system. This paper finds that different agents use different information at different lags to compute their forecast because agents are inattentive. For instance, on average, only

[2]Gibbs and Mariano (2015) provides an example that the Argentina government's annoucement of the inflation rate is lower than the actual inflation rate, so agents do not believe the official annoucement.

[3]Adaptive learning means agent uses past information (e.g. data) to form his own forecast. For example, they might use simple linear regression to compute prediction.

40% to 50% of the participants revise their forecast period-by-period. However, this inattentive behaviour does not necessary imply agents use past information to build their expectations; instead, they might simply treat the last period's prediction as the updated prediction. Nevertheless, explaining such inattentive behaviour, Reis (2006) suggests that agents are aware of the new information, but they do not incorporate updated information due the corresponding costs incurred, especially time, money and physical effort. Hence, agents might use this limited-information approach to form their expectations about future economic outcomes. Coibion and Gorodnichenko (2012) expand this idea to a macro-economic context. They find numerical evidence that various agents show delayed response to forecast any macro-economic shock. Based on this, I include up to 7 lags in the system, which, in economic terms, corresponds over two years of most recent information incorporated by agents in their forecasts.

In reality, the link between agents' expectations and the current economic outcomes is complex. For tractability, Evan and Honkapohja (2001), who use adaptive expectations, suggest a simplified linear representation. Therefore, a linearised state-space model is useful in modelling the macro-economy with the heterogenous expectations based on the linear approximation assumption. The latent part of the rational expectations model, which includes unobservable taste and preference, is reduced to the latent variable, Γ_t, which I discuss in details in the next section.

3.2 Model in state-space respentation

In this section, I provide a detail construction of my state-space model. Motivated by the justification in the previous section, equation 3.2 is expressed as the measurement equation representation below:

$$\begin{bmatrix} \triangle Y_t \\ \pi_t \\ \triangle U_t \\ \triangle C_t \\ \triangle EXR_t \\ \triangle I_t \\ \triangle E_t \\ \triangle R_t \end{bmatrix} = A_1 \begin{bmatrix} \triangle Y_{t-1} \\ \pi_{t-1} \\ \triangle U_{t-1} \\ \triangle C_{t-1} \\ \triangle EXR_{t-1} \\ \triangle I_{t-1} \\ \triangle E_{t-1} \\ \triangle R_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} \triangle Y_{t-2} \\ \pi_{t-2} \\ \triangle U_{t-2} \\ \triangle C_{t-2} \\ \triangle EXR_{t-2} \\ \triangle I_{t-2} \\ \triangle E_{t-2} \\ \triangle R_{t-2} \end{bmatrix} + \cdots + A_7 \begin{bmatrix} \triangle Y_{t-7} \\ \pi_{t-7} \\ \triangle U_{t-7} \\ \triangle C_{t-7} \\ \triangle EXR_{t-7} \\ \triangle I_{t-7} \\ \triangle E_{t-7} \\ \triangle R_{t-7} \end{bmatrix} + \gamma \begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{Y,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{U,t} \\ \varepsilon_{C,t} \\ \varepsilon_{EXR,t} \\ \varepsilon_{I,t} \\ \varepsilon_{E,t} \\ \varepsilon_{R,t} \end{bmatrix}$$

where A_i $(i = 1, .., 7)$ are 8 x 8 coefficient matrices and γ is a 8 x 3 coefficient matrix. I also assume that all unknown parameters are not varying overtime.

In the measurement equation, it expresses the relationship between heterogenous expectation and current economic outcomes in a linearised VAR system which contains adaptive learning and rational expectation components. First, the combination of Z_{t-i} $(i = 1, \ldots, 7)$ summarises the inattentive adaptive learning type agents using the past information of the economy which is used for computing their own perceived law of motion. Second, from Giacomini (2013), Γ_t provides a minimal state representation of the DSGE models specified in An and Schorfheide (2007), which summarises agents' tastes and preferences. Hence, minimal state representation of the DSGE models Γ_t can "approximate" agents' rational expectations. The description of the state variables is provided in Table 3.1.

Table 3.1: State variable

Variable	Meaning
z_t	Technology innovation
g_t	Government spending
r_t	Gross interest rate

Following the expression in Giacomini (2013), the DSGE model can be simplified as the following transition equation:

$$\begin{bmatrix} z_t \\ g_t \\ r_t \end{bmatrix} = \rho \begin{bmatrix} z_{t-1} \\ g_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} v_{z,t} \\ v_{g,t} \\ v_{r,t} \end{bmatrix}$$

The above matrices system can be simplified as the expression below :

$$\Gamma_t = \rho \Gamma_{t-1} + v_t, v_t \sim iid \quad \mathcal{N}(0, \Lambda_{v_t}) \tag{3.3}$$

where ρ are 3 x 3 state coefficients matrices, v_t is 3 x 1 vector containing independently and identically dritributed noise and Λ_{v_t} is a 3 x 3 variance-covariane matrix.

The transition equation captures the dynamic growth of the unobservable factors in an AR(1) process. This expression provides a flexible framwork that allows agents to "update" their rational forecast coping with an evolution of the fundamental factors. For example, using the inflation case again, agents might change their outlook from short-term to long-term. Hence, this particular change makes agents to be less responsive to the Reserve Bank announcement about the inflation rate, and results in actual inflation rate being higher than the target one. The AR(1) process is able to approximate the change of agents' taste and preference, based on the assumption of the DSGE model in Schorfheide and Del Negro (2012).

3.3 VARMA Approximation

In the previous section, I construct my model in state-space representation which can be estimated by the Kalman filter. However, obtaining the MLE estimates of each coefficient by the Kalman filter requires stationarity constraint for each coefficient, which is complex in the AR(7) process. Moreover, applying LASSO (penalised likelihood) for sparsity in this setting is more complicated which increases the computation cost. Also, the Kalman filter heavily relies on the Gaussian assumption of the innovations which might not necessarily hold in reality. Penzer and Shea (1997) suggest that rounding-off the errors term could generate negative-definite variance-covariance matrices. Hence, I cannot do the Cholesky decomposition of the variance-covariance, which makes the updating process computationally infeasible, see Ansley and Kohn (1990).

Nevertheless, the state space model can be re-written in the VARMA representation. By assuming time-constant coefficients, Jerz, Garcia-Hreisaux and Casals (2010) propose two algorithms to transform the state-space model into VARMAX representation. Moreover, Karapanagrotidis (2014) shows such transformation in a univariate case by combining the two different noises together.[4] Havery (2006) specifies a general transformation in which the state-space representation can be reduced into the VARMA representation which also inherites the original stationarity assumption. Such transformation requires decreasing the order of the AR component by one. Therefore, the reduced form VARMA is

$$Z_t + \tilde{A}_1 Z_{t-1} + \ldots + \tilde{A}_6 Z_{t-6} = \eta_t + \theta^* \eta_{t-1}, \eta_t \sim iid \quad \mathcal{N}(0, \Lambda_{\eta_t})$$

where η_t is the combined innovations with MA(1) process, θ^* is the 8 x 8 coefficient matrix with respect to the MA component, A_i is 8 x 8 coefficient matrix with respect to Z_{t-i}, $i = 1, \ldots, 6$.

One potential drawbacks of this transformation is losing the structural interpertation as the original structural parameters are re-specified into a single VARMA system.

[4]The two noises are from the state and measurement equations.

For example, the two noises from the transition and measurement are unrecoverable after merging into a single VMA(1) process. Nevertheless, the focus of this thesis is the forecasting performance of different models, including, regularised models, so sacrificing the structural interpretation is not of concern in this thesis.

3.4 Further approximation

Estimating VARMA(6,1) requires higher computational cost because the innovations with respect to the MA component are unobservable. The MA process cannot be simply estiamted by the OLS. Hence, the latent innovation process enhances the difficulty of estimating the parameters. Another way to estimate VARMA model is to re-specify the entire system into state-space representation again and implement the Kalman filter to construct the likelihood. However, specifying the stationarity constraints of the coefficients in the VARMA(6,1) model for the Kalman filter is still fairly complicated. The maximum likelihood estimates are inaccurate without the stationarity constraints. Nevertheless, theoretically, the VARMA(6,1) can be represented by VAR(∞).[5] Indeed, VAR(∞) is infeasible and I use VAR(10) to approximate it.[6]

Although the approximations alleviate the high computational cost of estimating the MA component, the drawback of doing approximation is over-parameterisation. Simply using the VAR(10) to appximate VARMA(6,1) includes large series of lags in the system, leading to over-parameterisation. This may provide good in-sample fit, but the out-of-sample forecasts will not be precise (Trevor and Thorp (1986) and Robertson and Tallman (1999)).

As I mentioned before, LASSO is a powerful tool that reduces the parameter space by setting some coefficients to zero. Hence, LASSO provides a feasible solutioin to improve simple over-parameterised models approximating the VARMA(6,1) model. In the next chapter, I discuss LASSO methodology in details.

[5] Using the Maclaurin series, the MA(1) component can be inverted into AR(∞) if the coefficient with respect the MA(1) is less than 1 which is assumed in this setting. I also find out that the estimation result of VARMA(6,1) aligns with this assumption.

[6] I try 12 and 13 lags, and it seems 10 lags is sufficient.

Chapter 4
Methodology

4.1 LASSO Regularisation

Before exploring the estimation methodologies, I provide a brief review of LASSO regularisation and specify my model in a LASSO problem representation.

4.1.1 LASSO intuition

In this section, I briefly review the intuition of LASSO and provide an example with two dimensions only for the sake of simple clarification.

Tibshirani (1996) introduces a norm-1 regularisation method in which modeller scarifices the unbiasness for low parameter space, so the LASSO estimates are bias. Neverthless, the main purpose of this thesis is to compute out-of-sample forecast, so unbiasness of the estimates is not my main concern. In a two dimensions case, the LASSO coefficients minimise the sum of squared residual with a penalty term (Objective loss function):

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \sum_{i=1}^{M} \{y_i - \beta_1 - \beta_2 x_i\}^2 + \lambda \sum_{j=1}^{2} |\beta_j|$$

or equivalently:

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \sum_{i=1}^{M} \{y_i - \beta_1 - \beta_2 x_i\}^2 + \lambda ||\beta||_1$$

Where y_i and x_i are data and λ is the tunning parameter that controls the amount of shrinkage : $\hat{\beta}^{LASSO}$ is equal to $\hat{\beta}^{OLS}$ if λ is 0; $\hat{\beta}^{LASSO}$ is equal zero if λ approaches ∞ (Tibshirani et al. (2002)). The LASSO problem can be equivalently written as :

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \sum_{i=1}^{M} \{y_i - \beta_1 - \beta_2 x_i\}^2$$

Subject to

$$|\beta_1| + |\beta_2| \leq f(\lambda) = c$$

The new expression shows that LASSO problem is a constrainted minmisation problem, which can be further explained graphically in Figure 4.1. [1]

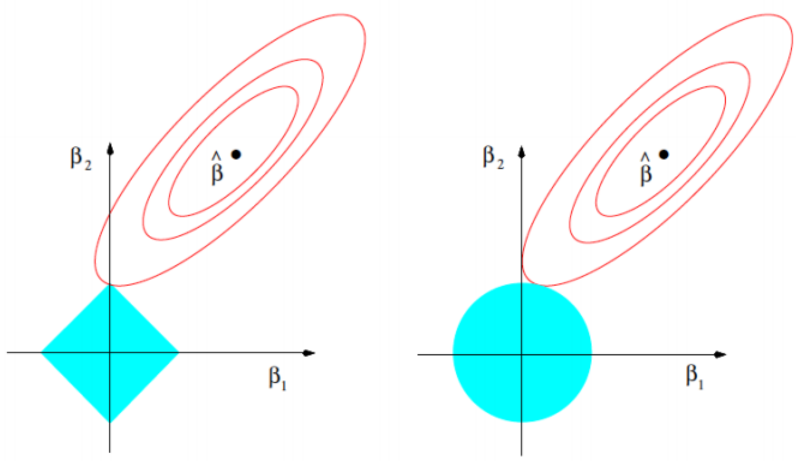

Figure 4.1: Graphical explanation of LASSO constrainted minimisation problem

Figure 4.1 shows a comparison between Lasso (left) and Ridge (Right) regression. The red ellispe curves on both pictures are the term of sum of residuals (SSR). Each curve represents different combinations of β_1 and β_2 with the same value of SSR. The SSR decreases when the curves are shrinking towards to the centre and therefore the centre point of the curves is the $\hat{\beta}^{OLS}$ where the lowest value of SSR is located. However, both optimisation problems change to constrainted minimisation problems when the LASSO and ridge regression introduces the constraint terms. The norm-1 penlty function forms a diamond shape constraint while the ridge regression's is just a circle.

As the optimal point is located inside the red curves in the unconstrainted optimasation problem, the solution of constrainted optimasation problem must lie on the boundaries in both pictures, which provide the smallest distance to the solution of the unconstrainted optimasation. For the LASSO problem, the solution is lied on the top edge of the diamond constraint, where $\hat{\beta}_1$ is zero and $\hat{\beta}_2$ is a positive number. This explains why the LASSO is able to provide sparse solution. In contrast, the

[1]This figure is cited from Tibshirani et al. (2002)

$\hat{\beta}^{Ridge}$ does not provide sparse solution since both $\hat{\beta}_1$ and $\hat{\beta}_2$ are non-zero estiamtes. To conclude, as long as the unique diamond shape of the LASSO's constraint, it is able to provide sparse solution and thus reducing the parameter space.

4.1.2 LASSO within the model

Without losing generality, LASSO is also appplicable in my model which is a VAR system (The two approximations of VARMA(6,1) by VAR(∞) will be explained in next section.). I estimate the coefficients in the VAR system by running the LASSO regression for each equation.[2] Hence, the LASSO solution of each equation can be expressed as:

$$\hat{a}_k^{LASSO} = \arg\min_{a_k} \sum_{j=1}^{M} \{z_{t,k,j} - \sum_{l=1}^{L} \sum_{i=1}^{p} a_{k,l,i} z_{t-i,l,j}\}^2 + \lambda_k ||a||_1$$

Here $z_{t,k,j}$ is the j-th observation of the k-th variable in Z_t (for instance, $z_{t,1}$ means the first log difference of real GDP at time t), p is the order of the AR process, M is the total number of observations, L is the total number of time series in the model, λ_k is a scaler that represents the tunning parameter with respect to the k-th equation, $a_{k,l,i}$ is the k-th row and l th column coefficient in the i-th lag coefficient matrix, $||a||_1$ is a norm-1 summation of all coefficients in the k-th equation and $\hat{a}_k$ is the LASSO estimates of all coefficients with respect to k-th equation in the VAR system. The coefficients in each equation are likely to be different from one another since the dependent variables are not the same. Therefore, imposing the restriction that each equation contains the same value of tunning parameter is too insensible and impractical. Hence, I assume that the tunning parameter of each equation is different from one another so running LASSO equation-by-equation to establish the system is reasonable in this setting.

The estimation section can be divided into two parts. The first part explains various ways to estimate VARMA(6,1) and how LASSO can be applied in this context. The second part provides a simple demonstration of the implementation of out-of sample predictions based on the models specified in the first part.

4.2 Estimation

I am applying LASSO to each approximation and compare the results with all unregularised models. The models that I estimate are specified as follows:

[2] I have eight variables so I have eight equations in the VAR system

1. **VAR(10)**
   - This model is using VAR(10) to approximate the VARMA(6,1) without any sparsity. I estiamte this model equation-by-equation in Matlab. The combination of eight AR(10) equations establishes the VAR(10) system.[3]
2. **LASSO-VAR(10)**
   - Estimating equation-by-equation, I apply LASSO for each AR(10) equation. This provides sparse solution in the VAR(10) model which is an approximation of VARMA(6,1).
3. **VARMA(6,1) -VAR(10)**
   - I first estimate and record the the residuals of VAR(10). As in Chan and Chen (2011), I treat the residual as the pesudo innovations with respect to the MA component in VARMA(6,1). This provides a more parsimonious solution without using LASSO. However, the drawback of this process is higher estimation error since it requires a two-stages regression.
4. **LASSO-VARMA(6,1) -VAR(10)**
   - Similarly, I treat the residuals from VAR(10) as the pesudo innovations with respect to the MA component in VARMA(6,1). The only difference is that I apply LASSO in this model for getting the sparse solution.

4.2.1 Tunning parameter selection

The difficulty of implmenting LASSO is to choose the optimal constraint, which is equivalent to choosing the optimal tunning parameter λ, that best fits the data. Tibshirani (1996) suggests using K-folds cross-validation method to select the optimal tunning parameter. However, since the data of dependent varible is also the data of independent variable, deleting a particular observation leads to the problem of missing observation. Nevertheless, I conduct an "augnmented" cross-validation. This is best illustrated with an example; consider an AR(2) process. Suppose I have observations y_i ($i = 1, \ldots, 20$). I want to predict y_{10} and use the rest observations to fit the model. Then, I exclude y_{10} as the depdendent variable but include it as the independent variable.

[3] I first estimate the AR(10) in which the dependent variable is the change of real GDP. Following the order in Table 2.2, I keep fitting the data by AR(10) model corresponding to different dependent variables. Then, I combine 8 AR(10) equations and construct the whole VAR(10) system. The equation-by-equation estimation on Matlab provides the same estiamtes as estimating as a system on Eviews.

This augnmented cross-validation method solves the problem of missing observations, but the observations might be serially correlated due to the time dependence.[4] Addressing the time dependence, I use similar method of Nicholson et al. (2014) - choosing the tunning parameter by minimising one-step ahead RMSE in a rolling based analysis.

I divide the whole sample into three parts. I use the first part to estimate the parameters, the second part to compute the RMSE for selecting the optimal tunning parameter and the last part for out-of sample forecast comparison (more details in chapter 6).[5] Figure 4.2 shows the procedure of tunning parameter selection in a rolling manner. The estimation procedures are specified as follow:

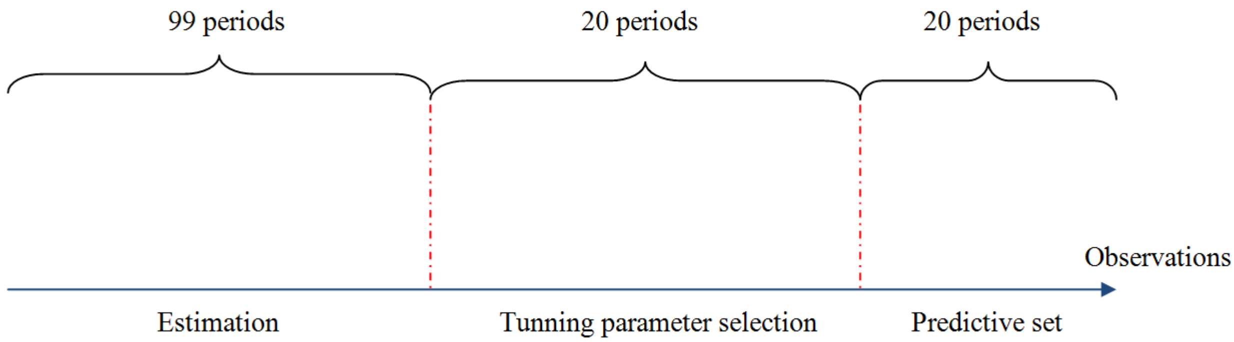

Figure 4.2: Illustration of rolling window for tunning parameter selection

1. Using the observations in the estimation set to fit the model.
2. Keep computing one-step ahead squared forecast error in a rolling manner till the end of the tunning parameter selection set and then compute the root-mean squared error (RMSE).
3. Find the optimal tunning which corresponds to the lowest RMSE.

To summarise, the CV and RMSE algorithms have different estimation advantages and disadvantages. The augnmented cross-validation is more computationally easy and stable, but the selections might be affected by the time dependence. RMSE algorithm can tackle the time dependence, but the optimal tunning parameter is

[4]In order to balance the computational cost and accurate tunning parameter selection, I use 100-folds in the augnmented cross-validation method.

[5]Nicholson et al. (2014) suggests to divide the sample into three parts equally. I do not follow it due to the constraint of degrees of freedom. Second, choosing an overly long parameter selection period will obtain larger tunning parameters, which "over-simplifies" the models and generates very flat forecasts.

very sensitive to selection of the length of the estimation set and the steps of the rolling analysis. Different lengths of the windows or predicitve periods generate different optimal tunning parameter. For the sake of brevity, 100-folds CV and one step ahead RMSE tunning parameter selection algorithms are abbreviated as "CV" and "RMSE" respectively.

4.3 Forecast

In this section, I compute 20 steps conditional forecast ahead of each variable of the corresponding model. I demonstrate the procedures to compute the predictions of VAR(10) and VARMA(6,1). I omit the demonstration of the computation of the predictions based on regularised models because the derviation is exactly the same. Consider VAR(10) first.

$$
\begin{aligned}
Z_t &= B_1 Z_{t-1} + B_2 Z_{t-2} + B_3 Z_{t-3} + B_4 Z_{t-4} + \ldots + B_{10} Z_{t-10} + \eta_t \\
Z_t &= \sum_{i=1}^{10} B_i Z_{t-i} + \eta_t
\end{aligned}
$$

where Z_t is a vector containing all variables mentioned in section 3, B_i are the coefficient matrices and η_t is a vector of error term.

When I compute 1 step ahead forecast:

$$\hat{E}(Z_{t+1}|\Omega_t) = \sum_{i=1}^{10} B_i Z_{t-i+1}$$

where the $\hat{E}(Z_{t+1}|\Omega_t)$ is the one-step ahead optimal forecast that conditions on all information up to time t.

Now, I compute 1 step ahead forecast further:

$$\hat{E}(Z_{t+2}|\Omega_t) = B_1 \hat{Z}_{t+1} + \sum_{i=2}^{10} B_i Z_{t-i+2}$$

where the $\hat{E}(Z_{t+2}|\Omega_t)$ is the two-steps ahead optimal forecast that conditions on all information up to time t and $\hat{Z}_{t+1}$ is $\hat{E}(Z_{t+1}|\Omega_t)$.

Therefore, keep repeating this recursion

$$\hat{E}(Z_{t+20}|\Omega_t) = \sum_{i=1}^{10} B_i \hat{Z}_{t-i+20}$$

Therefore, the implementation of 20 steps ahead conditional forecast is just simply repeating the recrusion above up to the 20th step ahead. Forecasting with VARMA(6,1) is very similar to the example above. However, as the mean of forecast errors is zero, the moving-average component is withdrawn after the first few step forecasts ahead. Consider the VARMA(6,1).

$$Z_t = B_1 Z_{t-1} + B_2 Z_{t-2} + B_3 Z_{t-3} + \ldots + B_6 Z_{t-6} + \eta_t + \theta^* \eta_{t-1}$$
$$Z_t = \sum_{i=1}^{6} B_i Z_{t-i} + \eta_t + \theta^* \eta_{t-1}$$

where Z_t is a vector containing all variables mentioned in section 3, B_i are the coefficient matrices of AR component, θ^* is the coefficient matrix of MA component and η_t is a vector of error term.

When I compute 1 step ahead forecast:

$$\hat{E}(Z_{t+1}|\Omega_t) = \sum_{i=1}^{6} B_i Z_{t-i+1} + \theta^* \eta_t$$

where the $\hat{E}(Z_{t+1}|\Omega_t)$ is the one-step ahead optimal forecast that conditions on all information up to time t.

Now, I compute 1 step ahead forecast further:

$$\hat{E}(Z_{t+2}|\Omega_t) = B_1 \hat{Z}_{t+1} + \sum_{i=2}^{6} B_i Z_{t-i+2}$$

where the $\hat{E}(Z_{t+2}|\Omega_t)$ is the two-steps ahead optimal forecast that conditions on all information up to time t and $\hat{Z}_{t+1}$ is $\hat{E}(Z_{t+1}|\Omega_t)$. The MA component is withdrawn in the model because η_{t+1} is outside the Ω_t. Hence, this is a VAR(6) process from now on.

Therefore, keep repeating this recursion

$$\hat{E}(Z_{t+20}|\Omega_t) = \sum_{i=1}^{6} B_i \hat{Z}_{t-i+20}$$

This completes the 20 steps ahead out-of-sample forecasts.

Chapter 5
Results

This section can be divided into two parts. In the first part, I show how LASSO provides sparse solution in the approximation models. In the second part, using the regularised and unregularised models, I practically compute 20-steps out-of-sample forecasts. I also highlight and discuss a few interesting observations in the results.

5.1 Regularisation of VARMA approximations by LASSO

Table A.1 and A.2 in appendix show the optimal tunning parameters for LASSO-VAR(10) by both CV and RMSE algorithms respectively. Surprisingly, both sets of the optimal tunning parameters are fairly different. Using the one step ahead RMSE method provides smaller values of the optimal tunning parameters compared to using 100-folds cross-validation. Table A.3 and A.4 show the optimal tunning parameters for LASSO-VARMA(6,1)-VAR(10) computed by both algorithms. Similarly to LASSO-VAR(10), on average, using RMSE method also provides smaller values of tunning parameters compared to CV's.

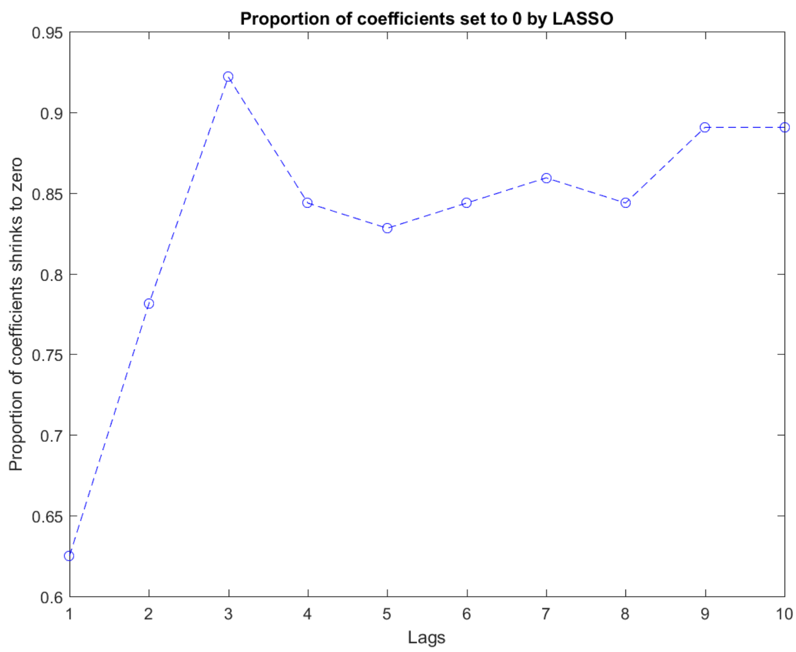

Figure 5.1: LASSO shrinkage (LASSO-VAR(10)) by using 100-folds CV

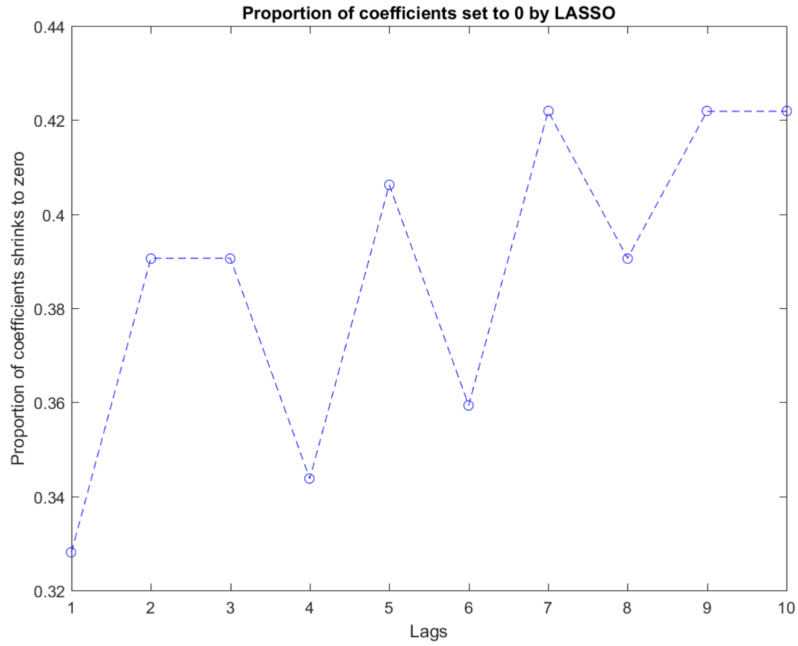

Figure 5.2: LASSO shrinkage (LASSO-VAR(10)) by using one step ahead RMSE method

Applying LASSO directly to regularise VARMA(6,1) requires plenty of computation effort because the innovations in the MA component are unobservable. One possible way to do this is to use the Kalman filter to construct the penalised likelihood expression with respect to VARMA(6,1), which is another way to implement LASSO in the Gaussian setting. However, since the likelihood of the non-linear part is not a quadratic function of the parameters and specifying the stationarity constraints of the coefficients in VARMA(6,1) is fairly complicated, optimising the penalised likelihood is computationally involved. Nevertheless, LASSO can still be applied on two approximation models, VAR(10) and VARMA(6,1)-VAR(10), which are linear model. Although the approximation simplifies the estimation process, more estimation errors are accumulated due to the enlarged parameter space. This situation leaves room for applying LASSO which automatically withdraws some uninformative variables by shrinking corresponding coefficients to zero.

In regard to the sparse solution of LASSO-VAR(10), Figure 5.1 and 5.2 show the shrinkage of coefficients by CV and RMSE method, that is the proportion of coefficients in the system reduced to zero corresponding to each specific lag. For instance, the first dot point on Figure 5.1 indicates that around 63% of the coefficients in the coefficient matrix corresponding to the first lag variable are reduced to zero.

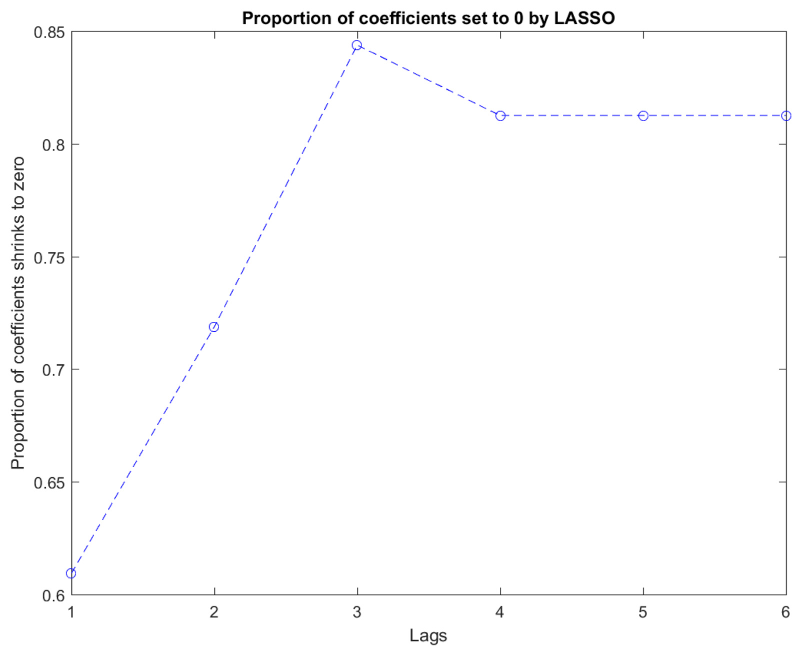

Figure 5.3: LASSO shrinkage (LASSO-VARMA(6,1)-VAR(10)) by using 100-folds CV method

In Figure 5.1, at the third lag, around 90% of the coefficients in the matrix are reduced to zero and the shrinkage flutuates around 85% thereafter. I define that the steady-state shrinkage is 85% after the cut-off (the 3rd lag). The cutoff can also be interpreted in economic terms. Mostly, agent use the past information with one year period. Importantly, the economic interpretation supports my initial assumption in the state-space model that the adaptive learning agents use small finite lag length to do forecast. However, from Figure 5.2, it seems that less coefficients are reduced to zero. The highest shrinkage, at 9th lag, is still less 50% and there is no obvious "cut-off". This happens because the optimal tunning parameters computed by the RMSE algorithm are lower those computed by CV. Therefore, less coefficients are set to zero due to the lighter penalty.

Figure 5.3 and 5.4 show the sparse solution of LASSO-VARMA(6,1)-VAR(10) by using CV and RMSE for tunning parameter selection.[1] In Figure 5.3, at the third lag, around 85% of the coefficients in the matrix are set to zero and the shrinkage keeps steady at 80% thereafter. Although the shrinkages of coefficients are different, the cut-off is consistent with the cut-off in LASSO-VAR(10). Figure 5.4 shows that the maximum shrinkage is 50% but I cannot observe any obvious steady-state shrinkage.

Furthermore, it is also important to check whether LASSO can reduce the model

[1] For an easier interpretation, I do not report the MA part in both figures.

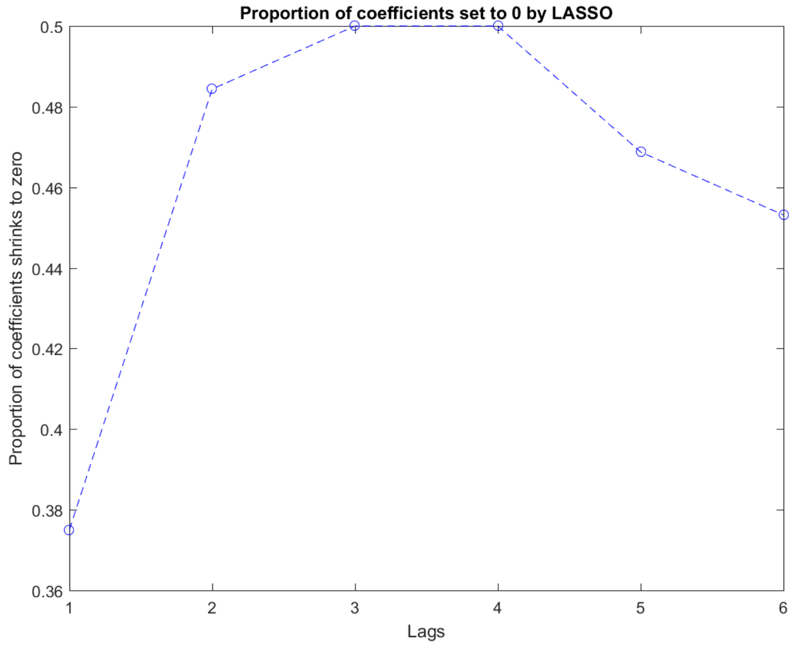

Figure 5.4: LASSO shrinkage (LASSO-VARMA(6,1)-VAR(10)) by using one step ahead RMSE method

Table 5.1: Parameter space

Model	Selection Algorithm	Number of non-zero coefficient
VARMA(6,1)	n.a.	392
LASSO-VAR(10)	100-folds CV	107
LASSO-VAR(10)	one step ahead RMSE	392
LASSO VARMA(6,1)-VAR(10)	100-folds CV	91
LASSO VARMA(6,1)-VAR(10)	one step ahead RMSE	210

space (set some coefficients to zero) in the approximations in order to tackle the problem of over-parameterisation. The reason of checking is to ensure that the model space of regularised models are at least equal or smaller than the orginal VARMA(6,1) model. All regularised models fulfill this prerequisite. Table 5.1 shows the number of non-zero coefficient in each system by using both tunning parameter selection algroithms. Coincidently, using RMSE to regularise VAR(10) generates the same number of non-zero coefficient as VARMA(6,1).[2] Using the same method to regularise VARMA(6,1)-VAR(10) is able to reduce more coefficients to zero, providing a smaller parameter space than VARMA(6,1)'s. Using 100-folds CV to regularise both approximations gives larger sparsity because the tunning parameters are generally higher. For LASSO-VAR(10), the sparsity is 3 times more than those computed by RMSE. Likewise, for LASSO-VARMA(6,1)-VAR(10), the

[2] The number of zero-coefficients in VARMA(6,1) is 0

sparsity is 2 times more than those computed by RMSE. Although imposing greater sparsity in models might improve out-of-sample forecast, in some cases, it might result in over-simplified model and hamper thr forecasts.

5.2 Forecast

I use the first 119 observations (before the lags adjustment) to fit the data and compute 20 steps ahead forecast (5 years). Then, I compare the predictive values with the last 20 observations (actual realisation). The reason of having a 20 steps ahead forecast is to show whether there is any significant discripency between long and short-term forecasts with different models.

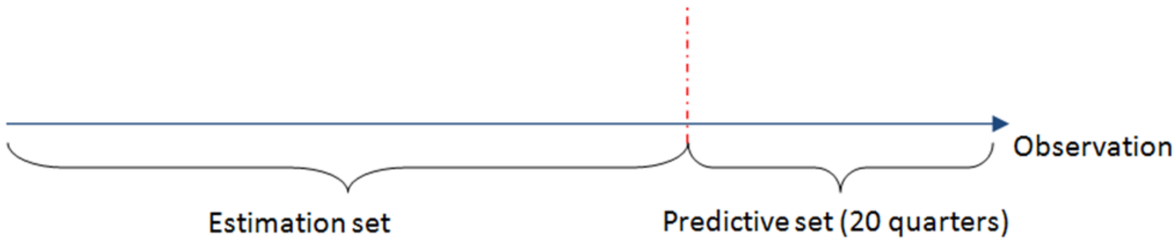

Figure 5.5: Illustration of estimation and predictive sets

Figure 5.5 describes the implementation of the out-of-sample (optimal) forecast graphically. To summarise, the estimation set contains observations to estimate the model and the observations in predictive set are used for assessing the forecast accuracy. All the predictions are shown in Figure B.1 to B.16 in apendix.

I make several remarks about these results. First, using CV in LASSO regularisation tends to generate less volatile forecasts. Figure B.1 and B.9 show the predictions of $\triangle Y_t$ in which the tunning parameters of the regularised models are computed by CV and RMSE. From Figure B.1, I observe that the both LASSO-approximations behave similarly. Both series show a small fluctuations at the begining, but quickly converage to stability thereafter. Based on Figure B.9, LASSO-VARMA(6,1)-VAR(10) behaves similarly as in the CV setting's but LASSO-VAR(10) is more volatile. This is because the tunning parameters computed by RMSE are relatively small, so less coefficients are reduced to zero and therefore the forecast is more sensitive.

Second, regularised models possess a greater capacity to incorporate non-stationary series.[3] Figure B.2 and B.4, both LASSO-predictions follow the decreasing trend of the actual series. Similarly, from Figure B.10 and B.12, both LASSO-predictions also demonstrate decreasing trends but in a more volatile manner.[4] In contrast, unregularised models' predictions of π_t and $\triangle C_t$ show significant deviations from

[3]In section 2, from the KPSS test, π_t and $\triangle C_t$ are potentially non-stationary.

[4]Using RMSE to select optimal tunning parameters tends to get smaller tunning parameters, so less coefficients are set to zero generating more volatile predictions.

the actual series.

Third, surprisingly, over-parameterised VAR(10) and VARMA(6,1) are good in incorporating any short term unexpected shocks. From Figure, B.1, B.3, B.6 and B.8, over-parameterised VAR(10) shows a big spike in each series at around 2009, but the other models do not react substantially to the Global Financial Crisis (GFC). Therefore, I can observe that VAR(10) is the most responsive model that successfully incorporates the big negative shock in the GFC. However, neither VAR(10) nor VARMA(6,1) do not exhibit predictive abilities for potential non-stationary series, namely π_t and $\triangle C_t$. Nevertheless, except for those two potential non-stationary series, the over-parameterised VAR(10) and VARMA(6,1) react substantially to the GFC effect on 5 out of 6 stationary economic series. Specifically, VARMA(6,1) is even more sensitive to short term shock, but it is also more likely to over-estimate the effect of the shock.[5] However, both VAR(10) ad VARMA(6,1) are only good for predicting short term potential shocks, but provide very inaccurate long term predictions.

Forth, using CV to do tunning parameter selection might give over-simplified model. From Figure B.5 and B.6, I can observe that the predictions computed by CV-LASSO regularisation are very close to zero throughout the whole forecasting horizon. Vice versa, from Figure B.13 and B.14, the predctions generated by RMSE-LASSO provide better forecasts which are closer to the true realisations.

These preliminarily results provide some insights about the forecasting performances of each model. Yet, it is important to implement statistical forecast evaluations in order to provide a more formal evidence of the predictive accuracy, which are demonstrated in the following chapters.

[5]Figure B.6 shows VARMA(6,1) "over-predicts" the GFC, compared to VAR(10).

Chapter 6
Forecast evaluation

In this chapter, I evaluate the forecast performances of different models based on the scaled RMSE and multivariate Diebold-Mariano tests.

6.1 Scaled RMSE criterion

6.1.1 Derivation of scaled RMSE

The most conventional way to assess models' predictive performances is to use root mean squared errors (RMSE) to compute the "distance" between the predictive and the actual series. It is possible that models perform differently in terms of different forecast horizons. Then, I divide the predictive set's into three blocks - Early, Mid and Late.[1] This separation provides more information rather than the overall RMSE. The RMSE for each block is computed as

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t,k}^2} \tag{6.1}$$

where T is the number of predictive values in a specific block, $e_{t,k}$ is the forecast error which is the difference between the true realisation of k variable $z_{t,k}$ and the corresponding predictive value $\hat{z}_{t,k}$.

Table 6.1 shows the sample standard deviations of the actual realisation in each block with respect to different series. The reason of checking the sample standard deviations is that the standard deviationse of the actual realisation might lead to unfair evaluation result. For example, higher standard deviations in the early but lower standard deviations in the later periods might provide high value of RMSE in the early block but low in the late block's. Figure B.5 provides an example that there is a huge unexpected shock in mid 2008 (early block), which is due to the global financial crisis (GFC), which drives the exchange rate significantly away from the unconditiona mean. From Table 6.1, we see that the early blocks' sample deviation

[1]As my total prediction is 20 periods, I put the first 7 predictive values into the early-block, second 7 predictive values into the mid-block and the last 6 values into the late-box.

Table 6.1: sample standard deviation

Variable	early	mid	late
$\triangle Y_t$	5.83E-03	5.71E-03	2.94E-03
π_t	6.63E-03	3.44E-03	5.13E-03
$\triangle U_t$	6.84E-02	2.71E-02	1.53E-02
$\triangle C_t$	5.39E-03	3.89E-03	3.98E-03
$\triangle EXR_t$	1.39E-01	4.13E-02	3.20E-02
$\triangle I_t$	7.43E-02	2.97E-02	2.95E-02
$\triangle E_t$	1.37E-01	6.38E-02	4.63E-02
$\triangle R_t$	2.82E-01	2.84E-02	7.70E-02

of $\triangle Y_t$, π_t, $\triangle U_t$, $\triangle EXR_t$, $\triangle I_t$, $\triangle E_t$ and $\triangle R_t$ are substantially greater than the respective late block's. It may seem that the long-term forecast is better than the short-term forecast, which is counter-intuitive. Motivated Billah et al. (2006), I propose a new evaluation indicator in ratio form - scaled RMSE (SRMSE):

$$SRMSE = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t,k}^2}}{\sqrt{\hat{\sigma}_k^2}}$$

Or equivalently:

$$SRMSE = \frac{\sqrt{\sum_{t=1}^{T}(z_{t,k} - \hat{z}_{t,k})^2}}{\sqrt{\sum_{t=1}^{T}(z_{t,k} - \bar{z}_{t,k})^2}}$$

where $\hat{z}_{t,k}$ is the predictive value and $\bar{z}_{t,k}$ is the mean of the actual realisation in a particular block. The above ratio has another useful interpretation. It compares the performances of the forecasting methods relatively to the relatively simple, but not feasible local mean forecasts within the forecast blocks.

6.1.2 Results of Scaled RMSE

Table A.5 and A.6 summarise the SRMSE result. The lowest value in each block is masked in bold. In addition to forecast from the describted models, I add a very simple forecast based on the unconditional sample mean. First, unconditional mean provides the best short, mid and long-term forecasts of the $\triangle Y_t$. Both LASSO approximations do not perform significantly better than the over-parameterised models - VAR(10). I attribute this to the fact that the variation and heterokedasticity of $\triangle Y_t$ are small, so using unconditional mean to forecast is better than using other models. Another noteworthy point is that both unregularised models - VARMA(6,1) provides the worst forecast regardless the predictive horizon.

Second, LASSO-VAR(10) provides the best forecasts in all forecasting horizons. Another LASSO approximation - LASSO-VARMA(6,1)-VAR(10) gives worse short-term and mid-term forecasts but achieve similar predictive accuracy in the long-run. Similarly for $\triangle Y_t$, the unregularised models provide worse forecasts, especially in the long-run. From chapter 2, the inflation rate is confirmed as stationary by the ADF test, but objected by the KPSS test, so it is potential non-stationary. This is a possible reason for why the unconditional mean does not provide good long term forecasts.

Again, the importance of sparsity is also shown in the forecasting result of $\triangle U_t$. Both LASSO approximations provide the best long-term forecasts, but the unregularised models give the best short-term prediction. Nevertheless, all the unregularised models give very low long-term forecast accuracy.

The forecasting performance of $\triangle C_t$ is similar to those in π_t. Both LASSO-approximations provide the best forecast regardless the forecast horizon. Other unregularised models provide worse forecasts regardless any predictive horizon. For instance, VARMA(6,1)-VAR(10) and VAR(10) give predictions with fairly low forecast accuracy regardless any forecasting horizon. Similarly for $\triangle EXR_t$, $\triangle I_t$, $\triangle E_t$ and $\triangle R_t$, the regularised models give the best long-term predictions, but the over-parameterised models give the best short-term predictions.

To summarise, first, LASSO approximations have a better flexibility to model slight non-stationary series, which is apparent in forecasting π_t and $\triangle C_t$. The LASSO forecasts are particularly well-performing in terms of long-term predictions. In contrast, over-parameterised models generate substantially worse predictions if the series are non-stationary, especially in the long-run. Second, estimating the MA component directly provides fairly inconsistent predictions, compared to other unregularised models. For example, both VARMA(6,1) and VAR(10) give similar predictive performances in the case of $\triangle I_t$. Yet, VARMA(6,1) gives fairly worse predictions than VAR(10) in the case of $\triangle Y_t$, especially in the long-run. I attribute this to the fact that estimating the MA component by non-linear technique requires high computational cost, which accumulates more estimation errors. Third, the over-parameterised models are particularly well-performing in predicting any short-term shocks. Except for non-stationary series, over-parameterised models react to the GFC effect on 5 out of 6 series in the VAR system.

6.2 Multivariate Diebold-Mariano tests

6.2.1 Derivation of the test

In this section, I evaluate the forecasting performance of each system jointly by multivariate Diebold-Mariano tests. Initially, Diebold and Mariano (1995) proposes a test (DM-test) to compare two univaraite models. This test computes the loss difference between two univaraiate models, given a particular loss function. Mariano and Preve (2012) extends the test to a multivariate version by which a system of time series predictions (e.g. predictions from VAR) can be evaluated jointly. Evaluating multivariate models' forecasting performances jointly can provide an understanding of the overall predictive performances of models. I assume the loss function is a squared deviation of the forecast from the realised value. Define the difference of squared error loss between the two forecasting models as:

$$d_{t,k} = (e^i_{t,k})^2 - (e^j_{t,k})^2 \quad i \neq j$$

where $e^i_{t,k}$ and $e^j_{t,k}$ are the forecast error of model i and j with respect to a paticular time series k in the system. $d_{t,k}$ is averaged by :

$$\bar{d_k} = \frac{1}{T}\sum_{t=1}^{T} d_{t,k}$$

where T is the total number of period in the predictive set and I can specify $\bar{\mathbf{d}}$ is a k x 1 vector containing all $\bar{d_k}$. The null hypothesis is defined as:

$$\bar{\mathbf{d}} = 0$$

with the test statistics

$$S = T\bar{\mathbf{d}}^{\intercal}\hat{\Omega}^{-1}\bar{\mathbf{d}} \longrightarrow \chi^2_k$$

$\bar{\mathbf{d}}^{\intercal}$ is the transpose matrix of $\bar{\mathbf{d}}$, $\hat{\Omega}$ is a consistent estimate of the variance-covariance matrix of $d_{t,k}$ and the degree of freedom of this setting is the number of time series k in the system. Mariano and Preve (2012) and Panchenko et al. (2010) suggest that using HAC estimate of Ω might provide correction of the time dependence structure without changing the distribution of the test statistics. Nevertheless, Giacomini (2006) provides Monte Carlo result which suggests that using sample variance still provide a reasonable estimator of Ω.

However, rejecting the null hypothesis can only indicate that the predictive abilities of the two models are significantly different. Checking which model is actually better requires examining the loss difference in $\bar{\mathbf{d}}$. If most of the elements $\bar{\mathbf{d}}$ are negative, I consider that model i is better than model j; otherwise model j is better than model i.

The comparisons are **(1)** LASSO-VAR(10) vs VAR(10), **(2)** LASSO-VARMA(6,1)-VAR(10) vs VARMA(6,1)-VAR(10), **(3)** LASSO- VAR(10) vs VARMA(6,1), **(4)** LASSO-VARMA(6,1)-VAR(10) vs VARMA(6,1) and **(5)** LASSO-VAR(10) vs LASSO-VARMA(6,1)-VAR(10).

The first two comparisons are designed for showing whether sparse solution of the regularised model is better than unregularised model. The second two comparisons are testing whether the regularised approximations have higher predictive accuracy than estimating VARMA(6,1) directly. The last comparison is to check which approximation is better. All results are summarised in Table 6.2 for LASSO results with different optimal tunning parameter algorithms.

6.2.2 Results of Multivariate Diebold-Mariano test

Table 6.2: Multivariate DM-test results

Algorithm	Comparison	Test statistics	Null rejection (5% sig. level)
CV	1	19.46	Yes
	2	28.09	Yes
	3	96.96	Yes
	4	94.39	Yes
	5	58.13	Yes
RMSE	1	46.73	Yes
	2	22.58	Yes
	3	111.79	Yes
	4	51.50	Yes
	5	18.54	Yes

The first stage is to test the null hypothesis at 5% significant level. I reject the null if the test statistics is greater than 15.507 (with degree of freedom = 8). From Table 6.2, all test statistics are greater than 15.507, so all the null hypothesis are rejected at 5% significant level. This result confirms four conjectures:

1. The loss differences between unregularised and regularised approximations are

statistically different from zero.

2. The loss differences between regularised approximations and the original VARMA model are statistically different from zero.

3. In terms of the regularised approximation models, the loss differences between LASSO-VAR(10) and LASSO-VARMA(6,1)-VAR(10) are statistically different from zero.

4. Different optimal tunning parameter selection does not give significant discrepancy in terms of the loss difference between LASSO and unregularised models.

In the second stage, I check the elements in $\bar{\mathbf{d}}$ for each comparison.[2] All the results are reported in Table 6.3. All negative values are masked in bold, implying model i has smaller loss than model j.

Table 6.3: Elements in $\bar{\mathbf{d}}$

		Model comparison				
Algorithm	**Variable**	**1**	**2**	**3**	**4**	**5**
CV	$\triangle Y_t$	**-1.60E-05**	**-8.35E-07**	**-7.12E-04**	**-7.08E-04**	**-3.71E-06**
	π_t	**-4.22E-05**	**-1.00E-05**	**-6.19E-05**	**-4.60E-05**	**-1.59E-05**
	$\triangle U_t$	1.01E-03	**-3.82E-04**	**-4.32E-04**	**-3.42E-04**	**-8.96E-05**
	$\triangle C_t$	**-1.01E-04**	**-1.51E-04**	**-6.98E-05**	**-6.75E-05**	**-2.39E-06**
	$\triangle EXR_t$	**-5.67E-04**	**-1.09E-03**	7.15E-04	8.02E-04	**-8.71E-05**
	$\triangle I_t$	**-1.19E-03**	**-5.89E-04**	**-4.11E-03**	**-4.11E-03**	7.75E-06
	$\triangle E_t$	**-5.52E-04**	**-3.29E-04**	**-1.59E-03**	**-1.60E-03**	8.74E-06
	$\triangle R_t$	**-1.83E-02**	9.29E-05	**-3.24E-02**	**-3.33E-02**	8.74E-04
RMSE	$\triangle Y_t$	**-2.29E-05**	**-4.48E-06**	**-7.19E-04**	**-7.12E-04**	**-6.94E-06**
	π_t	**-5.52E-05**	**-1.32E-05**	**-7.48E-05**	**-4.91E-05**	**-2.57E-05**
	$\triangle U_t$	1.85E-03	**-2.04E-04**	4.11E-04	**-1.64E-04**	5.75E-04
	$\triangle C_t$	**-1.31E-04**	**-1.04E-04**	**-9.95E-05**	**-2.05E-05**	**-7.90E-05**
	$\triangle EXR_t$	5.59E-05	**-9.52E-05**	1.34E-03	1.79E-03	**-4.54E-04**
	$\triangle I_t$	7.67E-04	**-2.79E-04**	**-2.15E-03**	**-3.80E-03**	1.65E-03
	$\triangle E_t$	2.76E-03	**-2.76E-04**	1.72E-03	**-1.54E-03**	3.26E-03
	$\triangle R_t$	**-1.24E-02**	**-1.13E-03**	**-2.66E-02**	**-3.45E-02**	7.93E-03

For the first comparison, by using RMSE, LASSO-VAR(10) does not perform significantly better than the over-parameterised VAR(10). Nevertheless, by using

[2]Each element in $\bar{\mathbf{d}}$ is the difference of squared forecast error between two models corresponding to a particular series. The variable order is the same as described in chapter 2. For example, the first element is the difference of squared forecast corresponding to $\triangle Y_t$.

CV, the forecasting performance of LASSO approximation is shown to be better than the over-parameterised VAR(10). The results of the second comparison do not show significant discrepancy between using CV and RMSE algorithm. This indicates that LASSO-VARMA(6,1)-VAR(10) gives better forecasts than VARMA(6,1)-VAR(10) regardless using any algorithm. For the third and forth comparison, similarly to the results of the first two comparisons, regularised models give better foreasting performances than VARMA(6,1) regardless using any algorithm. By using the CV algorithm, the fifth comparison shows that LASSO-VAR(10) provides a smaller loss than the LASSO-VARMA(6,1)-VAR(10). I attribute this to the fact that LASSO-VARMA(6,1)-VAR(10) might involve more estimation errors since it requires the two-stage optimisation. However, neither LASSO-VARMA(6,1)-VAR(10) nor LASSO-VAR(10) shows significantly better than each other in terms of their forecasting performance while using the RMSE algorithm.

There are two things that align with the results in the previous sections. First, although over-parameterised models are well-performing in forecasting short-term shocks, the overall predictive performances are bad. Second, the regularised models always give higher accuracies than unregularised models in forecasting non-stationary series.[3]

[3]The squared loss difference of π_t and $\triangle C_t$ are always negative, which are shown in Table 6.3.

6.3 Rolling Window Forecast

In this section, I describe a rolling-window based analysis for forecast evaluation. Using rolling window to compute forecast is to ensure that each one-step ahead optimal forecast is based the same sample size.

6.3.1 Introduction of rolling window

Initialising the rolling-window analysis is similar as the fixed sample forecast. First, I use the observations in the estimation set to fit the data. Then, I do one step ahead forecast and record the squared forecast errors of each series. The window shifts one step forward and I repeat the procedures of estimation, prediction and squared forecast error computation again.[4]

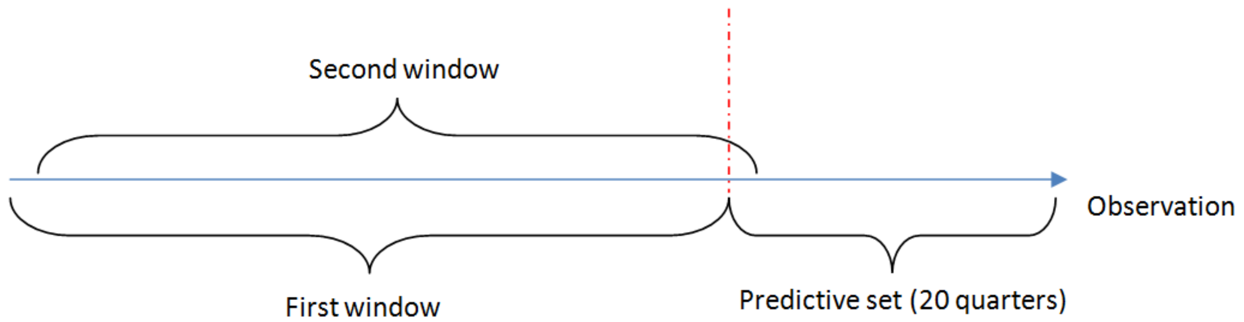

Figure 6.1: Illustration of rolling window analysis

Figure 6.1 provides a graphical ilustration of the rolling window analysis. The window keeps rolling by T-1 times where T is the total number of observation in the predictive set. I then compute the RMSE as:

$$RMSE_k = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(z_{t,k} - \hat{z}_{t,k})^2}$$

where t = 1, . . ., T and $RMSE_k$ is the root-mean squared error of with respect to a particular k-th time seires in the system.

6.3.2 Results of rolling window

Table A.7 and A.8 show the RMSE of each series computed by rolling window. For $\triangle Y_t$, both LASSO-VARMA(6,1)-VAR(10) models provide the best forecasting

[4]I drop the first observation in the estimation set but include the first observation in the predictive set. Hence, the sample size of the window does not change when the widnow is rolling.

performances. The RMSE of VARMA(6,1) is higher than all other regularised predictions'. Forecasting potential non-stationary series π_t and $\triangle C_t$ provide similar results in which regularised predictions give the best forecasting performance. Vice versa, directly estimating VARMA(6,1) provides the worst result in these cases. For $\triangle U_t$ and $\triangle E_t$, likewise, LASSO-VARMA(6,1)-VAR(10) provides the best forecasting performance. The second and third best performing models are both regularised predictions. For both $\triangle EXR_t$ and $\triangle I_t$, LASSO-VAR(10) give the best predictions. Likewise, unregularised VAR(10) and VARMA(6,1) give the worst forecasts. However, for $\triangle I_t$, unregularised VARMA(6,1)-VAR(10) model gives the best forecasting performance while LASSO- VARMA(6,1)-VAR(10) model gives the second best performance.

The rolling window results are summarised as follows. First, obviously, regularised models consistently give the highest forecast accuracy. According to the result, 7 out of 8 lowest RMSE predictions are computed by the regularised models. Second, over-parameterised and unregularised models consistently give predictions with low forecast accuracy. This aligns with the scaled RMSE and multivariate DM-test results. Third, estimating the VARMA(6,1) model directly gives fairly inaccurate predictions, which is worse than those computed by the VAR(10) model. According to the result, all predictions with the highest RMSE are generated by VARMA(6,1). This also aligns with the scaled RMSE results. Last, the result also suggests that using RMSE algorithm in LASSO models increase the forecast accuracy. 6 out of 7 LASSO predictions with the lowest RMSE are computed by the RMSE algorithm.

Chapter 7
Extension: Model averaging

In this chapter, I extend my prediction by using a simple model averaging (more detail below) technique. As shown in the previous results, over-parameterised models provide better short-term forecasts, which are particularly more sensitive to incorporating unexpected shock (e.g. GFC). In contrast, regularised models give better long-term predictions which are shown in the scaled RMSE results. Motivated by this reason, averaging the predictions of regularised and unregularised models might improve the overall forecast accuracy. Elliott and Timmermann (2004) suggest that combining the predictions by simple average with appropiate weights performs better in practice. Movitated by this reason, I simply average the predictions as follows:[1]

$$\hat{z}^{i,j}_{t,k} = \frac{\hat{z}^{i}_{t,k} + \hat{z}^{j}_{t,k}}{2}$$

where $\hat{z}^{i,j}_{t,k}$ is the mean of predictive values of model u and model j corresponding to variable k at time t. The entire procedure in this chapter is outlined as follow :

1. I compute the averaged predictions of **(1)** LASSO-VAR(10) and VAR(10) as well as **(2)**LASSO-VARMA(6,1)-VAR(10)and VARMA(6,1)-VAR(10). For the notation purpose, I denote $\mathbf{Z}^{*}$ and $\mathbf{Z}^{**}$ as the collection of averaged predictive values of **(1)** and **(2)** respectively.
2. Apply multivariate Deibold-Mariano tests again for the comparisons below:[2]
   - $\mathbf{Z}^{*}$ vs VAR(10)
   - $\mathbf{Z}^{*}$ vs LASSO-VAR(10)
   - $\mathbf{Z}^{**}$ vs VARMA(6,1)-VAR(10)
   - $\mathbf{Z}^{**}$ vs LASSO-VARMA(6,1)-VAR(10)
   - $\mathbf{Z}^{*}$ vs VARMA(6,1)

[1] I average the predictions of unregularised and regularised models because only these two different set of models provide very distinct features.

[2] The averaged predictions are not expected to give more accurate forecasts in either short or long-run because it is computed by two series with the opposite best forecasting features. Hence, evaluating the overall performance of the averaged predictions is more informative. Therefore, I choose the multivariate Deibold-Mariano tests rather than the scaled RMSE in this section.

- $\mathbf{Z}^{**}$ vs VARMA(6,1)
- $\mathbf{Z}^{**}$ vs $\mathbf{Z}^{*}$

3. Check whether both models in each comparison have the same loss based on the test statistics. I reject the null hypothesis at 5% level if the test statistics is greater than 15.507.
4. By using the same decision rule specified in chpater 6, I check the elements in $\bar{\mathbf{d}}$ in order to determine which model is better, given that the degree of freedom is k (number of series).

Table 7.1: Multivariate Diebold-Mariano test for average predictions

Algorithm	Model comparison	Test statistics	Null rejection (5% sig. level)
CV	1	72.838	Yes
	2	26.263	Yes
	3	44.873	Yes
	4	6.503	No
	5	136.565	Yes
	6	153.184	Yes
	7	38.752	Yes
RMSE	1	70.482	Yes
	2	24.695	Yes
	3	74.113	Yes
	4	16.752	Yes
	5	79.520	Yes
	6	59.302	Yes
	7	44.565	Yes

Before checking the multivariate Diebold-Mariano tests results, I can observe that the averaged predictions seem providing a better fit of the actual series. For example, Figure B.33 and B.35 show that the averaged predictions do not only "react" to the GFC shock but also give a better long-run predictions which are very closed to the actual realisations. The other averaged predictions possess similar results. All the averaged predictions are shown in Figure B.17 to B.48. Yet, using the multivariate Diebold-Mariano tests to compare models jointly peovide more comprehensive results with statistical evidence. The test results are shown in Table 7.1 below.

From Table 7.1, it shows that all models do not have the same loss except the third comparison of CV algorithm, given the 5% significant level. From Table A.9, by using CV, generally, the averaged predictions give better forecast accuracy than the regularised and unregularised models except LASSO-VARMA(6,1)-VAR(10).

Moreover, $\mathbf{Z}^{**}$ gives worse forecast than $\mathbf{Z}^{*}$. By using RMSE, likewise, the averaged predictions give better forecast accuracy than the regularised and unregularised models except LASSO-VARMA(6,1)-VAR(10). In contrast to using CV, $\mathbf{Z}^{*}$ gives better forecast than $\mathbf{Z}^{**}$.

There are few noteworthy points in Table A.9. First, averaged predictions computed by CV-LASSO improve the forecast accuracy. Excluding the seventh comparison, by using CV, $\bar{\mathbf{d}}$ contains more negative values than the RMSE's, which means that the "loss" corresponding to the averaged predictions is lower than for the regularised and unregularised models. Second, using the CV algorithm, averaged predictions consistently demonstrate better forecasting performance for predicting $\triangle EXR_t$, $\triangle E_t$. As I can see in Table A.9, all elements in the $\bar{\mathbf{d}}$ matrix corresponding to $\triangle EXR_t$ and $\triangle E_t$ are always negative. For π_t, $\triangle U_t$, $\triangle C_t$ and $\triangle I_t$, most of the corresponding elements in the $\bar{\mathbf{d}}$ matrix are negative, so I still recognise the averaged predictions can improve the forecast accuracy of these variables. However, averaged predictions of $\triangle R_t$ do not give any significant improvements.

Chapter 8
Conclusion and discussion

This thesis starts from motivating a reduced-form model incorporating agents' expectations in order to give a better understand of the process of agents' decision making. Then, I adopt LASSO regularisation method to undertake shrinkage and variable selection. Last, I compute 5 years ahead out-of-sample forecasts of each series and assess the forecast accuracies by several evaluation tests and criteria.

For in-sample fit, my numerical results show a clear "cut-off" at the 3rd lag when I use CV as the tunning parameter selection algorithm in the LASSO regularisation. This result is consistent with the the literatures that agents use past information to build their expectations about future economic outcomes because acquiring new information period-by-period is costly. For the out-of-sample forecasts and corresponding evaluations, my findings suggest that regularised models are particularly outstanding in long-term predictions. Surprisingly, they also suggest that over-parameterised models are reactive to a fairly short-term unexpectated shock, e.g. GFC. This result is particularly interesting since it is fairly uncommon in the literature. Last, as regularised and unregularised models possess distinct features in term of forecasting, I combine two predictions by doing simple model averaging. My findings suggest that averaged predictions of regularised and over-parametised models give more accurate predictions than simply using either one of them.

Furthermore, there is also a room to further extend my analysis. As shown in Table 6.1, although most of the series are confirmed to be stationary, heterokedasticity is present due to time-varying variance. Hence, a potential GARCH/ARCH effect might be present in the actual realisations. I suggest that a possible extension could be applying LASSO regularisation to GARCH/ARCH model.

Additionally, the most challenging task in this thesis is to select the optimal tunning parameter. Although the method I adopted is correct, the optimisation result heavily depends on the steps and length of the rolling window. Therefore, different settings in the rolling window length might provide different results. Although I adopt two tunning parameter selection algorithms, the limitations of these algorithms are worth considering. Using CV is computationally efficient but might give biased results if the time dependence is significantly high. Moreover, using RMSE requires to divide the estimation set into two parts, which is infeasible in a high dimensional model using low frequency small sample size macro-economic data. For small degrees of freedom, I overcome this difficulty by increasing the observations to fit the model but decreasing the predictive set for computing the RMSE of tunning parameter selection. However, this could potentially "under-estimate" the tunning parameter because shorter predictive period accumulates less predictive error. This problem suggests that using frequentist methodology might not be an optimal choice of tunning parameter selection. Park and Casella (2008) proposes Bayesian-LASSO in which the tunning parameter is not treated as a fixed number. Although the computational cost is higher by running the Markov Chain Monte-Carlo (MCMC), they show that the Bayesian-LASSO has a better shrinkage power. However, this approach is very subjective to the prior distribution of each parameters, which seriously affects the rate of converging to the true posterior estimates.

Last, some predictions are very flat and closed to zero (e.g. $\triangle R_t$). One of the reasons is that the effects from different explanatory variables on the dependent variable are different as some variables are scaled differently. Hence, using one tunning parameter in each equation to control the penalty could set some small but significant effects to zero if other large effects exist. One of the possible way to do extension is to use adpative-LASSO, which put different "weight" on the coefficients in the penalty function, proposed by Zou (2006). This approach takes various sizes of effects into account in order to avoid "over-simplified" results.

Appendix A
Tables

Table A.1: Value of tunning parameter λ of LASSO-VAR(10) by using 100-folds cross-validation

	Value of tunning parameter λ
$\triangle Y_t$	0.0011
π_t	0.0003
$\triangle U_t$	0.0054
$\triangle C_t$	0.0005
$\triangle EXR_t$	0.0113
$\triangle I_t$	0.0119
$\triangle E_t$	0.0059
$\triangle R_t$	0.0104

Table A.2: Value of tunning parameter λ of LASSO-VAR(10) by using one step ahead RMSE method

	Value of tunning parameter λ
$\triangle Y_t$	0.0014
π_t	0.0018
$\triangle U_t$	0.0002
$\triangle C_t$	0.0002
$\triangle EXR_t$	0.0023
$\triangle I_t$	0.0001
$\triangle E_t$	0.0001
$\triangle R_t$	0.0005

Table A.3: Value of tunning parameter λ of LASSO-VARMA(6,1) -VAR(10) by using 100-folds CV

	Value of tunning parameter
$\triangle Y_t$	0.0009
π_t	0.0002
$\triangle U_t$	0.0065
$\triangle C_t$	0.0008
$\triangle EXR_t$	0.0080
$\triangle I_t$	0.0109
$\triangle E_t$	0.0055
$\triangle R_t$	0.0072

Table A.4: Value of tunning parameter λ of LASSO-VARMA(6,1) -VAR(10) by using one step ahead RMSE method

	Value of tunning parameter
$\triangle Y_t$	0.0016
π_t	0.0017
$\triangle U_t$	0.0053
$\triangle C_t$	0.0001
$\triangle EXR_t$	0.0006
$\triangle I_t$	0.0001
$\triangle E_t$	0.0040
$\triangle R_t$	0.0001

Table A.5: Scaled RMSE

Variable	Model	Algorithm	***Early***	***Mid***	***Late***
$\triangle Y_t$	**VAR(10)**	**n.a.**	1.3726	1.8657	2.8767
	LASSO-VAR(10)	**rmse**	1.1702	1.5102	2.6648
	LASSO-VAR(10)	**CV**	1.3104	1.6387	2.5251
	VARMA(6,1) -VAR(10)	**n.a.**	1.9832	1.2598	1.6752
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	1.2284	1.6516	2.6770
	LASSO -VARMA(6,1) -VAR(10)	**CV**	1.3104	1.6387	2.5251
	VARMA(6,1)	**n.a.**	2.4426	4.0872	14.1175
	Unconditional mean	**n.a.**	**1.1619**	**0.9570**	**1.0465**
π_t	**VAR(10)**	**n.a.**	1.7780	1.7998	1.6392
	LASSO-VAR(10)	**rmse**	**1.0187**	**0.9110**	1.0423
	LASSO-VAR(10)	**CV**	1.3401	1.2290	**0.9466**
	VARMA(6,1) -VAR(10)	**n.a.**	1.2557	1.6792	1.9812
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	1.4747	1.6098	1.0987
	LASSO -VARMA(6,1) -VAR(10)	**CV**	1.6502	1.3322	0.9824
	VARMA(6,1)	**n.a.**	1.4126	3.2816	1.8713
	Unconditional mean	**n.a.**	1.0996	1.3465	1.5706
$\triangle U_t$	**VAR(10)**	**n.a.**	**0.5943**	1.4238	1.7872
	LASSO-VAR(10)	**rmse**	1.2831	1.2081	1.2099
	LASSO-VAR(10)	**CV**	1.0606	1.2791	**0.9813**
	VARMA(6,1) -VAR(10)	**n.a.**	1.2287	**0.8467**	1.4439
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	1.1071	1.4238	1.0009
	LASSO -VARMA(6,1) -VAR(10)	**CV**	1.0965	1.2150	1.0095
	VARMA(6,1)	**n.a.**	0.7985	1.9960	3.1515
	Unconditional mean	**n.a.**	1.1639	0.9489	1.2600
$\triangle C_t$	**VAR(10)**	**n.a.**	2.7835	2.1153	3.8284
	LASSO-VAR(10)	**rmse**	1.8287	**0.9925**	**0.6750**
	LASSO-VAR(10)	**CV**	1.5948	2.2795	1.9269
	VARMA(6,1) -VAR(10)	**n.a.**	3.3231	2.9022	3.7597
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	3.3012	1.0505	0.8905
	LASSO -VARMA(6,1) -VAR(10)	**CV**	**1.5528**	2.2480	2.1576
	VARMA(6,1)	**n.a.**	2.5724	3.4010	1.6144
	Unconditional mean	**n.a.**	2.1303	1.6598	2.3726

Table A.6: Scaled RMSE (cont.)

Variable	Model	Algorithm	*Early*	*Mid*	*Late*
$\triangle EXR_t$	**VAR(10)**	**n.a.**	0.8131	1.5337	2.1238
	LASSO-VAR(10)	**rmse**	0.9426	1.3101	0.9843
	LASSO-VAR(10)	**CV**	0.9262	1.0455	**0.9144**
	VARMA(6,1) -VAR(10)	**n.a.**	0.9825	1.2817	1.1044
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	0.9590	1.4311	1.0585
	LASSO -VARMA(6,1) -VAR(10)	**CV**	0.9330	**1.0441**	0.9147
	VARMA(6,1)	**n.a.**	**0.7445**	1.6517	1.3960
	Unconditional mean	**n.a.**	0.9259	1.0774	0.9163
$\triangle I_t$	**VAR(10)**	**n.a.**	0.9878	1.8378	1.4115
	LASSO-VAR(10)	**rmse**	1.2004	1.8746	1.1530
	LASSO-VAR(10)	**CV**	0.9192	1.1519	**0.9144**
	VARMA(6,1) -VAR(10)	**n.a.**	1.0346	0.9376	1.3872
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	1.0112	1.0340	0.9935
	LASSO -VARMA(6,1) -VAR(10)	**CV**	**0.9164**	1.1542	0.9154
	VARMA(6,1)	**n.a.**	1.3503	2.6737	1.5668
	Unconditional mean	**n.a.**	0.9985	**0.9318**	1.2166
$\triangle E_t$	**VAR(10)**	**n.a.**	**0.8529**	1.3220	1.5503
	LASSO-VAR(10)	**rmse**	1.1537	1.1788	1.1593
	LASSO-VAR(10)	**CV**	0.8962	1.2860	0.9909
	VARMA(6,1) -VAR(10)	**n.a.**	0.9096	1.1389	1.4225
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	0.8960	1.2986	0.9913
	LASSO -VARMA(6,1) -VAR(10)	**CV**	0.8926	1.2984	**0.9818**
	VARMA(6,1)	**n.a.**	0.8651	1.1921	2.1278
	Unconditional mean	**n.a.**	0.9380	**1.0971**	1.2213
$\triangle R_t$	**VAR(10)**	**n.a.**	1.3546	5.8660	3.4213
	LASSO-VAR(10)	**rmse**	1.2099	**5.4576**	**3.3546**
	LASSO-VAR(10)	**CV**	1.0718	5.9649	3.4633
	VARMA(6,1) -VAR(10)	**n.a.**	1.0797	5.7052	3.4152
	LASSO -VARMA(6,1) -VAR(10)	**rmse**	1.0269	6.2534	3.4250
	LASSO -VARMA(6,1) -VAR(10)	**CV**	1.0759	5.6653	3.4519
	VARMA(6,1)	**n.a.**	**0.9571**	9.3073	5.0771
	Unconditional mean	**n.a.**	0.9770	5.7929	3.4266

Table A.7: RMSE of rolling window

Variable	Algorithm	Model	***RMSE***
$\triangle Y_t$	**n.a.**	**VAR(10)**	1.45E-04
	rmse	**LASSO-VAR(10)**	6.77E-05
	CV	**LASSO-VAR(10)**	6.56E-05
	n.a.	**VARMA(6,1) -VAR(10)**	1.05E-04
	rmse	**LASSO -VARMA(6,1) -VAR(10)**	**6.02E-05**
	CV	**LASSO -VARMA(6,1) -VAR(10)**	6.46E-05
	n.a.	**VARMA(6,1)**	1.10E-02
	n.a.	**Unconditional mean**	1.31E-04
π_t	**n.a.**	**VAR(10)**	1.14E-04
	rmse	**LASSO-VAR(10)**	**1.79E-05**
	CV	**LASSO-VAR(10)**	4.38E-05
	n.a.	**VARMA(6,1) -VAR(10)**	5.28E-05
	rmse	**LASSO -VARMA(6,1) -VAR(10)**	1.90E-05
	CV	**LASSO -VARMA(6,1) -VAR(10)**	4.56E-05
	n.a.	**VARMA(6,1)**	7.26E-03
	n.a.	**Unconditional mean**	1.55E-04
$\triangle U_t$	**n.a.**	**VAR(10)**	3.36E-03
	rmse	**LASSO-VAR(10)**	8.60E-03
	CV	**LASSO-VAR(10)**	1.83E-03
	n.a.	**VARMA(6,1) -VAR(10)**	2.34E-03
	rmse	**LASSO -VARMA(6,1) -VAR(10)**	**1.75E-03**
	CV	**LASSO -VARMA(6,1) -VAR(10)**	1.89E-03
	CV	**VARMA(6,1)**	4.29E-02
	n.a.	**Unconditional mean**	2.23E-03
$\triangle C_t$	**n.a.**	**VAR(10)**	1.36E-04
	rmse	**LASSO-VAR(10)**	8.29E-05
	CV	**LASSO-VAR(10)**	**6.63E-05**
	n.a.	**VARMA(6,1) -VAR(10)**	1.26E-04
	rmse	**LASSO -VARMA(6,1) -VAR(10)**	1.14E-04
	CV	**LASSO -VARMA(6,1) -VAR(10)**	7.46E-05
	n.a.	**VARMA(6,1)**	1.10E-02
	n.a.	**Unconditional mean**	1.38E-04

Table A.8: RMSE of rolling window cont.

Variable	Algorithm	Model	*RMSE*
$\triangle EXR_t$	n.a.	VAR(10)	1.56E-02
	rmse	LASSO-VAR(10)	**6.27E-03**
	CV	LASSO-VAR(10)	7.39E-03
	n.a.	VARMA(6,1) -VAR(10)	7.30E-03
	rmse	LASSO -VARMA(6,1) -VAR(10)	6.68E-03
	CV	LASSO -VARMA(6,1) -VAR(10)	6.77E-03
	n.a.	VARMA(6,1)	8.62E-02
	n.a.	Unconditional mean	6.65E-03
$\triangle I_t$	n.a.	VAR(10)	1.16E-02
	rmse	LASSO-VAR(10)	8.14E-03
	CV	LASSO-VAR(10)	**1.55E-03**
	n.a.	VARMA(6,1) -VAR(10)	6.34E-03
	rmse	LASSO -VARMA(6,1) -VAR(10)	7.51E-03
	CV	LASSO -VARMA(6,1) -VAR(10)	1.69E-03
	n.a.	VARMA(6,1)	7.03E-02
	n.a.	Unconditional mean	2.26E-03
$\triangle E_t$	n.a.	VAR(10)	1.27E-02
	rmse	LASSO-VAR(10)	1.92E-02
	CV	LASSO-VAR(10)	6.75E-03
	n.a.	VARMA(6,1) -VAR(10)	8.39E-03
	rmse	LASSO -VARMA(6,1) -VAR(10)	**5.91E-03**
	CV	LASSO -VARMA(6,1) -VAR(10)	6.17E-03
	CV	VARMA(6,1)	8.69E-02
	n.a.	Unconditional mean	8.43E-03
$\triangle R_t$	n.a.	VAR(10)	8.55E-02
	rmse	LASSO-VAR(10)	1.18E-01
	CV	LASSO-VAR(10)	5.07E-02
	n.a.	VARMA(6,1) -VAR(10)	**4.05E-02**
	rmse	LASSO -VARMA(6,1) -VAR(10)	4.06E-02
	CV	LASSO -VARMA(6,1) -VAR(10)	5.12E-02
	n.a.	VARMA(6,1)	2.19E-01
	n.a.	Unconditional mean	5.88E-02

Table A.9: Elements in $\bar{\mathrm{d}}$ for average predictions

		Model comparison						
Algorithm	**Variable**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
CV	$\triangle Y_t$	**-5.74E-05**	**-4.14E-05**	6.74E-05	6.83E-05	**-7.54E-04**	**-6.40E-04**	1.13E-04
	π_t	**-3.25E-05**	9.70E-06	**-1.84E-05**	**-8.36E-06**	**-5.22E-05**	**-5.43E-05**	**-2.17E-06**
	$\triangle U_t$	6.45E-06	**-1.00E-03**	**-6.74E-04**	**-2.91E-04**	**-1.43E-03**	**-6.33E-04**	7.98E-04
	$\triangle C_t$	**-1.18E-04**	**-1.65E-05**	**-1.43E-04**	8.06E-06	**-8.63E-05**	**-5.94E-05**	2.69E-05
	$\triangle EXR_t$	**-6.03E-04**	**-3.57E-05**	**-1.99E-03**	**-9.05E-04**	**6.79E-04**	**-1.04E-04**	**-7.83E-04**
	$\triangle I_t$	**-1.44E-03**	**-2.55E-04**	**-3.78E-04**	2.11E-04	**-4.36E-03**	**-3.90E-03**	4.58E-04
	$\triangle E_t$	**-8.22E-04**	**-2.70E-04**	**-1.57E-03**	**-1.24E-03**	**-1.86E-03**	**-2.84E-03**	**-9.76E-04**
	$\triangle R_t$	**-1.35E-02**	4.80E-03	5.99E-03	5.90E-03	**-2.76E-02**	**-2.74E-02**	2.29E-04
RMSE	$\triangle Y_t$	**-5.78E-05**	**-3.49E-05**	6.83E-05	7.28E-05	**-7.54E-04**	**-6.39E-04**	1.15E-04
	π_t	**-3.51E-05**	2.02E-05	**-2.80E-05**	**-1.48E-05**	**-5.47E-05**	**-6.39E-05**	**-9.24E-06**
	$\triangle U_t$	1.44E-04	**-1.70E-03**	**-6.65E-04**	**-4.61E-04**	**-1.29E-03**	**-6.25E-04**	6.69E-04
	$\triangle C_t$	**-9.48E-05**	3.60E-05	**-1.26E-04**	**-2.22E-05**	**-6.35E-05**	**-4.26E-05**	2.09E-05
	$\triangle EXR_t$	**-3.55E-04**	**-4.11E-04**	**-1.62E-03**	**-1.52E-03**	9.27E-04	2.72E-04	**-6.55E-04**
	$\triangle I_t$	**-1.81E-05**	**-7.85E-04**	**-1.03E-04**	1.75E-04	**-2.94E-03**	**-3.63E-03**	**-6.91E-04**
	$\triangle E_t$	8.38E-04	**-1.92E-03**	**-1.65E-03**	**-1.37E-03**	**-1.99E-04**	**-2.91E-03**	**-2.72E-03**
	$\triangle R_t$	**-9.54E-03**	2.91E-03	5.56E-03	6.70E-03	**-2.36E-02**	**-2.78E-02**	**-4.14E-03**

Appendix B
Figures

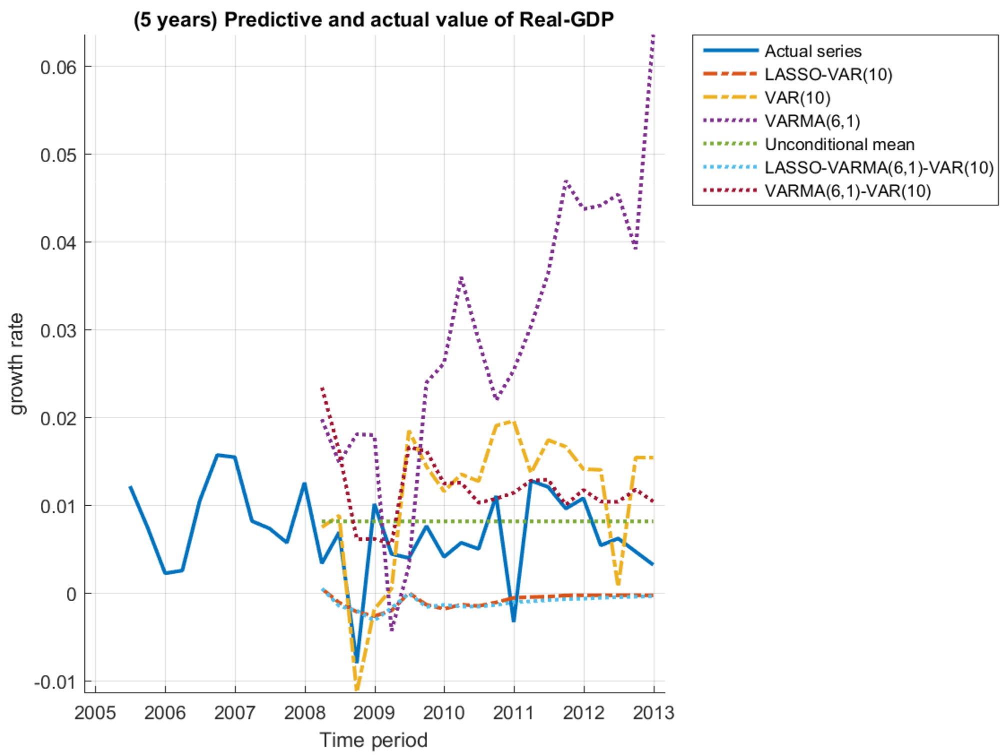

Figure B.1: (CV) 5 years ahead forecast of $\triangle Y_t$

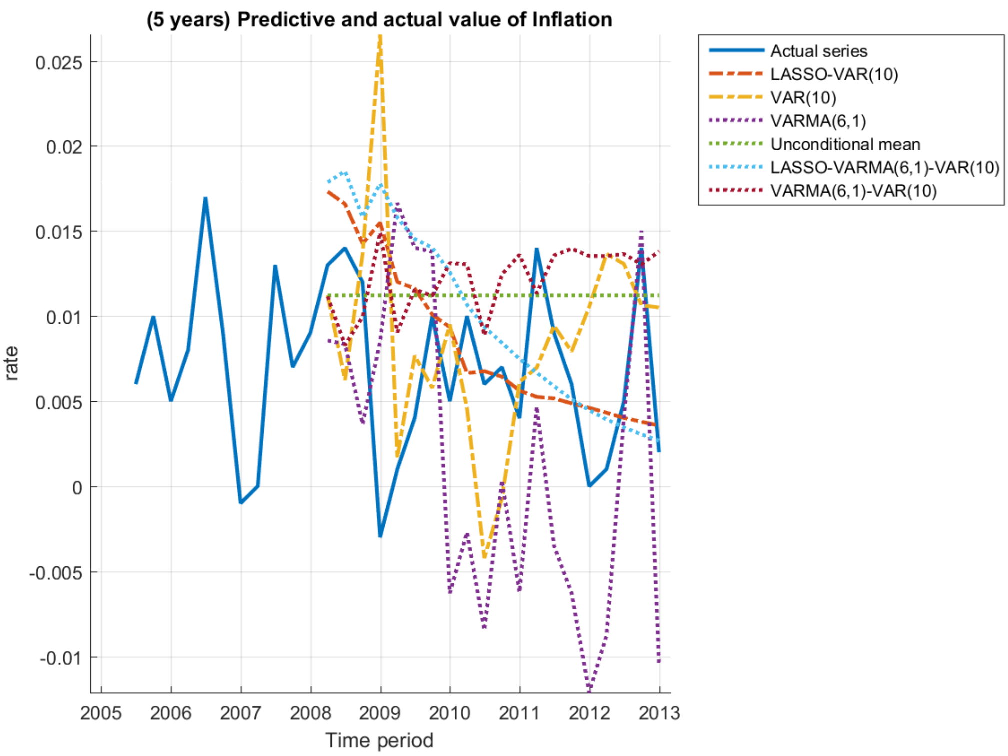

Figure B.2: (CV) 5 years ahead forecast of π_t

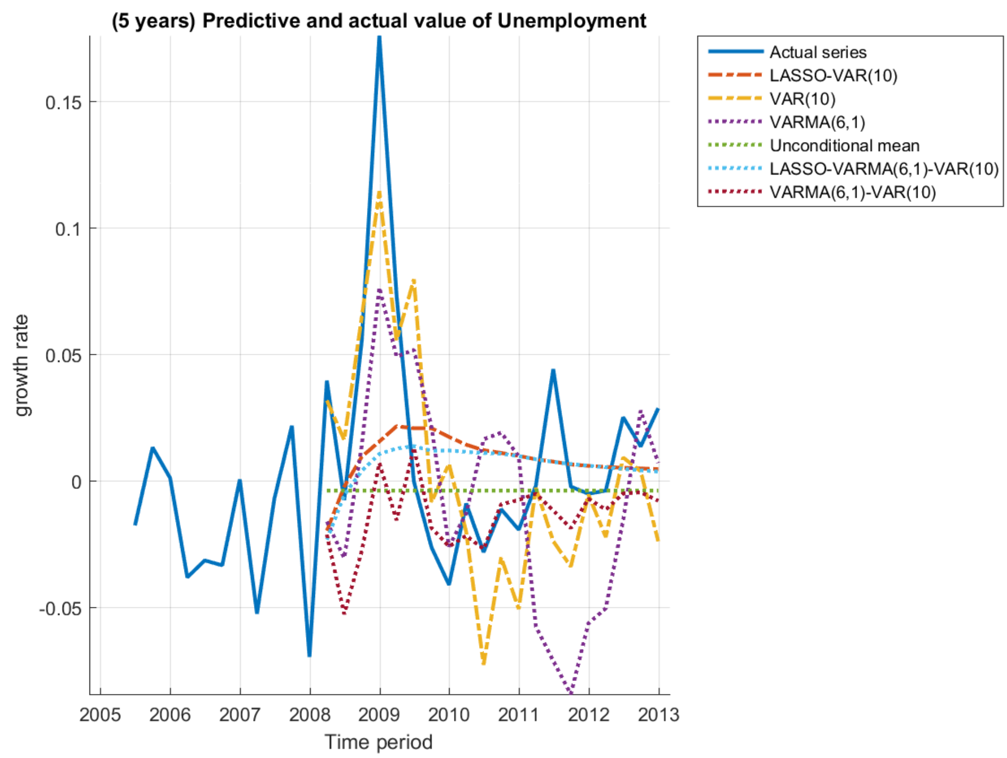

Figure B.3: (CV) 5 years ahead forecast of $\triangle U_t$

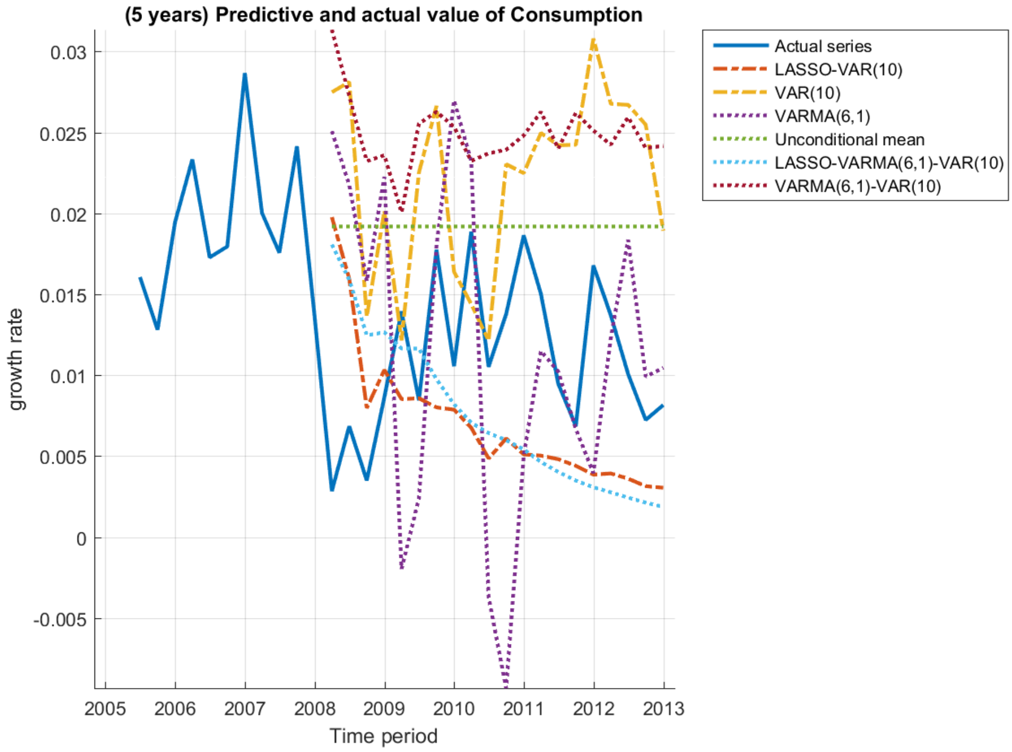

Figure B.4: (CV) 5 years ahead forecast of $\triangle C_t$

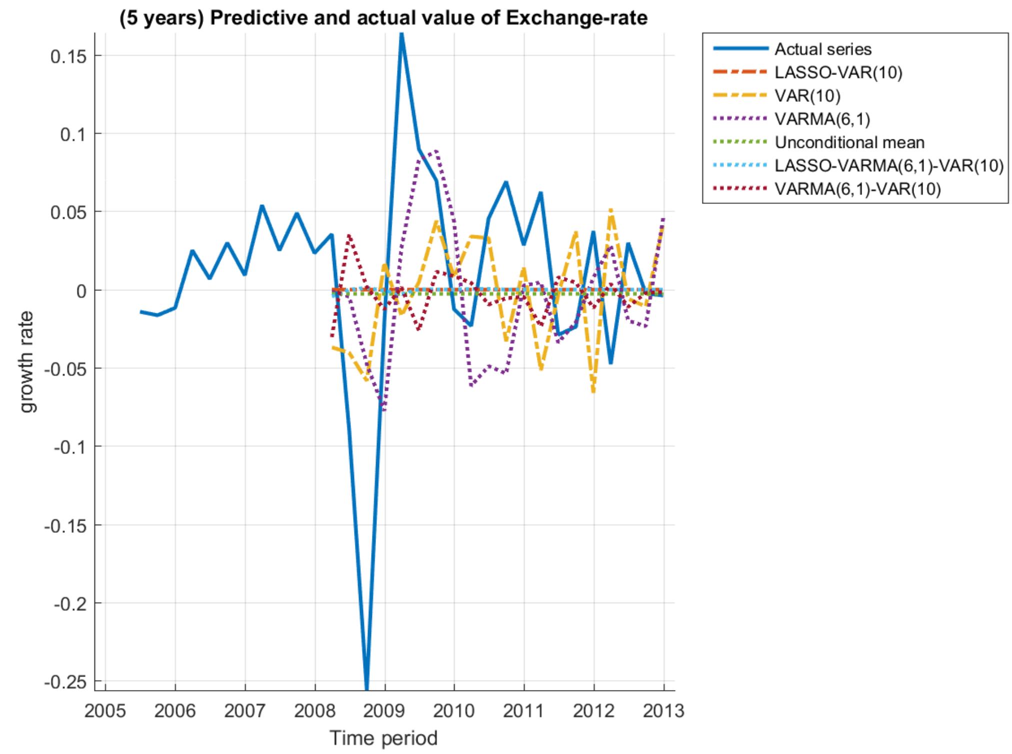

Figure B.5: (CV) 5 years ahead forecast of $\triangle EXR_t$

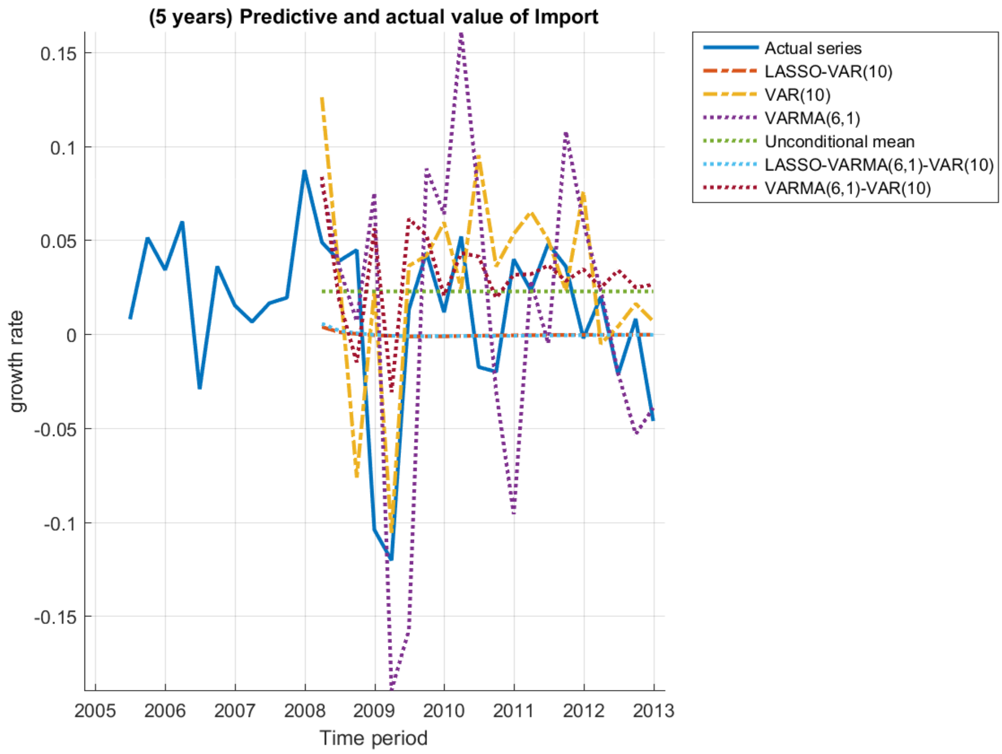

Figure B.6: (CV) 5 years ahead forecast of $\triangle I_t$

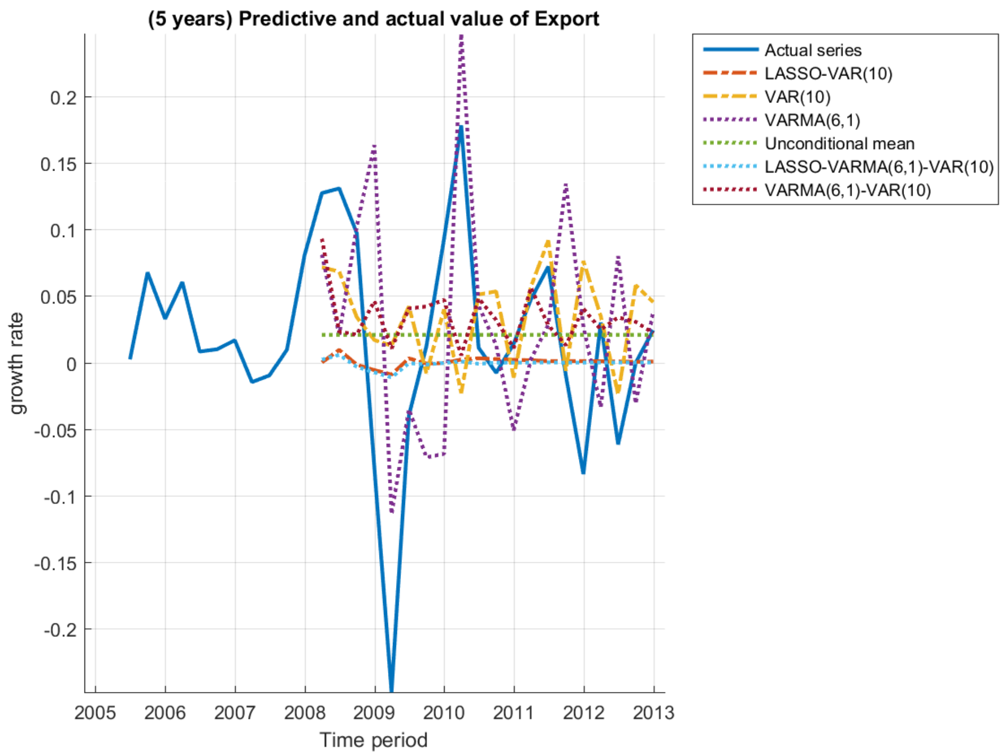

Figure B.7: (CV) 5 years ahead forecast of $\triangle E_t$

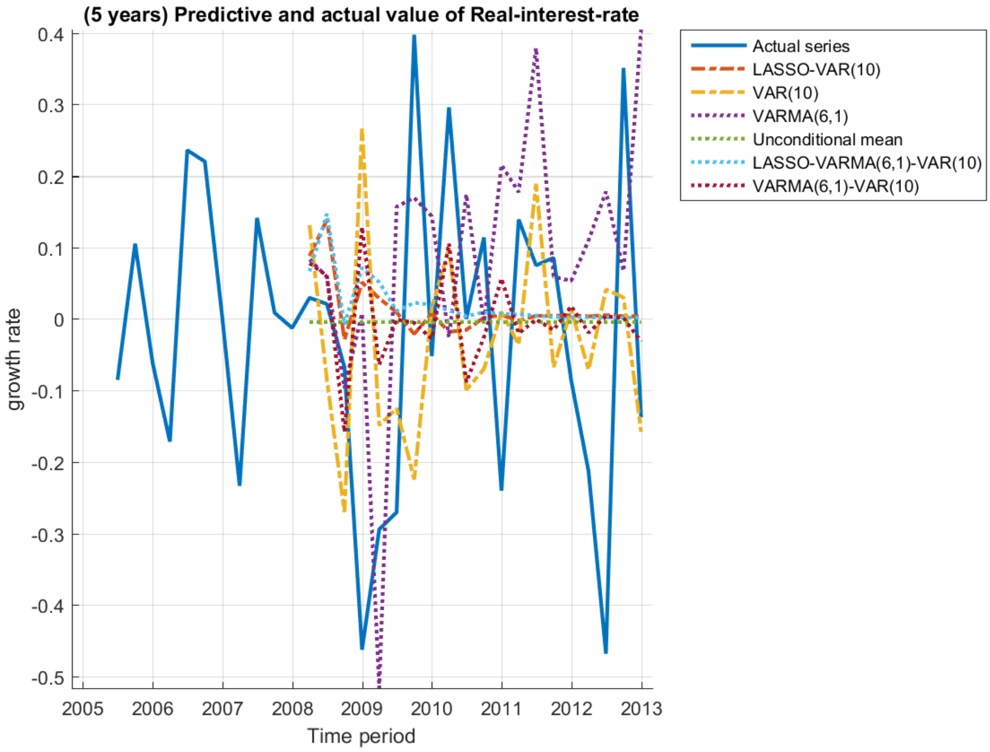

Figure B.8: (CV) 5 years ahead forecast of $\triangle R_t$

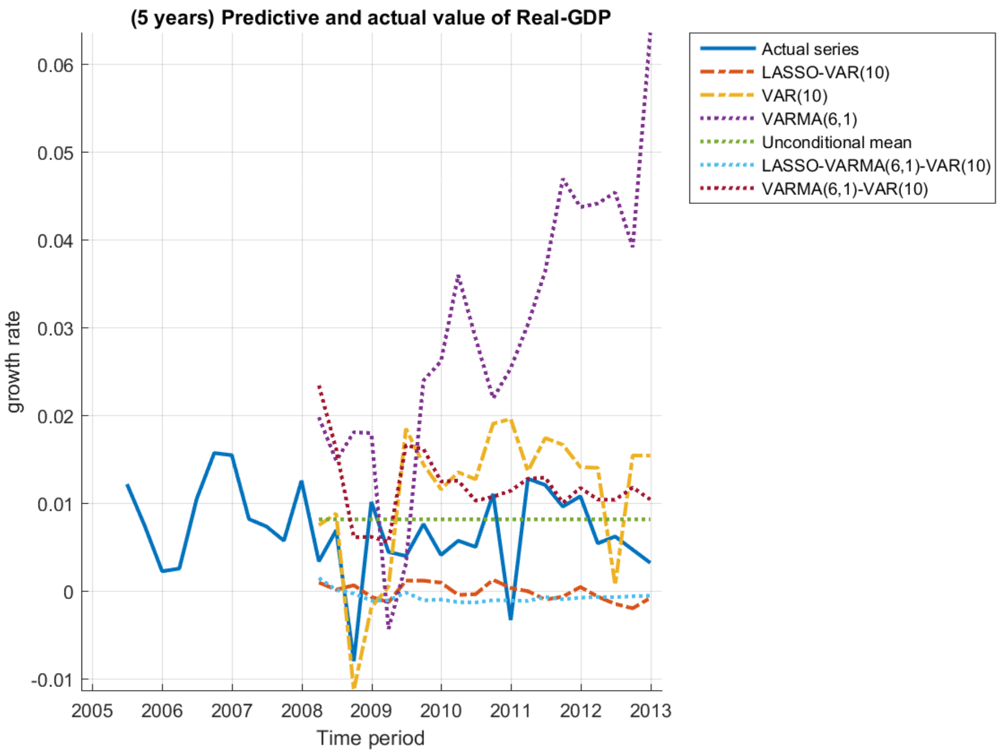

Figure B.9: (RMSE) 5 years ahead forecast of $\triangle Y_t$

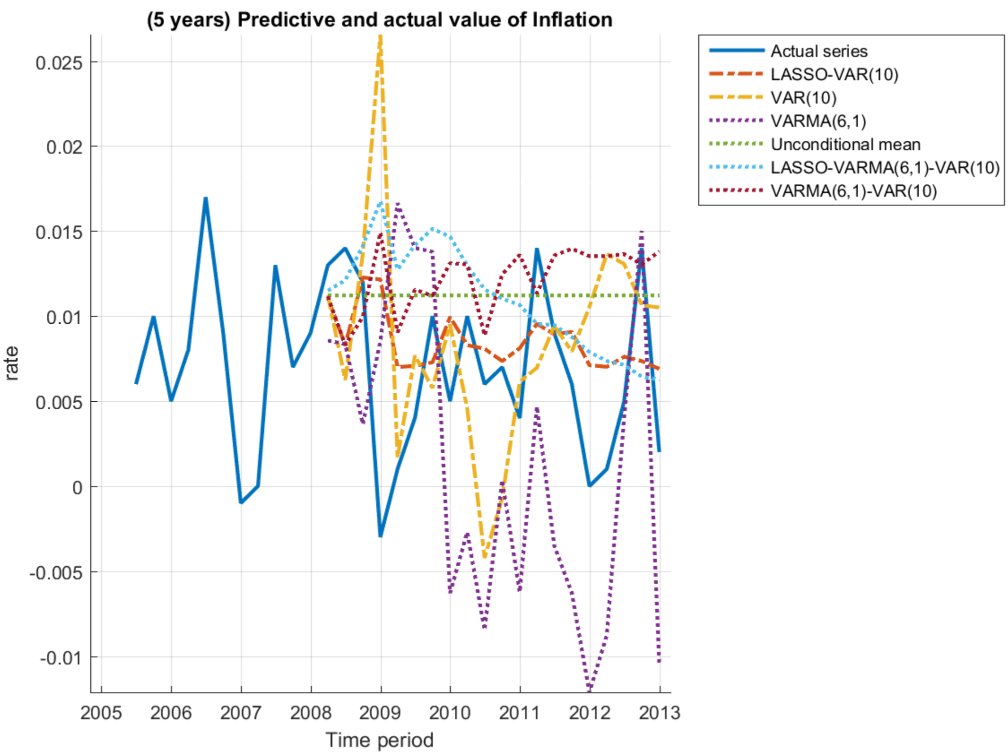

Figure B.10: (RMSE) 5 years ahead forecast of π_t

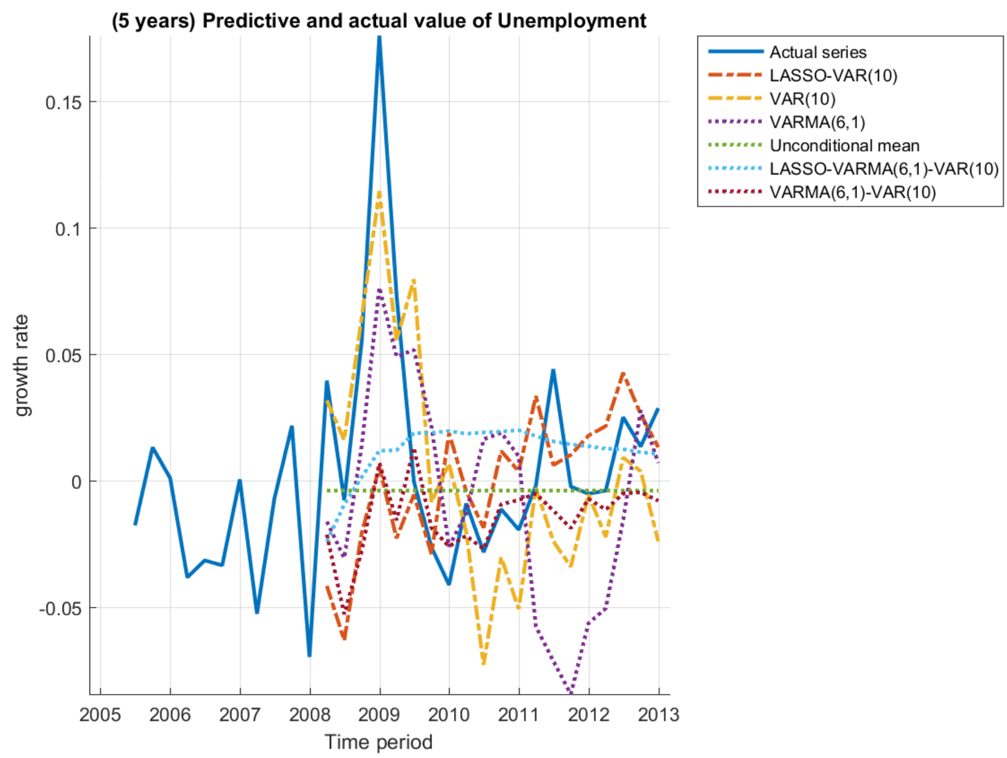

Figure B.11: (RMSE) 5 years ahead forecast of $\triangle U_t$

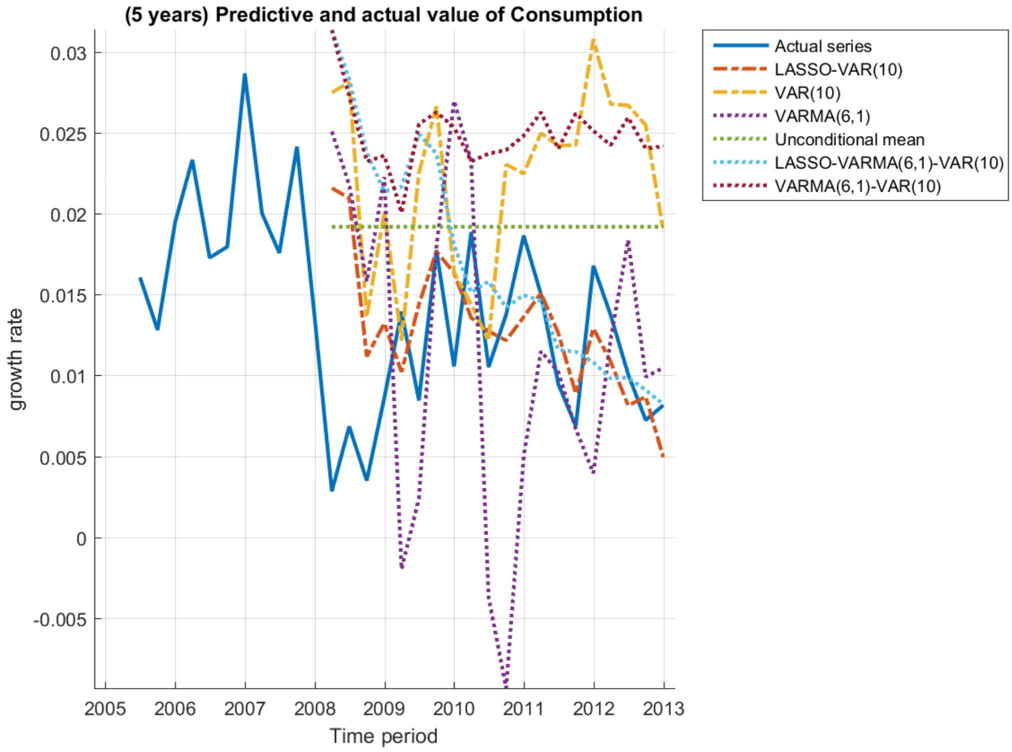

Figure B.12: (RMSE) 5 years ahead forecast of $\triangle C_t$

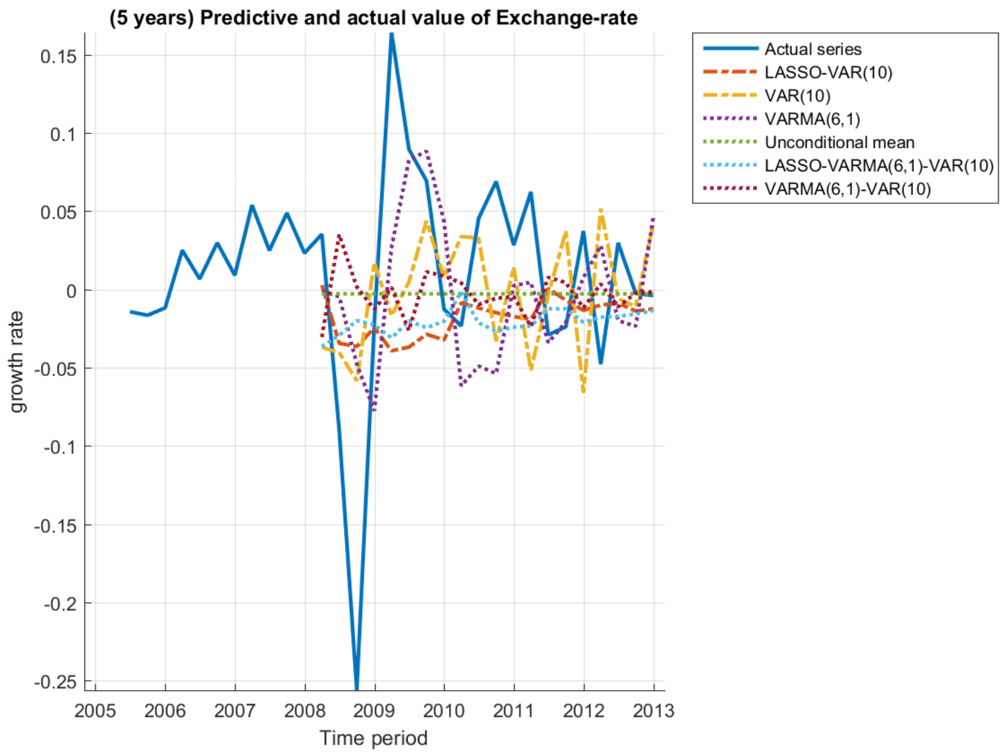

Figure B.13: (RMSE) 5 years ahead forecast of $\triangle EXR_t$

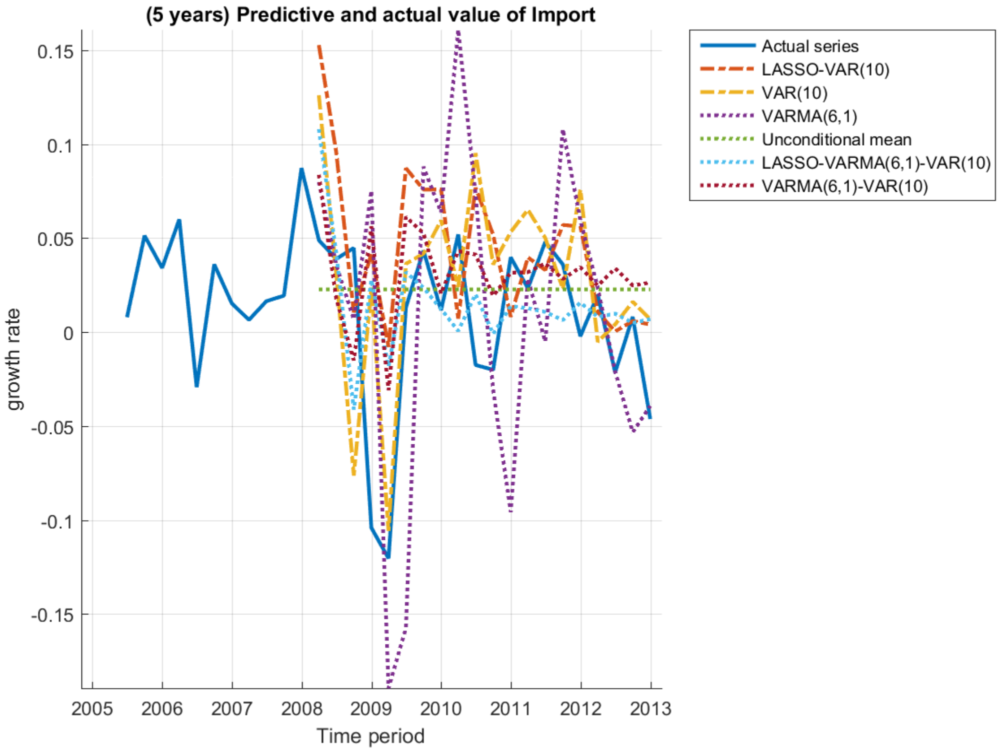

Figure B.14: (RMSE) 5 years ahead forecast of $\triangle I_t$

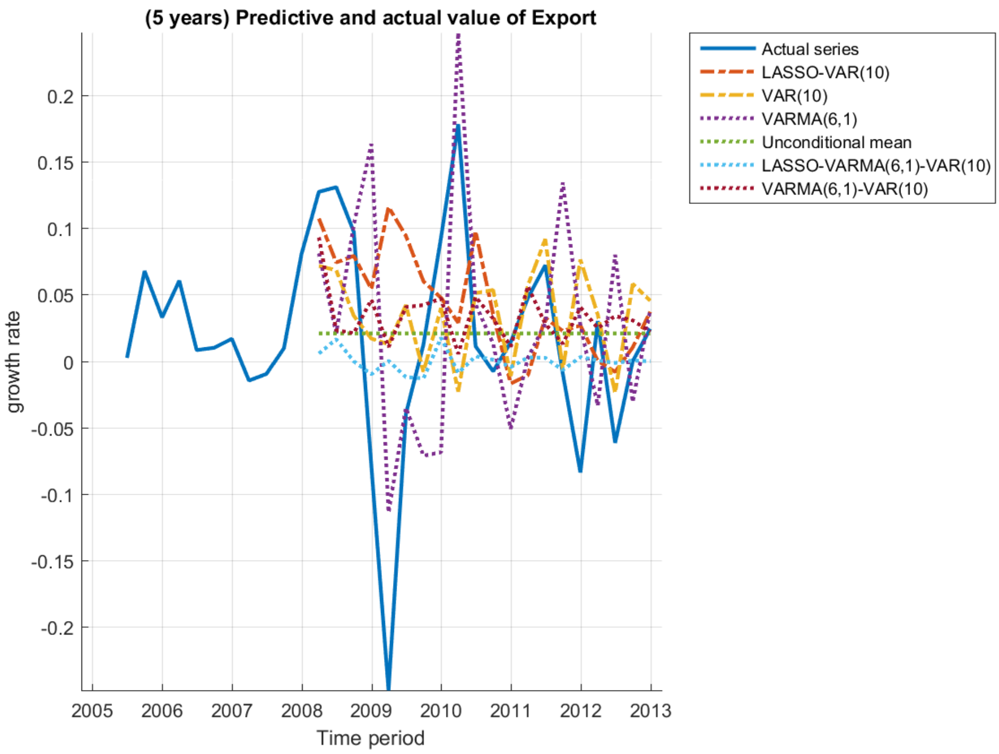

Figure B.15: (RMSE) 5 years ahead forecast of $\triangle E_t$

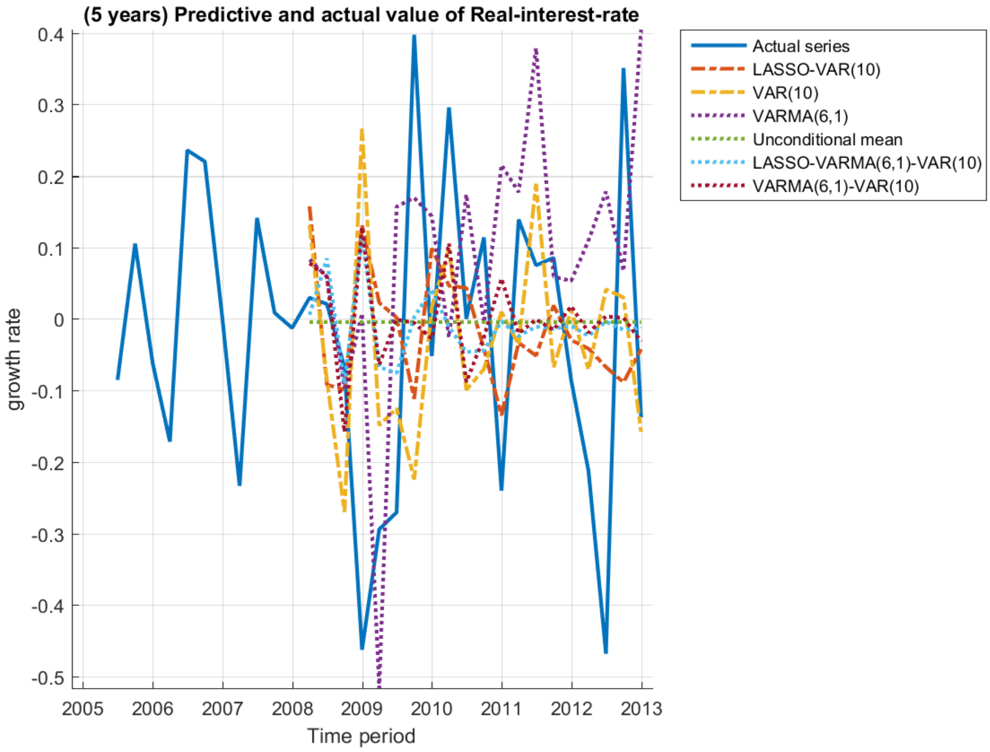

Figure B.16: (RMSE) 5 years ahead forecast of $\triangle I_t$

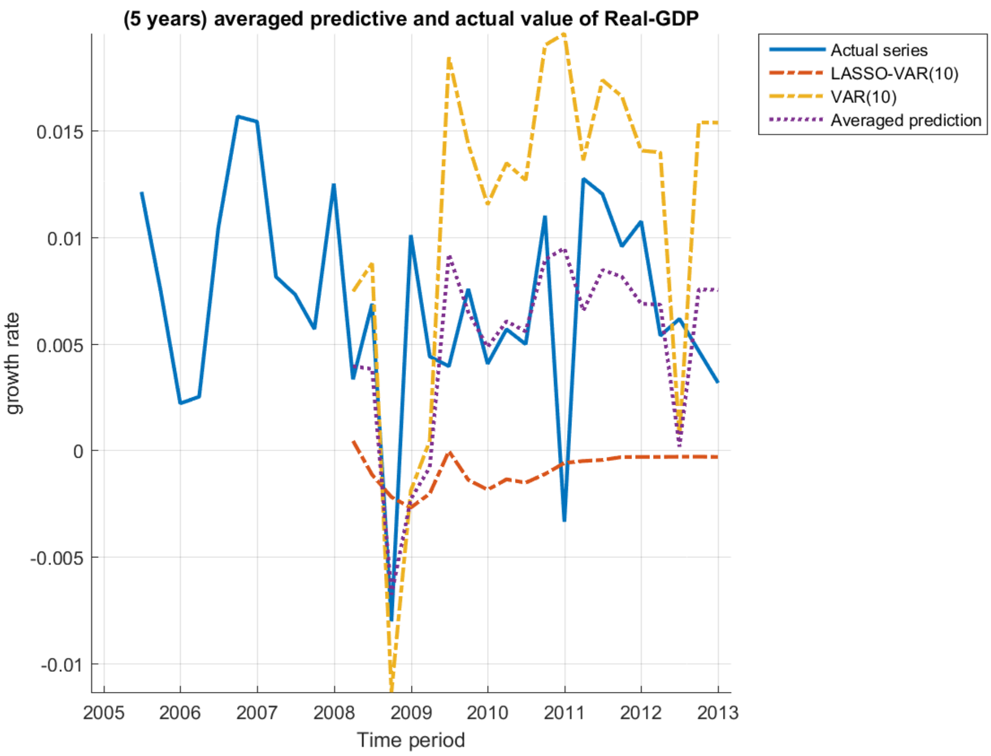

Figure B.17: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$

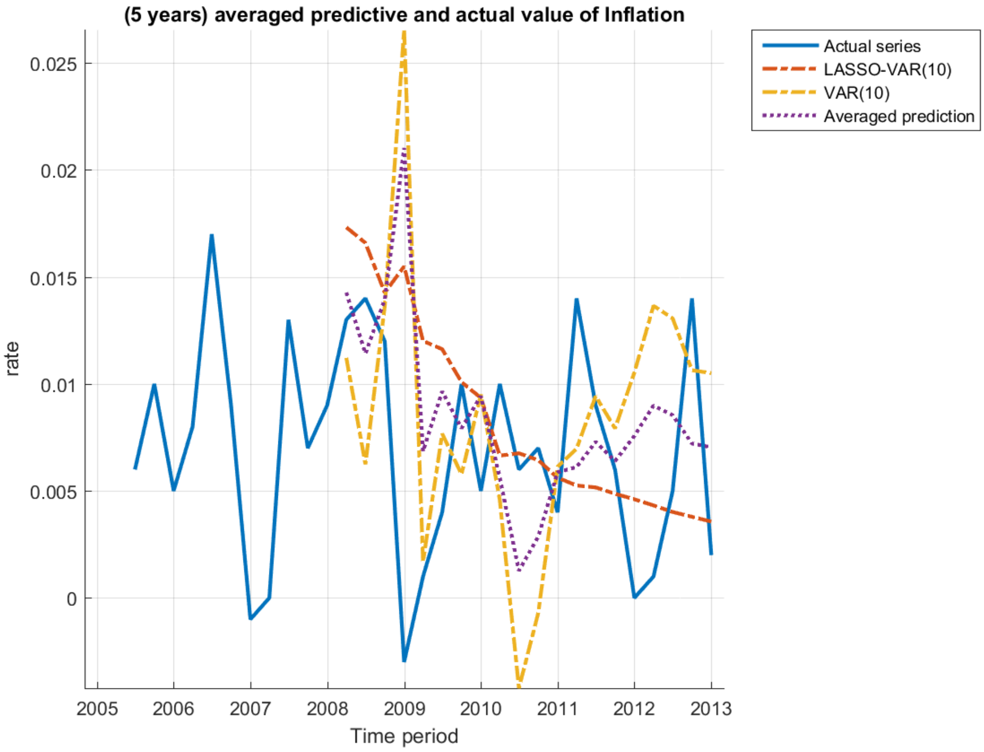

Figure B.18: CV- LASSO-VAR(10) 5 years ahead averaged forecast of π_t

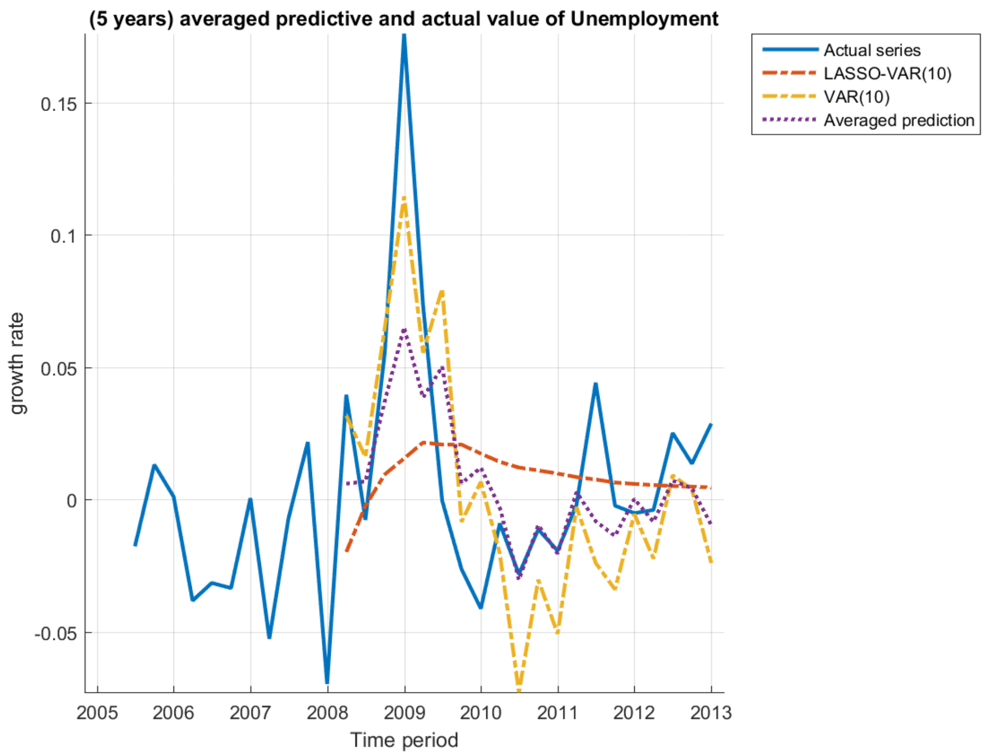

Figure B.19: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$

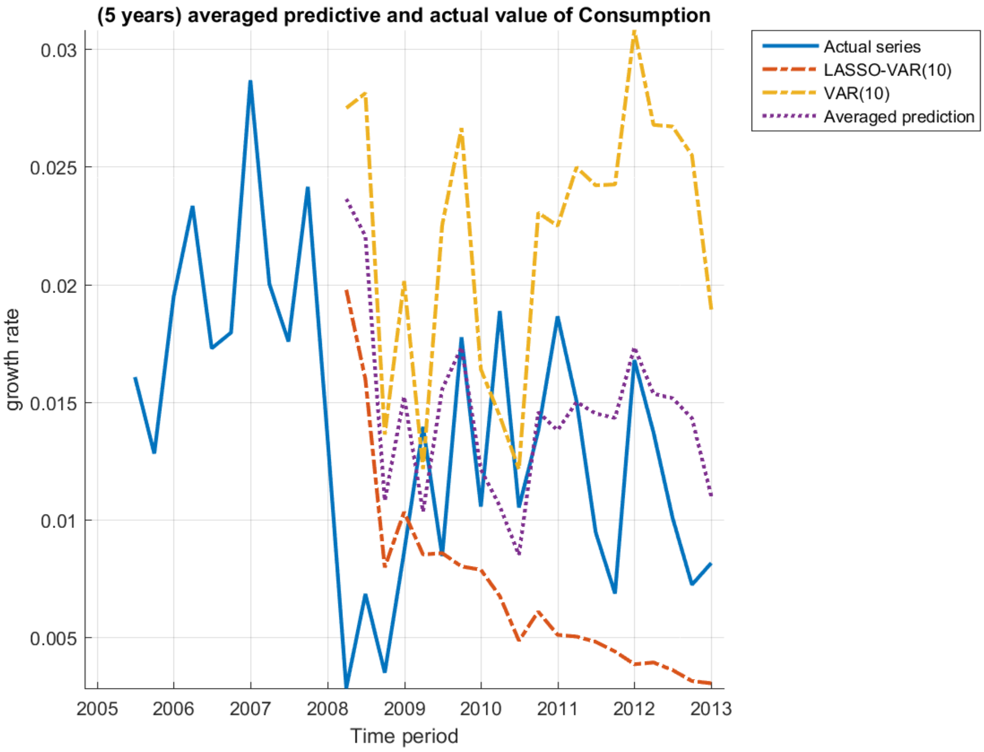

Figure B.20: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$

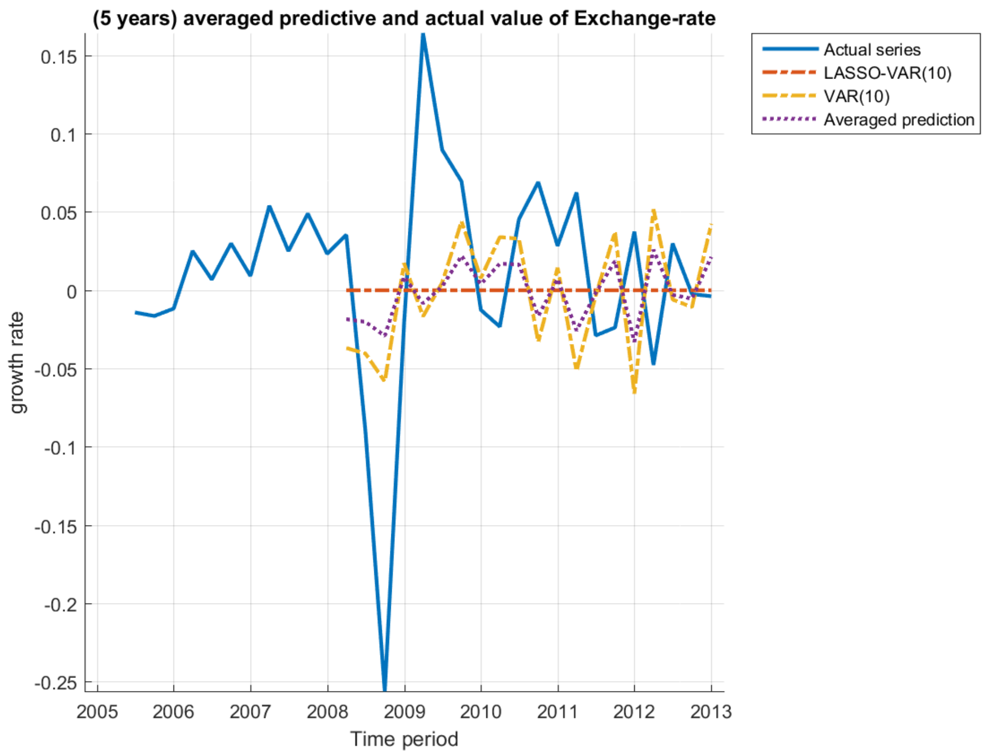

Figure B.21: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$

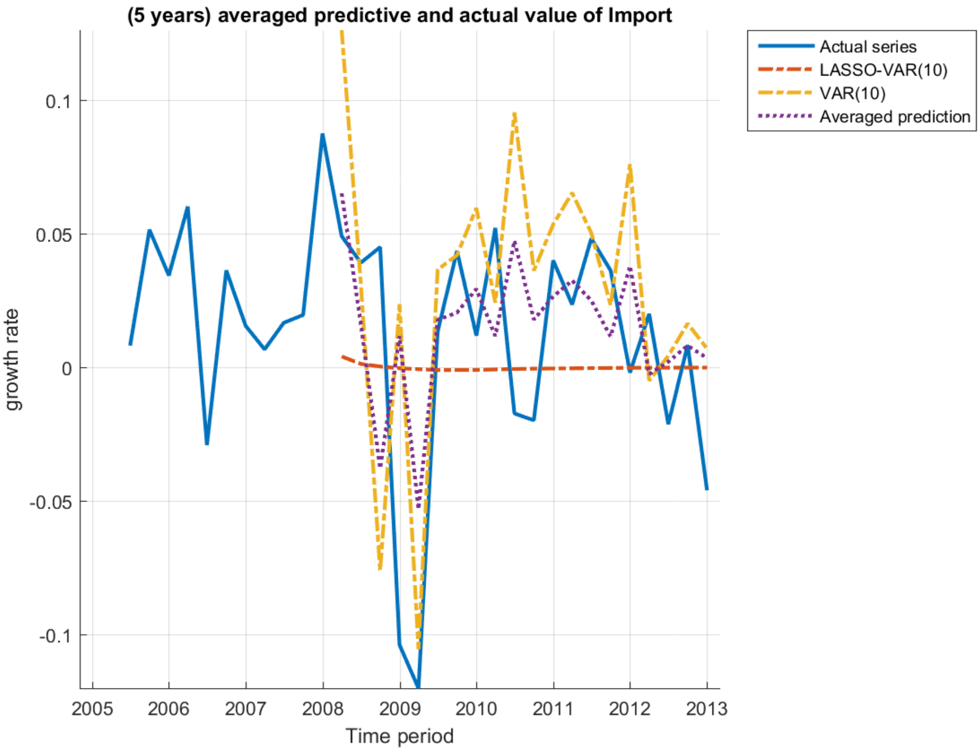

Figure B.22: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

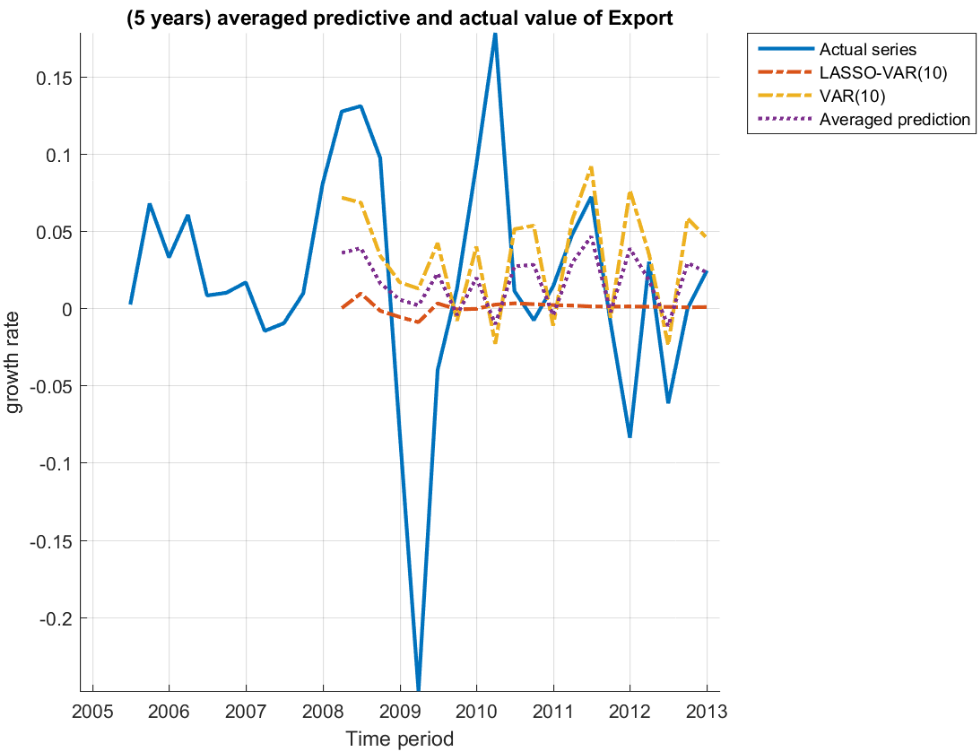

Figure B.23: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$

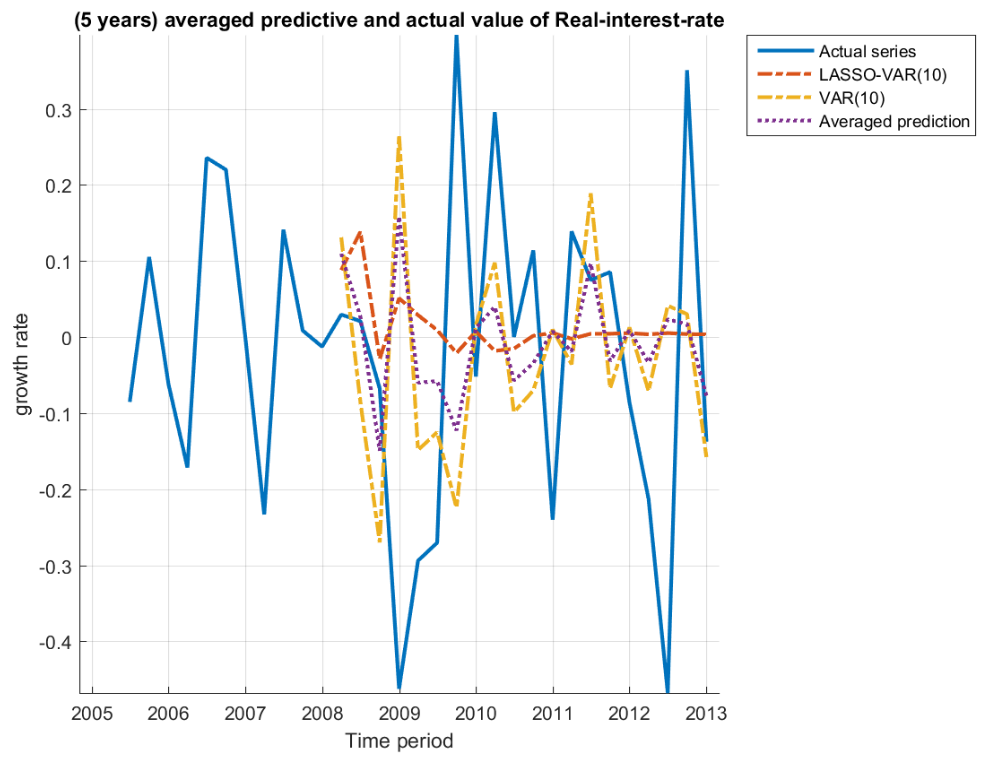

Figure B.24: CV- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

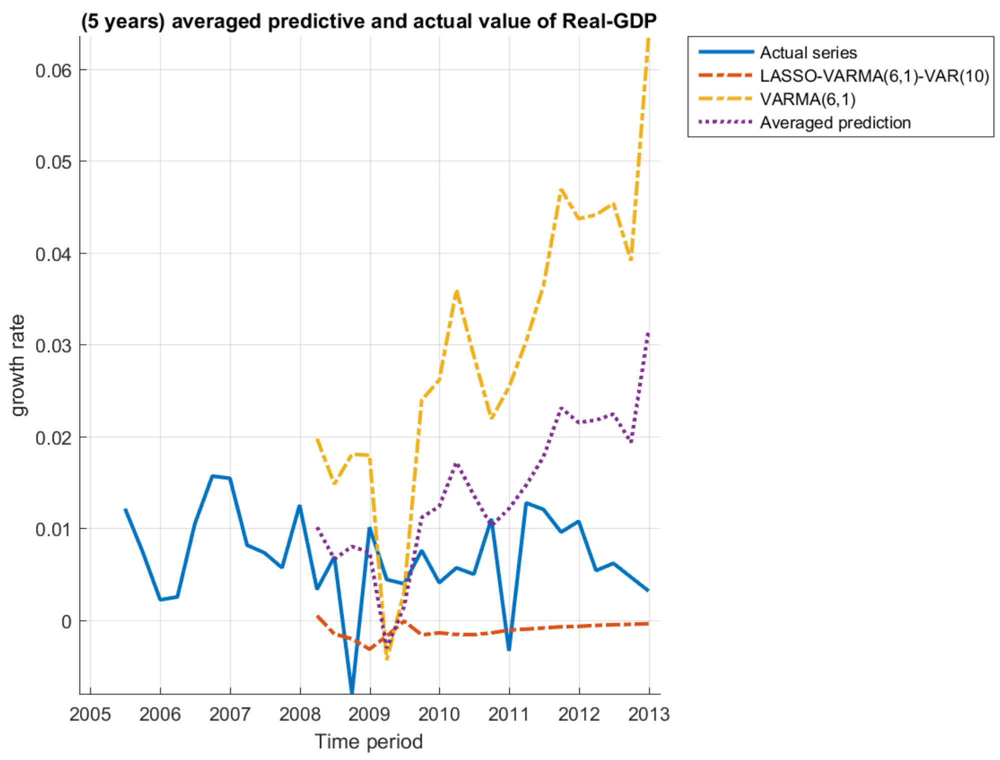

Figure B.25: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$

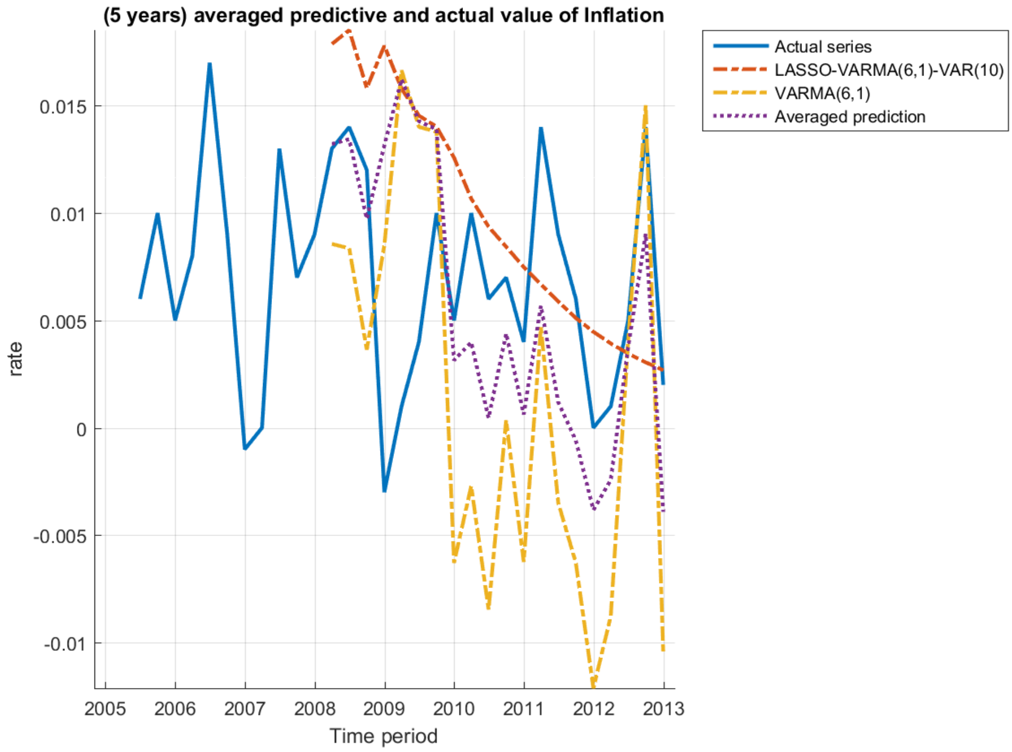

Figure B.26: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of π_t

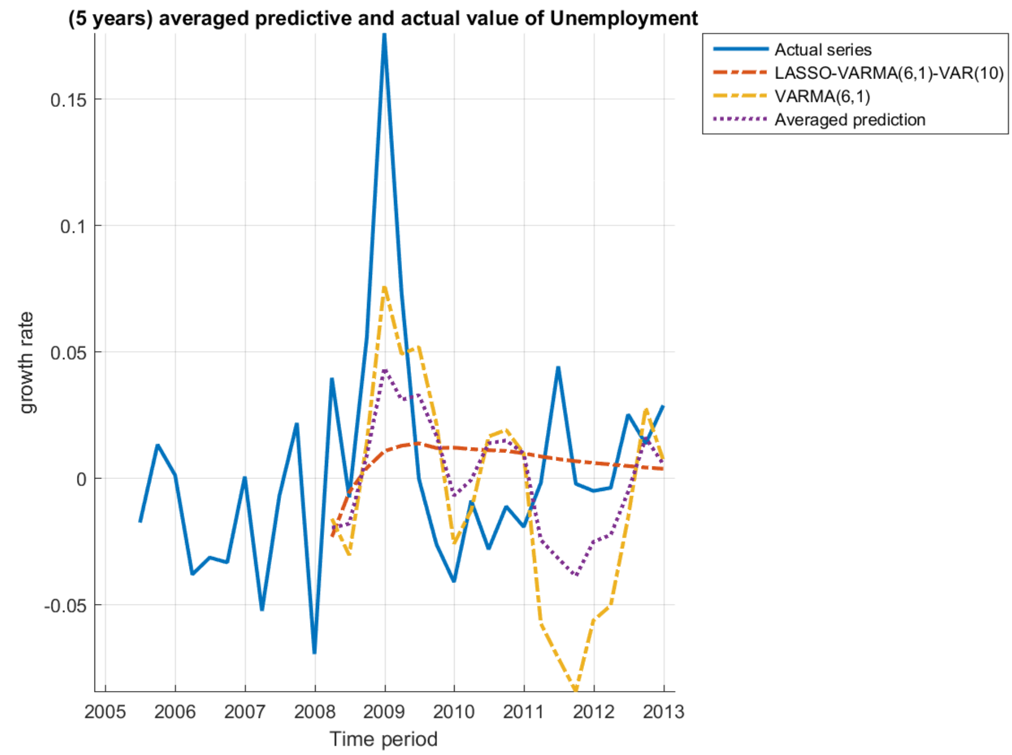

Figure B.27: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$

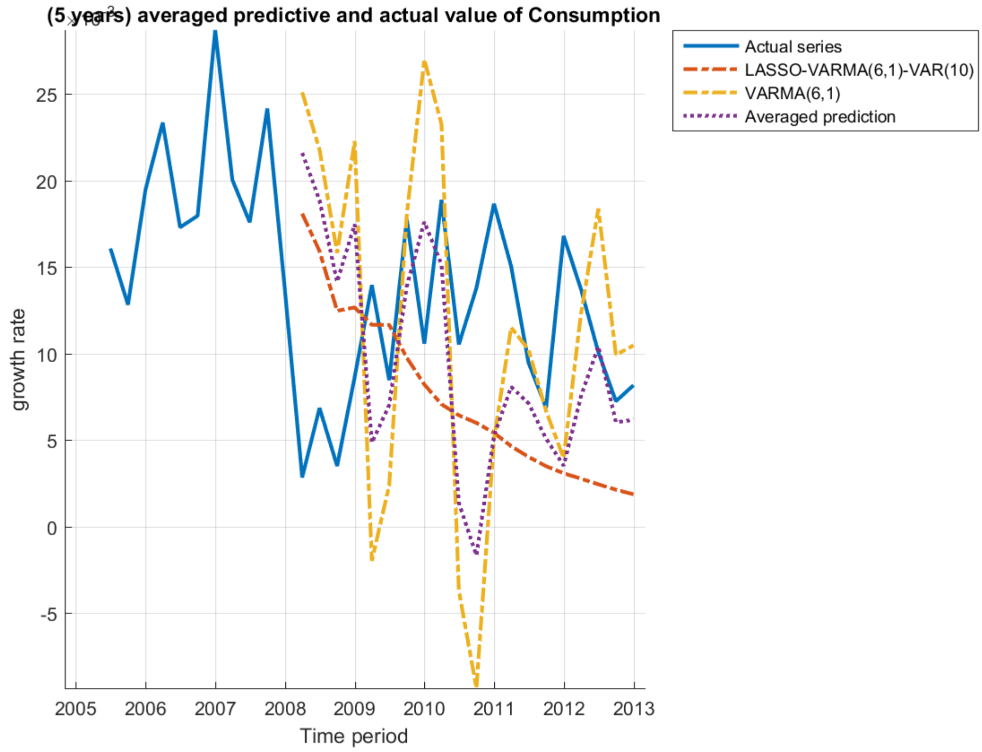

Figure B.28: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$

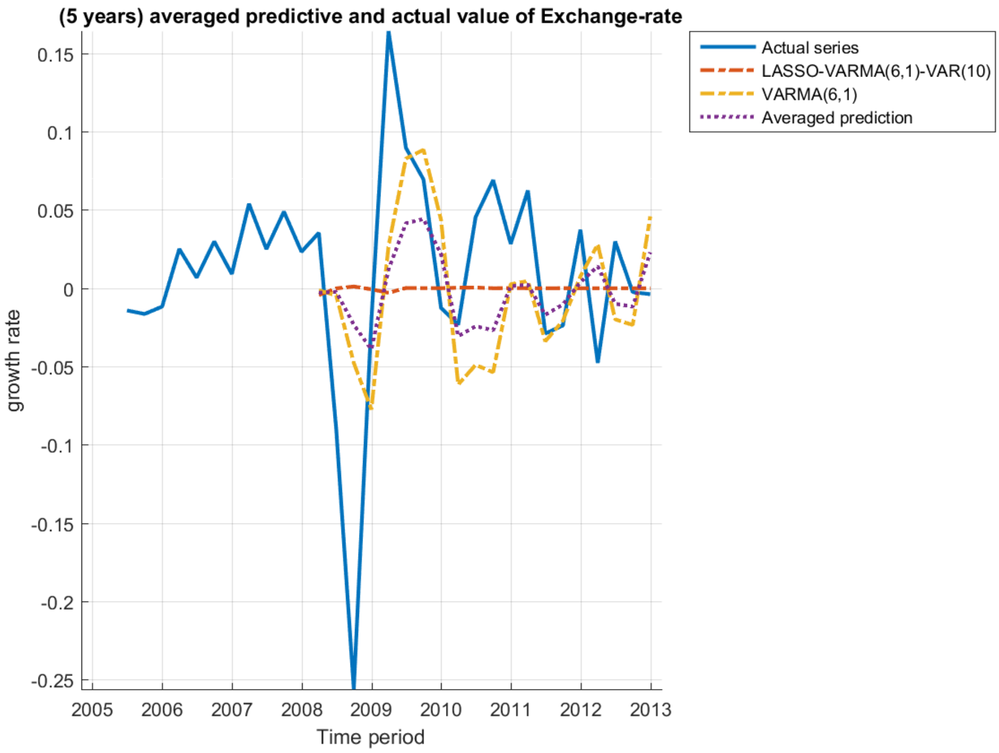

Figure B.29: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$

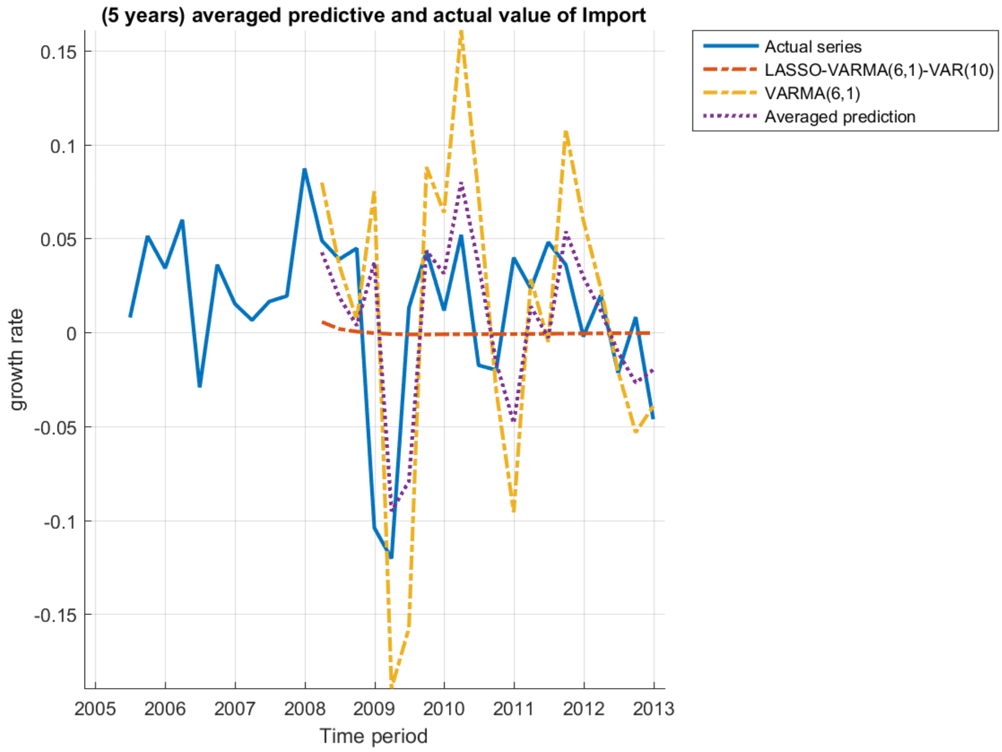

Figure B.30: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

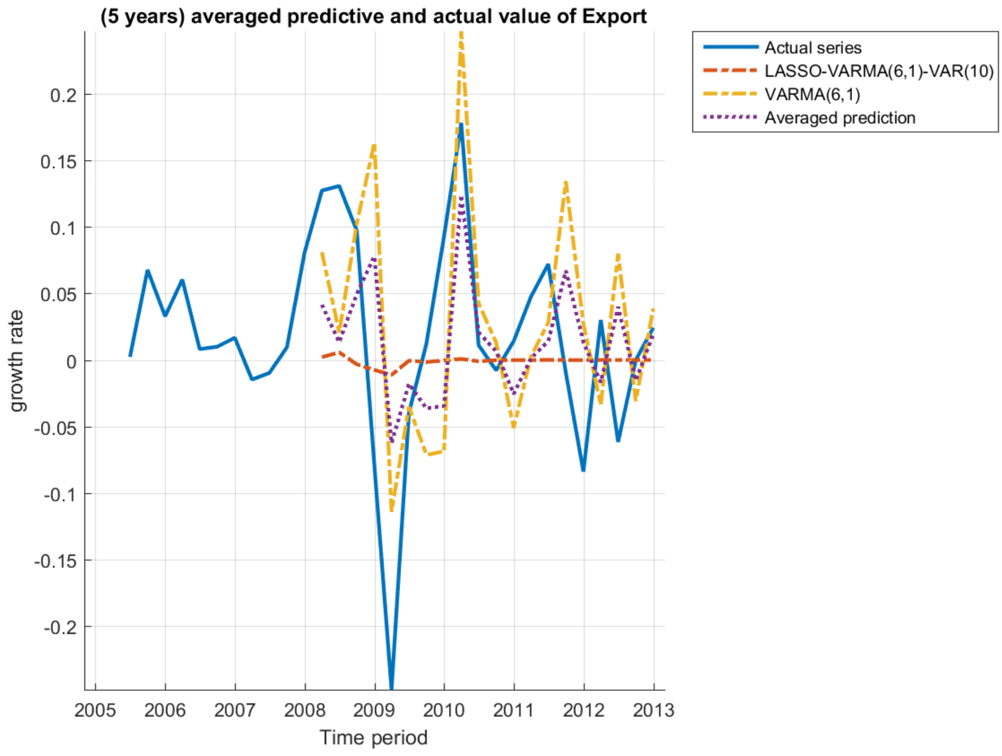

Figure B.31: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$

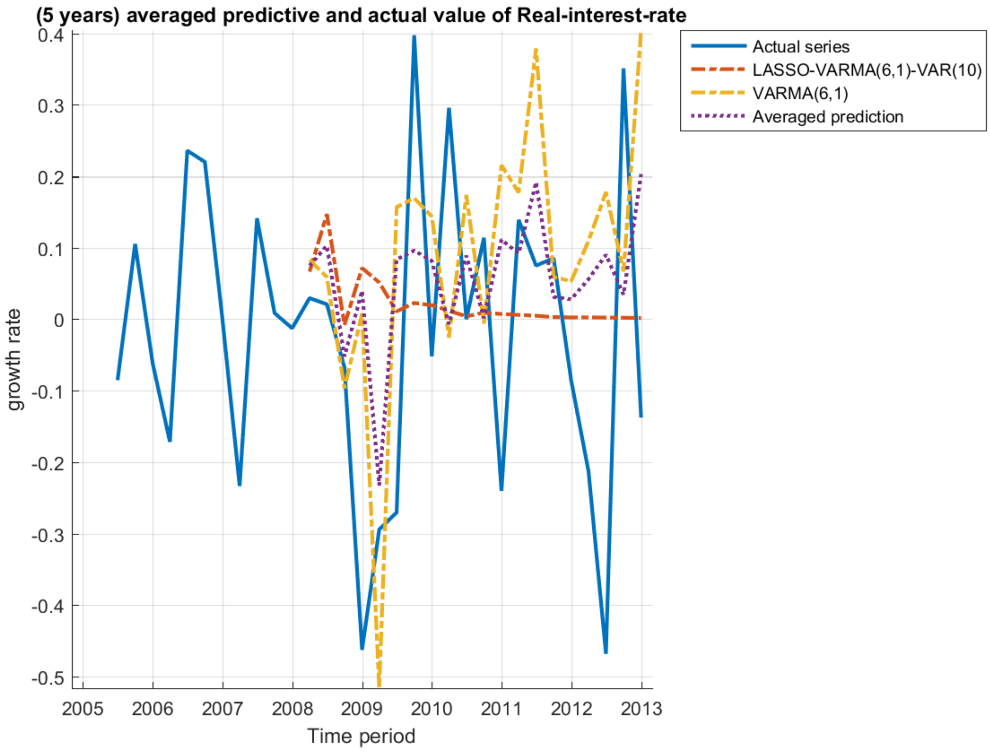

Figure B.32: CV- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle R_t$

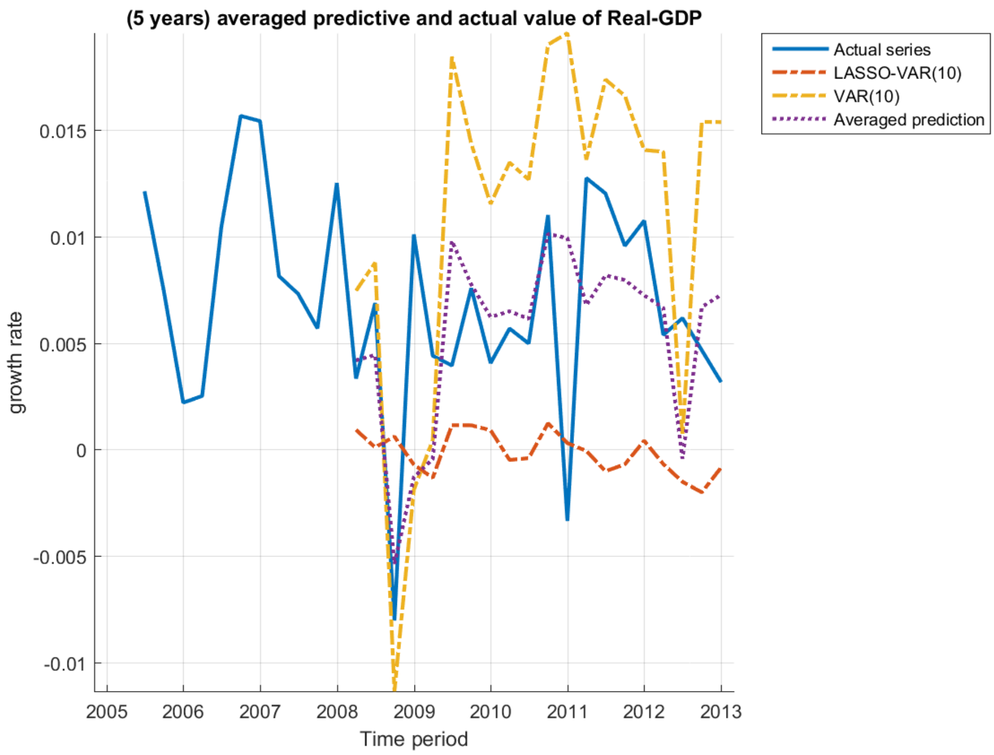

Figure B.33: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle Y_t$

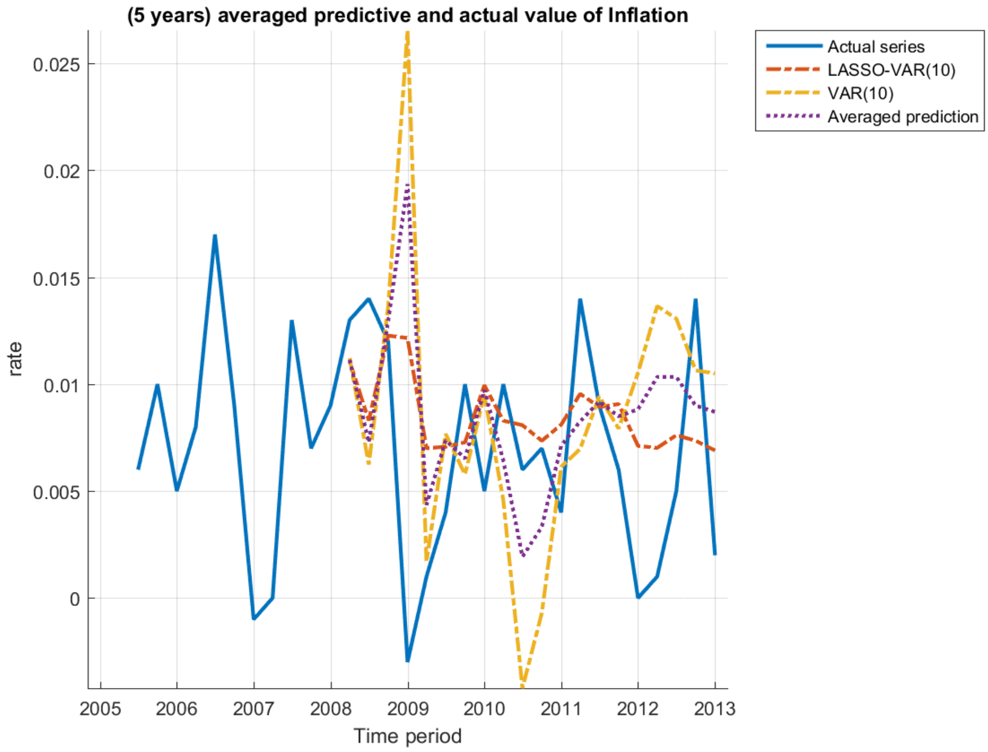

Figure B.34: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of π_t

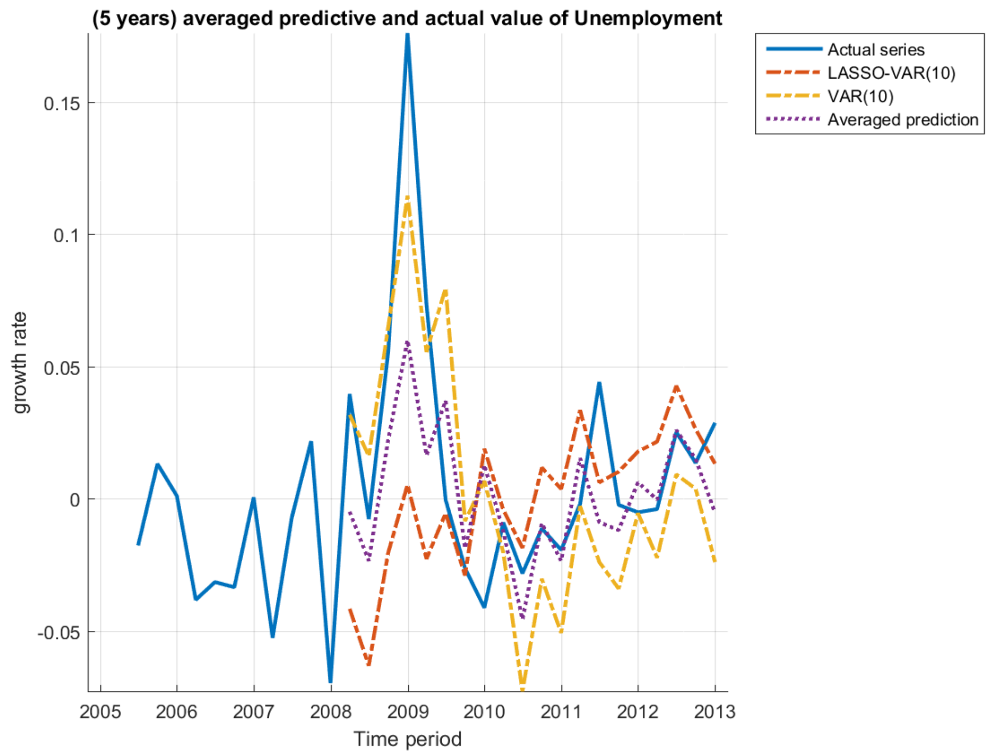

Figure B.35: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$

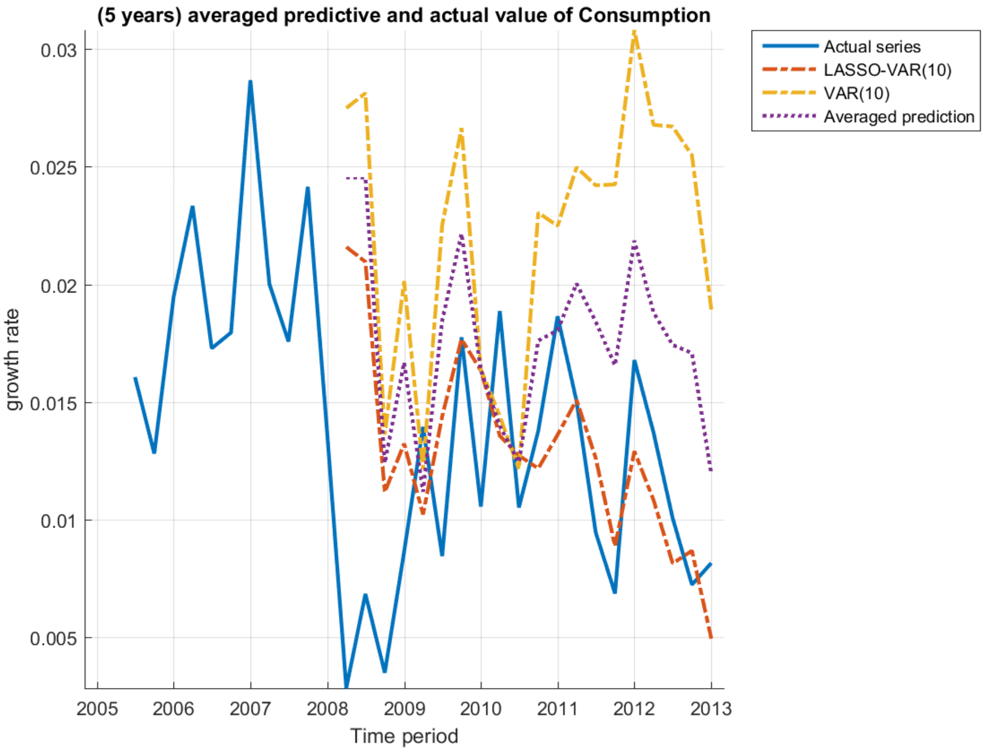

Figure B.36: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$

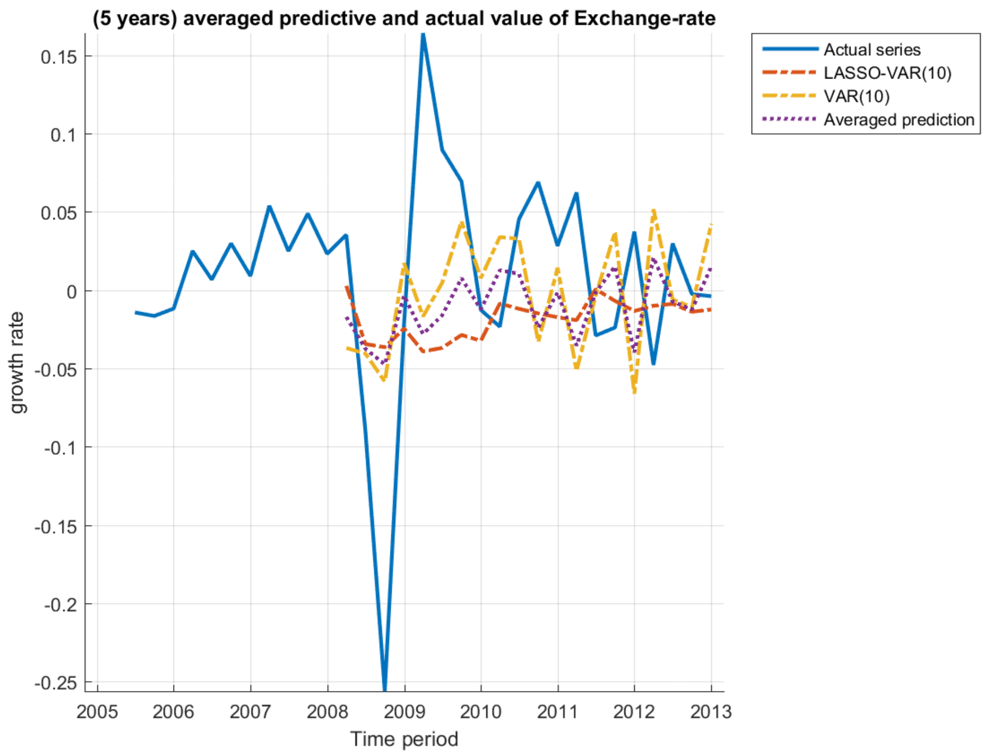

Figure B.37: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$

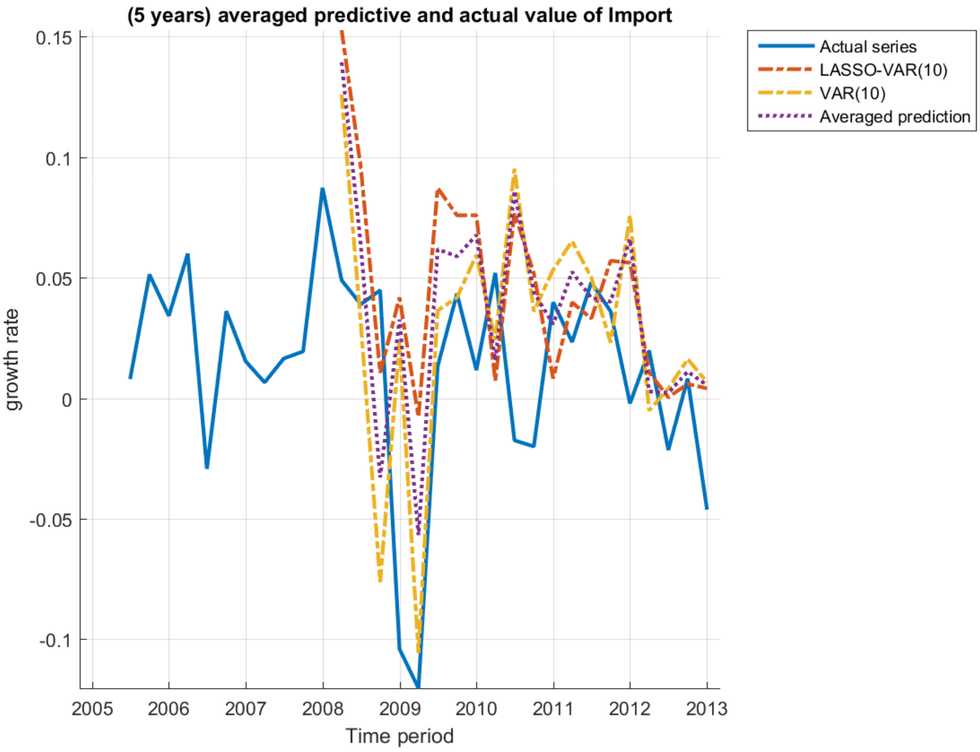

Figure B.38: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

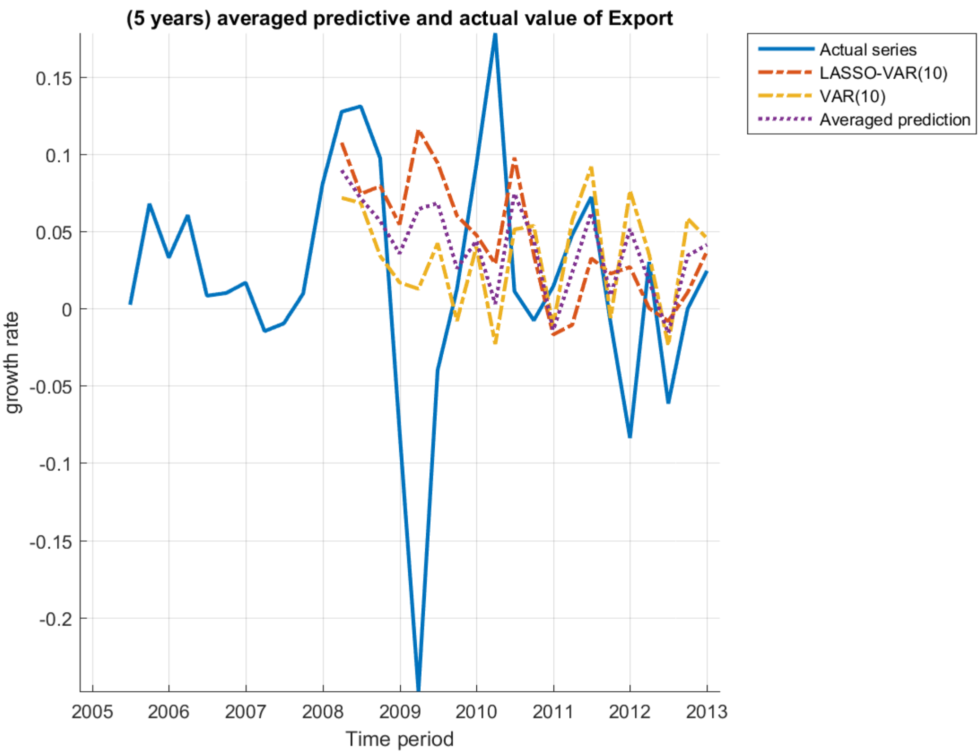

Figure B.39: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$

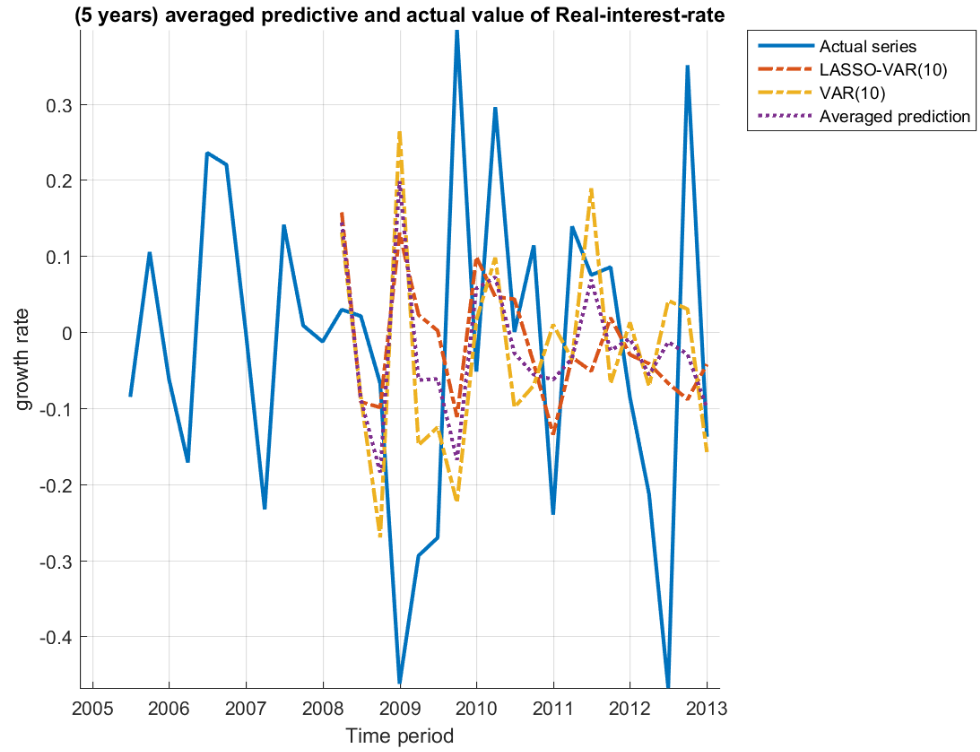

Figure B.40: RMSE- LASSO-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

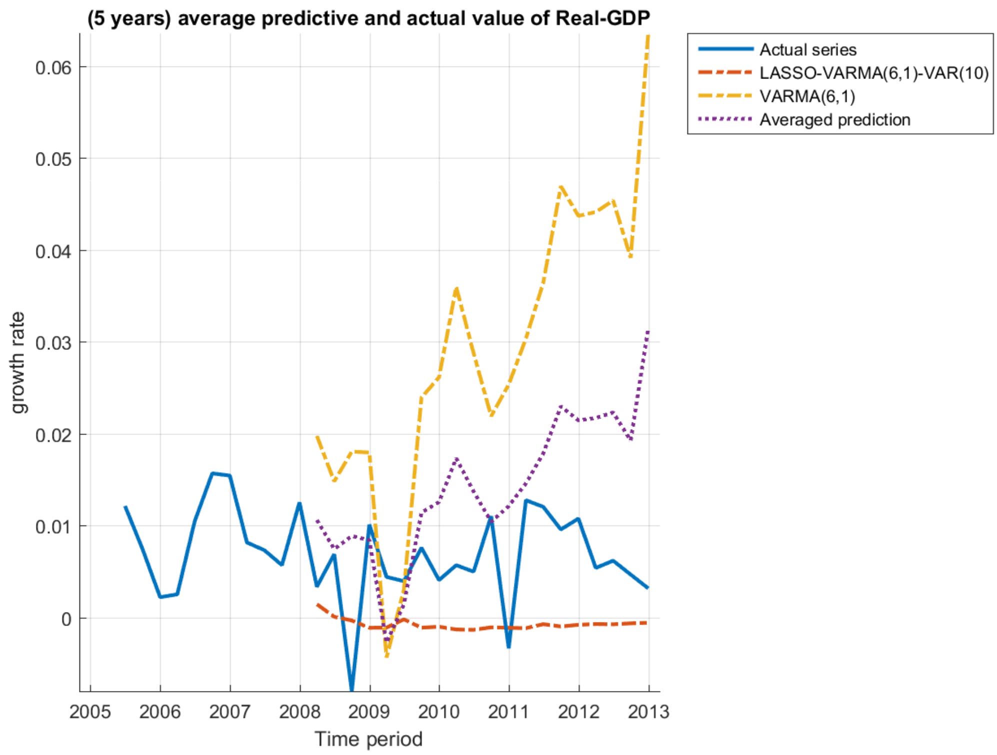

Figure B.41: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years steps ahead averaged forecast of $\triangle Y_t$

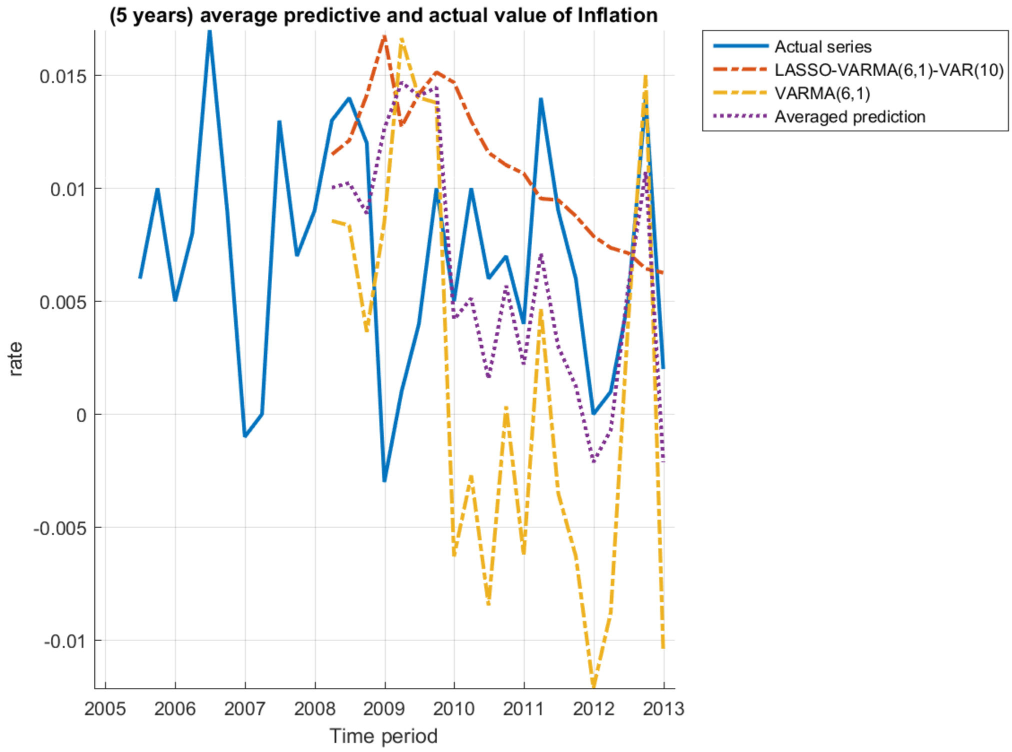

Figure B.42: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of π_t

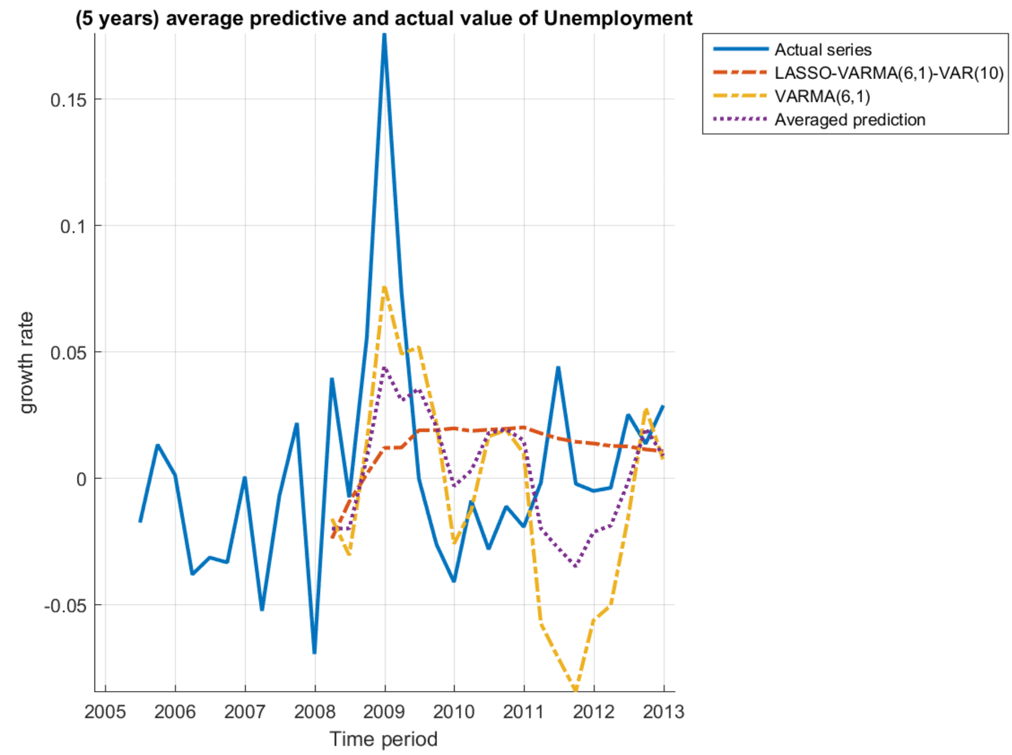

Figure B.43: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle U_t$

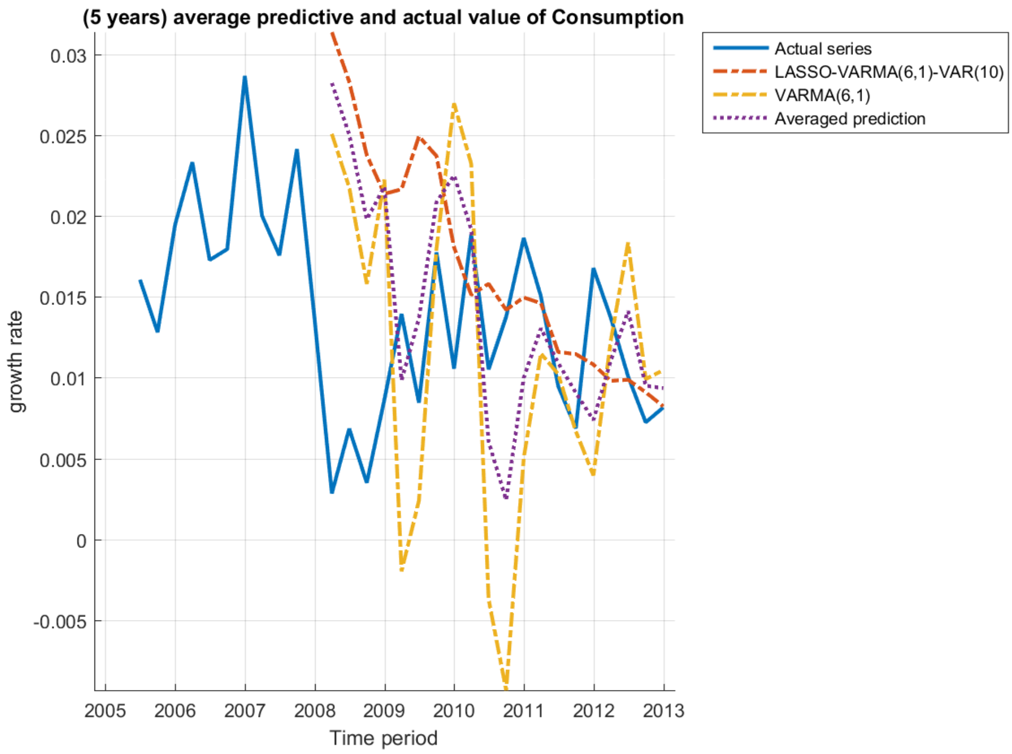

Figure B.44: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle C_t$

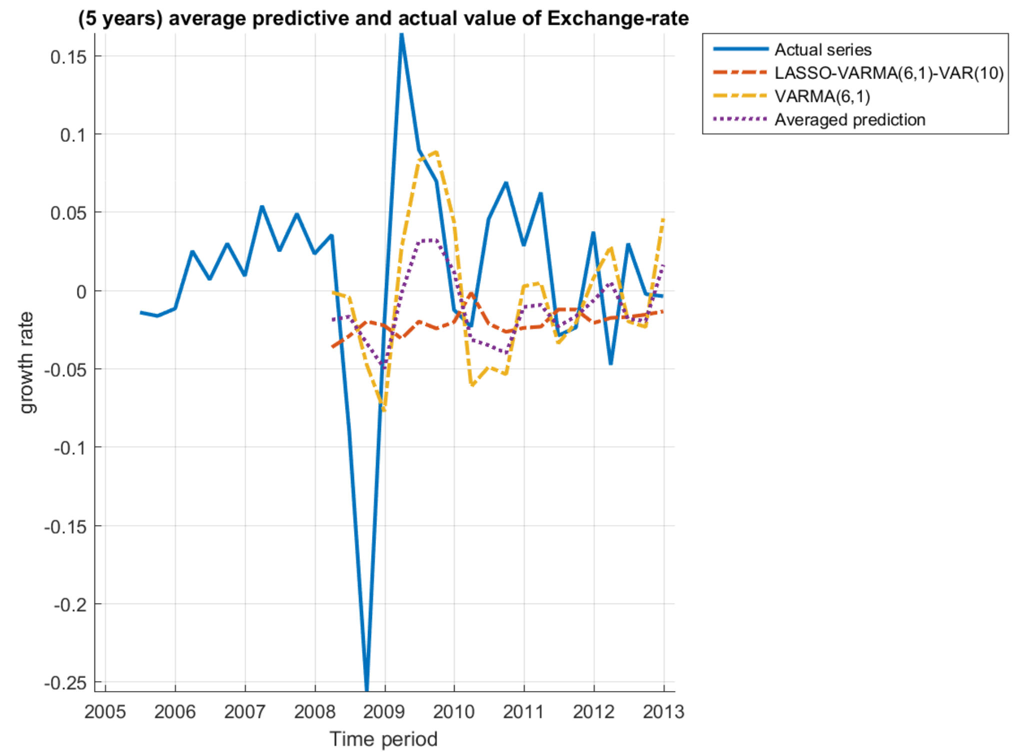

Figure B.45: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle EXR_t$

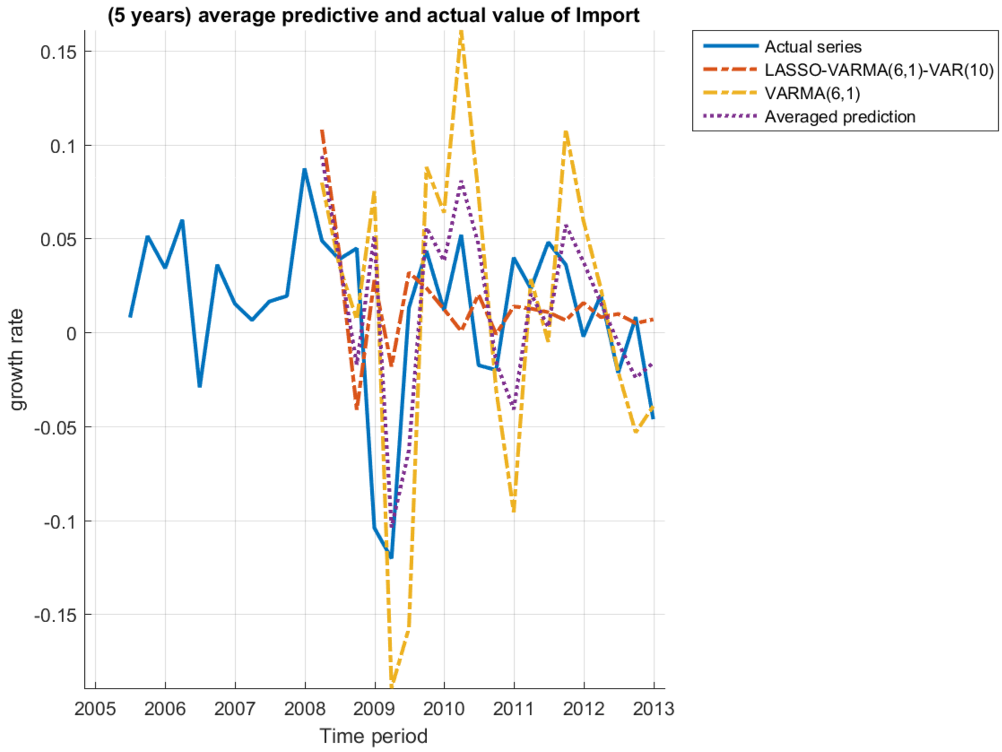

Figure B.46: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle I_t$

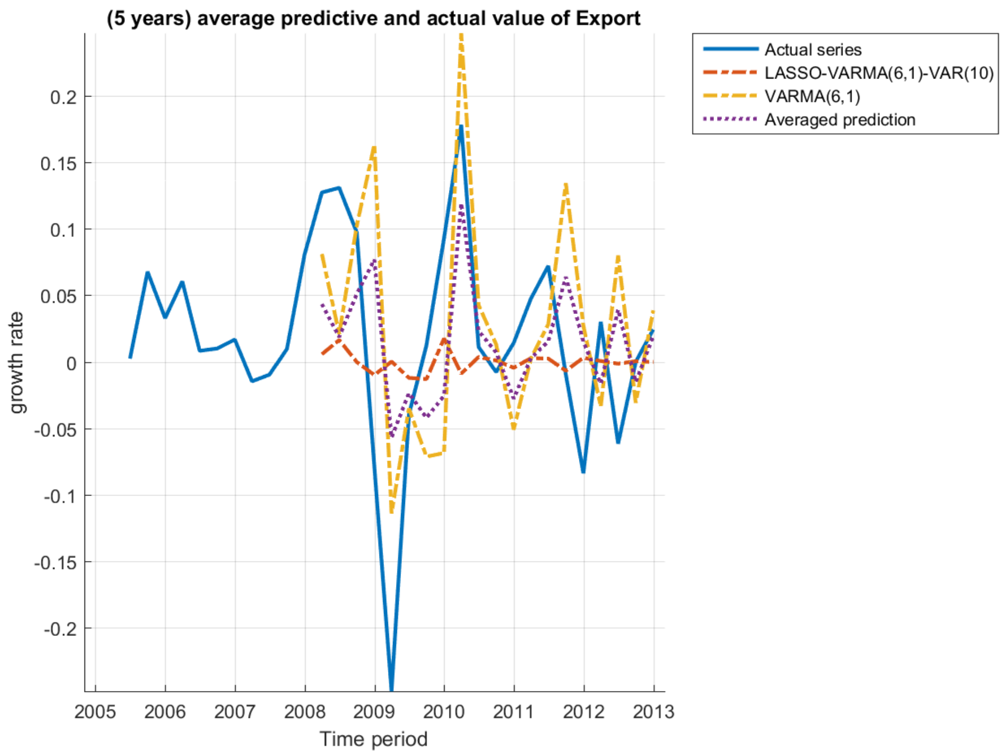

Figure B.47: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle E_t$

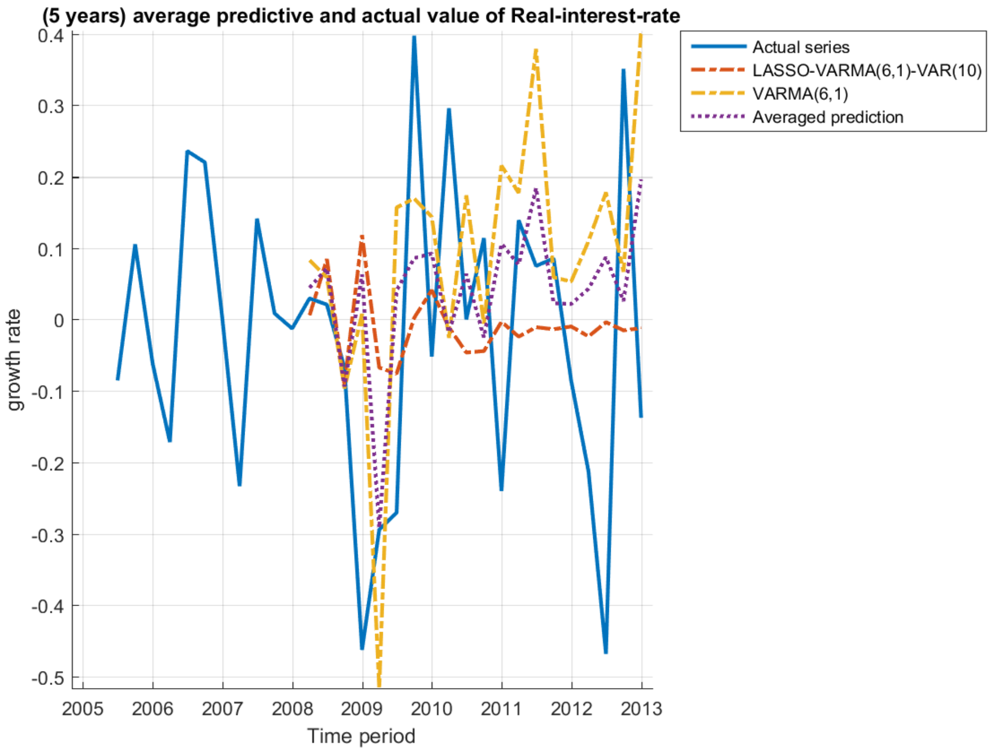

Figure B.48: RMSE- LASSO- VARMA(6,1)-VAR(10) 5 years ahead averaged forecast of $\triangle R_t$

Bibliography

An, S. and F. Schorfheide (2007). Bayesian analysis of DSGE models. *Econometric reviews 26*(2-4), 113–172.

Ansley, C. F. and R. Kohn (1990). A note on square root filtering for vector autoregressive moving-average models. *Journal of Time Series Analysis 11*(3), 181–183.

Basu, S., A. Shojaie, and G. Michailidis (2015). Network granger causality with inherent grouping structure. *Journal of Machine Learning Research 16*, 417–453.

Billah, B., M. L. King, R. D. Snyder, and A. B. Koehler (2006). Exponential smoothing model selection for forecasting. *International journal of forecasting 22*(2), 239–247.

Bodman, P. M. (1998). Asymmetry and duration dependence in australian GDP and unemployment. *Economic Record 74*(227), 399–411.

Broze, L., C. Gouriéroux, and A. Szafarz (1995). Solutions of multivariate rational expectations models. *Econometric theory 11*(02), 229–257.

Casals Carro, J., A. García Hiernaux, and M. Jerez Méndez (2010). From general state-space to varmax models.

Chen, K. and K.-S. Chan (2011). Subset ARMA selection via the adaptive lasso. *Statistics and its Interface 4*(2), 197–205.

Coibion, O. and Y. Gorodnichenko (2008). What can survey forecasts tell us about informational rigidities? Technical report, National Bureau of Economic Research.

Del Negro, M. and F. Schorfheide (2012). Dsge model-based forecasting. *Available at SSRN 2018451*.

Diebold, F. X. and R. S. Mariano (2012). Comparing predictive accuracy. *Journal of Business & economic statistics*.

Diks, C., V. Panchenko, and D. Van Dijk (2010). Out-of-sample comparison of copula specifications in multivariate density forecasts. *Journal of Economic Dynamics and Control 34*(9), 1596–1609.

Elliott, G. and A. Timmermann (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics 122*(1), 47–79.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics Springer, Berlin.

Giacomini, R. (2013). The relationship between DSGE and VAR models. *Advances in Econometrics 31.*

Giacomini, R., V. Skreta, J. Turén, et al. (2015). Models, inattention and expectation updates. Technical report.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gibbs, C. G. and M. Kulish (2015). Disinflations in a model of imperfectly anchored expectations. *Available at SSRN*.

Harvey, A. (2006). Forecasting with unobserved components time series models. *Handbook of economic forecasting 1*, 327–412.

Haufe, S., G. Nolte, K.-R. Mueller, and N. Krämer (2009). Sparse causal discovery in multivariate time series. *arXiv preprint arXiv:0901.2234*.

Honkapohja, G. E.-S. and G. Evans (2001). Learning and expectations in macroeconomics. *Princeton Uni-versity Press, Princeton.*

Hsu, N.-J., H.-L. Hung, and Y.-M. Chang (2008). Subset selection for vector autoregressive processes using LASSO. *Computational Statistics & Data Analysis 52*(7), 3645–3657.

Karapanagiotidis, P. (2014). Dynamic state-space models.

Keynes, J. M. (1937). The general theory of employment. *The quarterly journal of economics*, 209–223.

Krolzig, H.-M. (2003). General-to-specific model selection procedures for structural vector autoregressions. *Oxford Bulletin of Economics and Statistics 65*(s1), 769–801.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Mariano, R. S. and D. Preve (2012). Statistical tests for multiple forecast comparison. *Journal of econometrics 169*(1), 123–130.

Neumeyer, P. A. and F. Perri (2005). Business cycles in emerging economies: the role of interest rates. *Journal of monetary Economics 52*(2), 345–380.

Nicholson, W., D. Matteson, and J. Bien (2014). Structured regularization for large vector autoregressions. *Cornell University*.

Nossman, M. and A. Wilhelmsson (2009). Is the VIX futures market able to predict the VIX index? a test of the expectation hypothesis (digest summary). *Journal of Alternative Investments 12*(2), 54–67.

Park, T. and G. Casella (2008). The Bayesian LASSO. *Journal of the American Statistical Association 103*(482), 681–686.

Penzer, J. and B. Shea (1997). The exact likelihood of an autoregressive-moving average model with incomplete data. *Biometrika 84*(4), 919–928.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–19571. *economica 25*(100), 283–299.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies 73*(3), 793–821.

Ren, Y. and X. Zhang (2010). Subset selection for vector autoregressive processes via adaptive LASSO. *Statistics & probability letters 80*(23), 1705–1712.

Robertson, J. C. and E. W. Tallman (1999). Vector autoregressions: forecasting and reality. *Economic Review-Federal Reserve Bank of Atlanta 84*(1), 4.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. In *Soviet Math. Dokl.*, Volume 5, pp. 1035–1038.

Trevor, R. G. and S. J. Thorp (1988). VAR forecasting models of the australian economy: A preliminary analysis. *Australian Economic Papers 27*(s1), 108–120.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.