Bayesian Vector Autoregressions and its Applications in Macroeconomics

By

Hengwei Qi

Submitted to the Department of Economics and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members

Professor Shu Wu, Chairperson

Professor Terry Soo

Professor Ted Juhl

Professor Jianbo Zhang

Professor Pym Manopimoke

Date defended: May 13, 2015

The Dissertation Committee for Hengwei Qi certifies
that this is the approved version of the following dissertation :

Bayesian Vector Autoregressions and its Applications in Macroeconomics

Professor Shu Wu, Chairperson

Date approved: May 13, 2015

Abstract

This dissertation presents three essays which are organized as chapters. Each chapter, with its own feature, is focusing on parsimonious estimation of Vector Autoregressive (VAR) models. The newly presented models are applied to some issues in macroeconomics.

VARs are widely used in economic analysis and forecasting probably due to its easy tractability and the power in exposing and exploring complicated dynamic process in a modern economy such as shocks, channels and their links. Researchers sometimes would like to include more observations in VARs in order to have a broad investigation into the economy. In addition, this analysis, in particular, currently has extended to time variation in parameters so as to capture the possibly changing dynamics. These two will inevitably cause parameter proliferation. The over-fitting caused by hoping including more variables and extending to time variation in parameters and wide applications because of tractability and detecting power motivate my research interest in parsimonious modeling, estimation and applications as well as comparisons in this dissertation.

In the first chapter, I extend the previous parsimonious estimation on constant coefficients to time varying coefficients. It is desirable since when time dimension is taken into account, the number of parameters rises with time periods while number of observations still fixed there, the same as in constant parameter estimation. I use stochastic variable selection method over the time dimension, that is, the time varying dynamics of each coefficient in a TVP-VAR model is checked. I estimate the model via Bayesian

method based on the sensitivity of variation of conditional likelihood when the coefficient considered in the model or not. If the coefficient has more contribution to the likelihood, the posterior tends to give more probability of staying in the system. Nevertheless, if the parsimonious check is imposed on every coefficient in the model, the computational burden will substantially increase because this estimation is essential a mixed model estimation such that every possible model needs to be checked. The number of candidate models increases with time dimension. I suggest two ways to alleviate it. One is from model setting that we can check the variables we are interested in collected together as a whole. In other words, we can do block checking rather than single checking. The other is from the efficiency of the numerical computation which is conducted in two dimensions. We construct large matrices to replace Kalman forward filter and backward smoother in order to reduce the procedures in each iteration in Bayesian simulation. On the other hand, the structure of the large matrix is sparse which further make the computation efficient. The single-checking-based TVP–VAR with stochastic volatility is used to estimate the changes in monetary policy stance and agents' behavior to policy shocks over time. With the most parsimonious estimation, I still cannot find significant changes both in policy stance and in the reaction of economic agents to the non-systematic monetary policy shocks.

In the second chapter, I present a general parsimonious estimation on the time varying parameter VAR with stochastic volatility via factor idea. That is, the far lags are driven by recent lags; the time variation in coefficients on regressors is driven by several factors and therefore the covariance matrix of innovations to the coefficients become reduced rank. Lastly, we use a latent factor, namely, the common volatility to represent full stochastic volatility based on the empirical evidence that the estimated volatilities of most macroeconomic variables share the similar pattern. Note that the model I present concentrates on how to reduce the dimension of the parameters, not the dimension of large data set such as factor models like dynamic factor models or factor

augmented models. The model is estimated by Bayesian simulation. Each estimation procedure or block is presented in this chapter. Based on the general treatment, the estimation procedures can be freely combined with some proper modification depending on the specific research object. I give an empirical analysis by the factor driven model. The analysis is based on the typical small scale monetary VAR. Principal component analysis shows that small scale TVP-VAR is still over-fitting very much, can be driven by several factors and that early lags are not suitable to drive far lags that will cause dynamic contamination. Therefore parsimonious estimation via factors on time varying coefficients and common volatility is used to estimate the small scale monetary VAR. Focusing on agent response to monetary policy shocks, I cannot find significant difference among different time periods.

In the last chapter, I consider a large Bayesian VAR that contains 28 variables. The variables cover a broad range of the U.S economy including labor market, housing market, bonds market and so on. High dimensional observations are desirable by researchers because this setting can give a strong background of the whole economy, reducing potential missing variables that are critical for the transmission of some shocks of interest. A large number of endogenous variables will increase the degree of parameter proliferation. I use proper priors that can shrink values of coefficients to conduct an empirical analysis on monetary policy shocks, financial shocks and uncertainty shocks. I find that for the effects of monetary policy shocks, the impulse response functions are almost perfectly in line with theoretical predictions. For the financial shocks and uncertainty shocks, we analyze them jointly. Both positive financial and uncertainty shocks have negative effects on real activities, however, financial shocks have more persistent effects on these variables than uncertainty shocks. We also find that financial variables care more for uncertainty shocks compared to financial shocks.

Acknowledgements

I wish to express my deep and sincere gratitude to my advisor, Professor Shu Wu, for his excellent guidance, care, patience, and for providing me with an excellent atmosphere for doing research. His invaluable suggestions and help hasten the cross of Bayesian econometrics and macroeconomics in my dissertation.

I would like to thank Professor Terry Soo, in Department of Mathematics, for reading this thesis word by word from grammar to contents and giving me lots of important suggestions. Special thanks go to Professors Ted Juhl, Jianbo Zhang and Pym Manopimoke for being in my dissertation committee.

Understanding the theory of Bayesian econometrics is one thing, putting it into practice is another thing. I would also like to thank Professor Dimitris Korobilis, at University of Glasgow, for his suggestion in email contact and code sharing and explanation on his website. Without his code and suggestion, I cannot have a good and deep understanding of Bayesian theory and improve my code writing skill. I learn much from his website.

Finally, I would like to express my deepest gratitude to all the people who have helped me in my research and dissertation writing during the Ph.D life at University of Kansas. Without them, I would never have been able to finish my dissertation. I owe much to them.

Contents

1 The Restricted Bayesian Vector Autoregressions and Monetary Policy Estimation 1

1.1 Introduction . . . 2

1.2 Models in stochastic variable selection . . . 7

1.2.1 Stochastic variable selection Bayesian VAR . . . 9

1.2.2 Stochastic variable selection partial TVP-VAR . . . 11

1.2.3 Stochastic variable selection full TVP-VAR . . . 14

1.3 Bayesian estimation . . . 19

1.3.1 Stochastic volatility . . . 19

1.3.2 Stochastic selection . . . 20

1.3.3 Exercise . . . 21

1.4 Monetary policy in stochastic selection models . . . 22

1.4.1 Data description . . . 22

1.4.2 Lags and significance . . . 24

1.4.3 Impulse responses . . . 30

1.4.4 Systematic or exogenous change . . . 32

1.4.4.1 Agents bahavior to monetary shocks . . . 33

1.4.4.2 Systematic monetary policy . . . 33

1.4.4.3 Non-systematic monetary policy shocks . . . 42

1.4.4.4 An interpretation . . . 42

1.5 Robust check . . . 48

1.6 Concluding remarks . . . 55
Appendix . . . 61
A. The basic Kalman filter and state smoother . . . 61
B. Modified Kalman filter and state smoother with breaks . . . 63
C. Estimation of states without Kalman filter and state smoother . . . 64
D. Efficient estimation of SVS-Full-TVP-VAR . . . 66

2 A General Parsimonious Estimation of Time-varying Vector Autoregressions 70

2.1 Introduction . . . 71
2.2 Model specification . . . 77
2.3 Bayesian estimation procedure . . . 80
2.3.1 Draw a history of $\{b_t\}_{t=1}^T$ . . . 80
2.3.1.1 Draw a history of $\{\tilde{b}_t\}_{t=1}^T$ . . . 81
2.3.1.2 Draw $M_b b_0$ as a whole . . . 82
2.3.1.3 Draw a history of $\{b_t\}_{t=1}^T$ . . . 84
2.3.2 Draw reduced rank covariance matrix Q_b . . . 84
2.3.3 Draw constant matrix B . . . 84
2.3.4 Draw constant matrix G_1 . . . 86
2.3.5 Draw structural impact matrix A . . . 87
2.3.6 Draw diagonal elements of S . . . 88
2.3.7 Draw common stochastic volatility $\{\lambda_t\}_{t=1}^T$ . . . 88
2.3.8 Draw coefficient F . . . 91
2.3.9 Draw variance Q_h . . . 92
2.4 Empirical analysis . . . 93
2.4.1 Are factors important in time varying parameters? . . . 94
2.4.2 Structural analysis and model evaluation . . . 102
2.5 Concluding remarks . . . 113
Appendix . . . 114

A. Bayesian estimation of restricted linear regression model 114

3 Structual Analysis in a Large Bayesian VAR **117**

3.1 Introduction 118

3.2 The literature 120

3.3 The model 124

3.3.1 The likelihood of VAR 124

3.3.2 The priors 126

3.3.3 The posterior 129

3.4 Empirical analysis 129

3.4.1 The data 130

3.4.2 Empirical findings 130

3.5 Concluding remarks 133

Appendix 137

A. Data specification 137

List of Figures

1.1 IRs for model 2 34
1.2 IRs for model 3 35
1.3 IR-inflation-comparison for model 3 36
1.4 IRs-unemployment rate-comparison for model 3 37
1.5 IRs-inflation-comparison for model 2 38
1.6 IRs-unemployment rate-comparison for model 2 39
1.7 3D-IRs for model 2 40
1.8 3D-IRs for model 3 41
1.9 Policy response to inflation for model 3 43
1.10 Policy response to inflation for model 2 44
1.11 Policy response to unemployment rate for model 3 45
1.12 Policy response to unemployment rate for model 2 46
1.13 Volatility for each variable in model 3 47
1.14 Posterior probability for each coefficient in policy rule 49
1.15 Posterior probability of each coefficient in inflation equation 50
1.16 Poterior probability of coefficient in unemployment rate equation 51
1.17 Posterior probability of each regime 54
1.18 IRs of inflation and their difference between different regimes 56
1.19 IRs of unemployment rate and their difference between different regimes 57
1.20 IRs of interest rate to inflation shocks 58

1.21 IRs of interest rate to unemployment shocks 59

2.1 Contribution and cumulative contribution for lag 1 96
2.2 Contribution and cumulative contribution for lag 2 97
2.3 Contribution and cumulative contribution for lag 3 98
2.4 Contribution and cumulative contribution for lag 4 99
2.5 TVP-VAR-CV-lag1-factor7-IR-inflation 103
2.6 TVP-VAR-CV-lag1-factor7-IR-unemployment rate 104
2.7 TVP-VAR-CV-lag2-factor3-IR-inflation 105
2.8 TVP-VAR-CV-lag2-factor3-IR-unemployment rate 106
2.9 TVP-VAR-CV-lag3-factor3-IR-inflation 108
2.10 TVP-VAR-CV-lag3-factor3-IR-unemployment rate 109
2.11 TVP-VAR-CV-lag4-factor3-IR-inflation 110
2.12 TVP-VAR-CV-lag4-factor3-IR-unemployment rate 111
2.13 Common volatility in different model settings 112

3.1 IRs to monetary policy shocks 134
3.2 IRs to uncertainty shocks 135
3.3 IRs to financial shocks 136

List of Tables

1.1 Exercise result 23

1.2 Lag check1 25

1.3 Lag check2 26

1.4 Lag check3 26

1.5 Lag check4 27

1.6 Lag check5 27

1.7 Lag check6 28

1.8 Lag check7 28

1.9 Lag check8 29

1.10 Lag check9 29

1.11 Lags in model 2 30

2.1 Contribution for each factor via principal component analysis on covariance matrix Q_b for TVP-VAR-CV from lag one to lag two 95

2.2 Contribution for each factor via principal component analysis on covariance matrix Q_b for TVP-VAR-CV from lag three to lag four 95

3.1 Data specification 138

Chapter 1

The Restricted Bayesian Vector Autoregressions and Monetary Policy Estimation

1.1 Introduction

Many economists have found that there is strong evidence that the U.S. economy experienced higher and more volatile unemployment and inflation in the last century 70's and the early 80's than the following years. It's natural to relate the U.S. monetary policy to the bad performance and ask the question of whether high inflation and slow growth during that period were due to bad policy or bad luck. The literature generally gives two main explanations. One (e.g. Blanchard and Simon, 2001; Koop, Leon-Gonzalez and Strachan, 2009; Primiceri, 2005; Sims and Zha, 2006; Stock and Watson, 2002a) focuses on exogenous shocks, which have been much more volatile in the 1970's and early 1980's than in the rest of the sample. They argued that the variance of the exogenous shocks has changed over time and that this alone may explain many apparent changes in monetary policy. This is 'bad luck' story, i.e. in the 1970s volatility was high, whereas later policymakers had the good fortune of the Great Moderation of the business cycle. Koop, Leon-Gonzalez and Strachan (2009) and Primiceri (2005) using the method of time-varying vector autoregression, they still find exogenous shocks play an overwhelming role. Therefore, I put them in this respect. The second emphasizes the changes in the transmission mechanism—the way macroeconomic variables respond to shocks. Particular attention gives to monetary policy reaction function. Some authors (e.g. Boivin and Giannoni, 2006, Cogley and Sargent, 2001, and Lubik and Schorfheide, 2004)) have argued that the way the Fed reacted to inflation has changed over time. They think that the Fed was less active in fighting inflation pressures under the chairmanship of Burns than under Volker and Greenspan. This is 'bad policy' story.

However, these explanations are controversial. For example, Bernanke and Mihov (1998), Hanson (2003), Leeper and Zha (2003) found that litter evidence of changes in the systematic part of monetary policy and Sims (1999, 2001) found no evidence of unidirectional drifts in policy toward a more active behavior.

From the methodological perspective, many versions of Bayesian Vector Autoregressions (BVAR) in this field are well developed. Here, I just pick up some representatives. Sims and Zha (2006) used Bayesian VAR with multivariate stochastic volatility but no time-varying parameters. Cogley

and Sargent (2001) developed a time-varying parameters VAR model without stochastic volatility. Later they (2005) further extend a Bayesian VAR model with time-varying parameters and multivariate stochastic volatility. The volatility however is too restrictive. This model allows the simultaneous relations among variables are time invariant i.e. covariance is constant and variance is dynamic. In impulse response analysis, it can be shown that this restriction implies that a shock to the i^{th} variable has an effect on the j^{th} variable which is constant over time. In some macroeconomic applications, relationships might not be the case. Boivin (2001) considered time varying simultaneous relations, but ignored potential heteroscedasticity of innovations. Ciccarelli and Rebucci (2003) extend a t-distribution of errors based on the Boivin (2001) model; Finally, Primiceri (2005) got the time varying parameters and loose multivariate volatility together. Loose multivariate volatility means no restriction imposed on the covariance and variance parameters in error matrix, totally time-varying now. This flexible approach can be regarded as today's standard time varying VAR workhorse and is widely used in empirical macroeconomics.

It seems surprising, however natural that all models above and further modifications are all Bayesian. First, if the model is a time invariant parameter VAR, in terms of classical estimation, for instance, a VAR involving 5 dependent variables with 4 lags contains 105 parameters. Generally, in a quarterly macroeconomic data set, the number of observations on each variable might be a few hundred and for some reason, usually a segment, not all range is picked up for a specific analysis. That implies that maybe a few data is shared for each parameter estimation—the information density is too low. If the model is extended to a time variant one, apparently more parameters will be created. Think how many observations of data we should have to make the estimation and further features of interest like forecasts and impulse responses precise. It's impossible! Second, if the variance of the time varying coefficients is small, the classical maximum likelihood estimator of this variance has a point mass at zero. This is so called pile-up problem (Sargent and Bhargava, 1983; Shephard and Harvey, 1990; Stock and Watson, 1998). The third, the maximum likelihood method often meets optimization problem when model dimensionality is multiple and no longer linear. Such a complicated model will highly probably have a multivariate likelihood function of

regions containing many local peaks and it's difficult to distinguish which one is global. Moreover, if these peaks are very narrow, the likelihood may reach particularly high values, not at all representative of the model's fit on a wider and more interesting parameter region. Altogether, rich parameters and potential problems in maximum likelihood with limited observations of data sets make the classical estimation not suitable for some researches. A Bayesian method is naturally introduced. In a Bayesian setting, not as many observations are needed, of course the more, the better; Given a uninformative prior together with likelihood function, posterior mean and variance of parameters and features of interest can be obtained from convergent therefore stable distributions by using Markov chain Monte Carlo (MCMC) method. The Gibbs sampler is a variant of MCMC and widely used in these models. It consists of drawing from lower dimensional conditional posteriors as opposed to the high dimensional joint posterior of the whole parameter set.

Bayesian vector autoregressions with Markov Chain Monte Carlo are a good and suitable approach to estimate parameters, especially in the form of time variation. However, implicitly it is assumed that all parameters of every draw from some stable posterior distributions are efficient and should exist in the model. Simply speaking, all parameters in this Bayesian model are given significant before you start estimate, i.e., every one is counted in. In reality, it's not the case. Let's compare it with classical vector autoregressions first. In the context of classical estimation, the usual way to test a coefficient in a model whether or not significant relative to zero is t-test or F-test. We get the p-value of the coefficient and then compare it with significant level in some distribution to decide reject or accept the null hypothesis. Anyhow, a specific way of testing parameter significance is already at hand in the field of classical estimation. Nevertheless, when the same concern comes to Bayesian VAR and determine which coefficient is significant by some methodology, how to deal with this problem? I am very interested in it.

To my best knowledge, I only find two articles about this issue in Bayesian field. One is Kuo and Mallick (1997), the other is Korobilis (2013b). The former presented the coefficients to be precisely zero if the indicator of the corresponding coefficient is zero (for instance, γ_j is the corresponding indicator for β_j, when $\beta_j = 0$ if $\gamma_j = 0$); The latter has extended the use of such methods

to VARs for forecasting considering possible economic structure change and policy regime switch that has been well documented in related literature, especially in the field of monetary policy and business cycles. A Markov chain Monte Carlo method (I will give details later) is created to obtain the posterior probability of each indicator after many times of iterations, therefore the posterior probability of parameter corresponding to this indicator. A criteria probability is set to compare with the probability of each indicator, namely, corresponding parameter. If the posterior probability of the parameter is greater than the criteria probability, then the parameter is said to be significant with high confidence and should stay in the Bayesian model. It looks like a test as in classical estimation.

This restricted vector autoregression method chosen by indicators has three advantages. First, variable selection is automatic, meaning that along with estimates of parameters we get associated probabilities of inclusion of each parameter in the 'best model'. For instance, in a VAR of 5 dependent variables with 4 lags I used before, the indicators tells us which elements of coefficient matrices should be included or excluded from the final optimal model, thus implementing a selection among all possible VAR model combinations, without the need to estimate each and everyone of these models. Second, this form of Bayesian variable selection is independent of the prior assumptions about parameter matrices. That is, if the researcher had defined any desirable prior for her parameters of the unrestricted model, adopting the variable selection restriction needs no other modification than one extra block in the posterior sampler that draws from the conditional posterior of the indicators. Finally, unlike other proposed stochastic search variable selection algorithms for VAR models (George et al., 2008), this form of variable selection may be adopted in many nonlinear extensions of VAR models.

The contribution of this paper is threefold. One is on methodology. In Korobilis (2013b), he extends stochastic variable selection method to constant parameter VAR and limited time-varying parameter VAR without stochastic volatility, respectively. This type of time variation in the context of TVP-VAR framework in Korobilis (2013b) only refer to one of coefficients either keep existing and time varying over time or not. Here, based on his model, I extend the potential time variation

to each and every coefficient, namely, it allows each and every coefficient of TVP-VAR to shift over time with corresponding posterior probability of present or not, but not necessarily the next time. Whether or not the coefficient is time varying is totally determined by data, period by period. Furthermore, I could also add loose stochastic volatility, i.e. both variance and covariance parameters of innovation matrix are time-variant – some econometricians call this volatility model heteroskedastic TVP-VAR – with several extra blocks embedded in original iteration route. One thing worth mention is that since the method looks into determining whether every and each coefficient in the TVP-VAR exist or not over time, the computational burden is high and thus time consuming. To increase computational efficiency, we use the precision based algorithm of Chan and Jeliazkov (2009) to replace typical procedure of Kalman filter and Smoother (forward filtering and backward smoothing) for an estimation of linear and gaussian state-space models. The second is that perhaps the first one implements the stochastic variable selection time-varying parameters autoregression with stochastic volatility on monetary policy analysis. We find that, more precisely than extant literature, there are no significant systematic shifts in monetary stance and economic agents' behavior due to exogenous monetary policy shock. The last but not the least is that we give an Bayesian statistic proof of the stability of small monetary VAR system. This is done by two steps. we first check variable significance over time due to proliferation of parameters in VAR system under limited data availability; second, we only focus on large breaks because generally, only the large ones have high potential to be candidates of structural change.[1] Since typical TVP-VAR is a type of model whose latent parameters vary every time, the accounting of small breaks may take over some strength from original true large breaks, which may cause misleading. We solve this problem with a modified TVP-VAR augmented with a discrete Markov process.

The paper is organized as follows. In the second section, we present three stochastic variable selection models step by step from static to time variation and from without stochastic volatility to with it. We give basic Bayesian estimation procedure for each model and more importantly, also explore their internal relations among these models; The third section discusses estimation

[1]Note that breaks are not necessarily equivalent to systematic changes. Systematic changes imply some 'deep parameters' have changed in DSGE models.

issues on stochastic volatility and stochastic variable selection as well as efficient implementation of computation; In section 4, we use these models to estimate and evaluate whether there were regime switches in monetary policy and structural changes in the U.S. economy, then give our findings and interpretation; Section 5 conducts robust check and the last section 6 concludes.

1.2 Models in stochastic variable selection

This section gives three models that are mutually relevant and become fledged step by step, modeling gradually close to economic reality. I present a basic description of each model. The first is stochastic variable selection time-invariant parameter model, namely stochastic variable selection Bayesian VAR for which I use SVS-BVAR for short. The second is stochastic variable selection partial time-varying parameter model, in which setup parameters either exist all time or not. I use SVS-Partial-TVP-VAR for short for this model. The last, newly developed in this chapter, is stochastic variable selection full time-varying parameter model where parameters either exist or not period by period. I call it SVS-Full-TVP-VAR. The three models can be imbeded with a stochastic volatility part, taking into account dynamics of exogenous shocks.

At the starting point, I should clarify some notations associated with these models.[2] A usual form of VAR should be transformed into a reduced form for convenient analysis. The VAR(p) of $p-$lag and M dimensionality with constant parameters can be written as:

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t \tag{1.1}$$

where y_t is an $M \times 1$ vector, c is an $M \times 1$ constant, $\{A_i\}_{i=1}^{p}$ are $M \times M$ matrices, and ε_t is an $M \times 1$ vector of shocks following normal distribution $\varepsilon_t \sim N(0, \Sigma)$, for $t = 1, \ldots, T$. The (1.1) can be transformed into a compact form

$$y_t = Z_t \beta + \varepsilon_t \tag{1.2}$$

[2]These notations apply to all the three chapters.

where $Z_t = I_M \otimes [1, y_{t-1}^{'}, \ldots y_{t-p}^{'}]$ and $\beta = vec([c, A_1, \ldots A_p]^{'})$, with $\otimes$ and *vec* Kronecker product and column stacking operator, respectively.

Above is just simple expression for writing convenience. If in terms of estimation, i.e. nesting all data together in big matrices, two forms are found. One is Canova (2007) and others; the other is, for example, Kadiyala and Karlsson (1997). The former arises if we use an $TM \times 1$ vector y which stack all T obsevations on the first dependent variables, then all T observations on the second dependent variables, etc., i.e. $y = [y_1^{1:T'}, \ldots, y_M^{1:T'}]'$ where $y_i^{1:T}$ denote observations of y_i in a column from time 1 to time T. The latter arises if we define Y to be $T \times M$ matrix which stacks the T observations on each dependent variable in columns next to one other, i.e. $Y = \left[y_1^{1:T}, \ldots, y_M^{1:T}\right]$. ε and E denote stackings of the errors in a manner conformable for y and Y, respectively. Define $x_t = [1, y_{t-1}^{'}, \ldots, y_{t-p}^{'}]^{'}$ and $X = [x_t, \ldots, x_T]^{'}$. Note that if we let $K = 1 + M \cdot p$ be the number of coefficients in each equation of the VAR, the X is a $T \times K$ matrix. Finally, if $A = [c, A_1, \ldots, A_p]^{'}$, we define $\beta = vec(A)$ that is a $K \cdot M \times 1$ vector which stacks all the VAR coefficients including the intercepts into a column vector. With all these definitions, we can write the VAR either as

$$Y = XA + E \text{ KK version} \tag{1.3}$$

or

$$y = (I_M \otimes X)\beta + \varepsilon \text{ Canova version} \tag{1.4}$$

where $y = vec(Y)$ from above definitions and $\varepsilon \sim N(0, \Sigma \otimes I_T)$.[3]

If parameters are time variant, the form of VAR accordingly becomes

$$y_t = c_t + A_{1,t} y_{t-1} + \cdots + A_{p,t} y_{t-p} + u_t \tag{1.5}$$

or compactly

$$y_t = Z_t \beta_t + u_t$$

[3]Equation (1.4) can be derived from equation (1.3) via $vec\,(ABC) = (C' \otimes A)\, vec\,(B)$.

where $Z_t = I_M \otimes \left[1, y'_{t-1}, \ldots, y'_{t-p}\right]$, $\beta_t = vec\left([c_t, A_{1,t}, \ldots, A_{p,t}]'\right)$ and $u_t \sim i.i.dN(0, \Sigma_t)$ for $t = 1, \ldots, T$. The stochastic volatility of Σ_t is introduced in the next subsection. Other details and modifications will be given in specific models.

Below we get into the stochastic variable selection models with notations defined above.

1.2.1 Stochastic variable selection Bayesian VAR

The VAR model in simple form of (1.2) can be written as

$$y_t = Z_t \theta + \varepsilon_t \tag{1.6}$$

where $\theta = \Gamma\beta$, $\Gamma = diag(\gamma) = diag([\gamma_1, \ldots, \gamma_{KM}]')$ and $\varepsilon_t \sim i.i.dN(0, \Sigma)$. We denote γ_j the j^{th} element of the vector γ, which is also j^{th} diagonal element of the matrix Γ, and γ_{-j} the vector γ with the j^{th} element removed. Priors for parameters in the model are set as

$$\beta \sim N_{MK}(\underline{\beta}, \underline{V})$$

$$\gamma_j | \gamma_{-j} \sim Bernoulli(1, \pi_{0,j})$$

$$\Sigma \sim IWishart(\underline{S}, \underline{\nu})$$

where $\underline{\beta}$, $\underline{V}$, $\pi_{0,j}$, $\underline{\nu}$ and $\underline{S}$ are hyperparameters set by researchers.
Given the priors above, the full conditional posteriors are

1. Sample β from the posterior density

$$\beta | \gamma, \beta, \Sigma, y, Z \sim N_{MK}\left(\bar{\beta}, \bar{V}\right)$$

where

$$\bar{V} = \left(\underline{V}^{-1} + \sum_{t=1}^{T} Z_t^{\star\prime} \Sigma^{-1} Z_t^{\star} \right)^{-1}$$

$$\bar{\beta} = \bar{V} \left(\underline{V}^{-1} \beta + \sum_{t=1}^{T} Z_t^{\star\prime} \Sigma^{-1} y_t \right)$$

and $Z_t^{\star} = Z_t \Gamma$.

2. Sample γ_j from the posterior density

$$\gamma_j | \gamma_{-j}, \beta, \Sigma, y, Z \sim Bernoulli\left(1, \bar{\pi}_j\right)$$

preferably in random order j, where $\bar{\pi}_j = \frac{l_{0j}}{l_{0j}+l_{1j}}$, with

$$l_{0j} = p(y|\theta_j, \gamma_{-j}, \gamma_j = 1)\pi_{0j}$$

and

$$l_{0j} = p(y|\theta_j, \gamma_{-j}, \gamma_j = 0)(1 - \pi_{0j})$$

The expression $p(y|\theta_j, \gamma_{-j}, \gamma_j = 1)$ and $p(y|\theta_j, \gamma_{-j}, \gamma_j = 0)$ are conditional likelihood expressions. Here we define $\theta^{\star}$ to be equal to θ but with the j^{th} element $\theta_j = \beta_j$ in the case of $\gamma_j = 1$. Similarly, we define $\theta^{\star\star}$ to be equal to θ but with the j^{th} element $\theta_j = 0$ when $\gamma_j = 0$. Then in terms of the likelihood of simple form of VAR for each period, namely equation (1.6), we can write l_{0j}, l_{1j} analytically as

$$l_{0j} = \exp(-\frac{1}{2} \sum_{t=1}^{T} (y_t - Z_t\theta^{\star})'\Sigma^{-1}(y_t - Z_t\theta^{\star}))\pi_{0j} \tag{1.7}$$

$$l_{1j} = \exp(-\frac{1}{2} \sum_{t=1}^{T} (y_t - Z_t\theta^{\star\star})'\Sigma^{-1}(y_t - Z_t\theta^{\star\star}))(1 - \pi_{0j}) \tag{1.8}$$

3. Sample Σ from the posterior density

$$\Sigma|\beta,\gamma,y,Z \sim IWishart(\bar{\nu},\bar{S})$$

where

$$\bar{\nu} = T + \underline{\nu}$$

and

$$\bar{S} = \underline{S} + \sum_{t=1}^{T} (y_t - Z_t\theta)(y_t - Z_t\theta)'$$

4. Go back to setp 1 again, start the next iteration.

This model is fundamental to the following model extensions in three aspects. Firstly, the essential part that determines the posterior probability of presence for each coefficient is the difference between the two conditional likelihood l_{0j} if the coefficient exist and the one l_{1j} if not. When l_{0j} dominate l_{1j}, the j^{th} coefficient shows up with high posterior possibility supported by data; Secondly, do not forget the role of prior probability π_{0j} for each coefficient β_j in that prior also give information to the posterior outcome. When it is assigned with very informative prior due to some economic theory, say, a very low value for π_{0j}, the information involved in the conditional likelihood l_{0j}, to some extent, will be weakened, while l_{1j} will be strengthened; Lastly, the Markov Chain Monte Carlo (MCMC) method based on this model is flexble to accomodate other modification, such as time-varying parameters, stochastic volatility widely existed in the time series.

1.2.2 Stochastic variable selection partial TVP-VAR

In this model, compared with the model SVS-BVAR above, the difference is that coefficients are time variant of the type as in Korobilis (2013b). This time variation as mentioned previously for some coefficient either exists for the whole time period or not; that is each indicator refers to a coefficient from period one to the last period, which is illustrated in details below. The model

specification is

$$y_t = Z_t \theta_t + \varepsilon_t \tag{1.9}$$

$$\beta_t = \beta_{t-1} + u_t \tag{1.10}$$

where $\theta_t = \Gamma\beta_t$, $\Gamma = daig(\gamma) = diag\left([\gamma_1, \ldots, \gamma_{KM}]'\right)$, $\varepsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, Q)$ which are uncorrelated with each other at all leads and lags. The priors for this models are

$$\beta_0 \sim N_{MK}\left(\underline{\beta}, \underline{V}\right)$$

$$\gamma_j | \gamma_{-j} \sim Bernoulli\left(1, \bar{\pi}_{0j}\right)$$

$$Q \sim IW\left(\underline{\xi}, \underline{R}\right)$$

$$\Sigma \sim IW\left(\underline{S}, \underline{\nu}\right)$$

Estimating these parameters means sampling sequentially from the following conditonal densityies

1. Sample $\beta_t | \beta_{t-1}, Q, \Sigma, y_t, Z_t^\star$ for all t, where $Z_t^\star = Z_t\Gamma$, using the Carter and Kohn (1994) filter and smoother for state-space models. For details on this, please refer to appendix of Primiceri (2005). This step, for computational efficiency, could be replaced by precision based algorithm of Chan and Jeliazkov (2009), taking full advantage of sparse matrix computation in commonly used maths and econometrics software.

2. Sample γ_j from the density

$$\gamma_j | \gamma_{-j}, \beta, \Sigma, y, Z \sim Bernoulli\left(1, \bar{\pi}_j\right)$$

preferably in random order j, where $\bar{\pi}_j = \frac{l_{0j}}{l_{0j}+l_{1j}}$, with

$$l_{0j} = p(y|\theta_j^{1:T}, \gamma_{-j}, \gamma_j = 1)\pi_{0j}$$

$$l_{0j} = p(y|\theta_j^{1:T}, \gamma_{-j}, \gamma_j = 0)(1-\pi_{0j})$$

The expression $p(y|\theta_j^{1:T}, \gamma_{-j}, \gamma_j = 1)$ and $p(y|\theta_j^{1:T}, \gamma_{-j}, \gamma_j = 0)$ are conditional likelihoods, where $\theta_j^{1:T} = \left[\theta_{1,j}, \ldots, \theta_{t,j}, \ldots, \theta_{T,j}\right]$. Define $\theta_t^{\star}$ to be equal to θ_t but with the j^{th} element $\theta_{t,j} = \beta_{t,j}$ when $\gamma_j = 1$ for all $t = 1, \ldots, T$. Similarly, we define $\theta_t^{\star\star}$ to be equal to θ_t but with j^{th} element $\theta_{t,j} = 0$, namely when $\gamma_j = 0$, for all $t = 1, \ldots, T$. Then in the case of TVP-VAR likelihood of the model, we can write l_{0j}, l_{1j} analytically as

$$l_{0j} = \exp(-\frac{1}{2}\sum_{t=1}^{T}(y_t - Z_t\theta_t^{\star})'\Sigma^{-1}(y_t - Z_t\theta_t^{\star}))\pi_{0j} \tag{1.11}$$

$$l_{1j} = \exp(-\frac{1}{2}\sum_{t=1}^{T}(y_t - Z_t\theta_t^{\star})'\Sigma^{-1}(y_t - Z_t\theta_t^{\star}))(1-\pi_{0j}) \tag{1.12}$$

3. Sample Q from the posterior density

$$Q|\beta, \gamma, \Sigma, y, Z \sim IW\left(\bar{\xi}, \bar{R}\right)$$

where

$$\bar{\xi} = T + \underline{\xi} - 1$$

and

$$R = \bar{R} + \sum_{t=2}^{T}\left(\beta_t - \beta_{t-1}\right)\left(\beta_t - \beta_{t-1}\right)'$$

4. Sample Σ from the posterior density

$$\Sigma|\beta, \gamma, y, Z \sim IWishart(\bar{\nu}, \bar{S})$$

where

$$\bar{\nu} = T + \underline{\nu}$$

and

$$\bar{S} = \underline{S} + \sum_{t=1}^{T} (y_t - Z_t\theta_t)(y_t - Z_t\theta_t)'$$

5. Go back to step 1, start a new iteration.

One thing to note from step 2 above is that the indicator γ_j associate with coefficient β_{jt} from $t = 1, \ldots, T$, i.e. Γ matrix is invariant over time. This model specification extending from constant coefficients to time varying coefficient may be good to forecast, but probably too restrictive for modeling time series dynamics as some coefficients may appear and disappear mutually during whole time, not simply keeping existing or not. A nature idea is to assign time variant indicator to each coefficient in a TVP-VAR model, to capture possible distinct dynamic feature for each coefficient. We will refer to model 3 below.

1.2.3 Stochastic variable selection full TVP-VAR

This model is based on the stochastic variable selection partial time-varying parameter VAR (Korobilis, 2013b), i.e. the second model above, and incorporates the multivariate stochastic volatility. The specification of the model has the same formula as equation (1.9) and equation (1.10) but with time dimension $\theta_t = \Gamma_t\beta_t$, $\Gamma_t = diag(\gamma_t) = diag\left([\gamma_{1,t}, \ldots, \gamma_{KM.t}]'\right)$ and $\varepsilon_t \sim N(0, \Sigma_t)$. $\gamma_{j,t}$ is indictor for corresponding $\beta_{j,t}$, either 0 or 1, meaning including and excluding the corresponding $\beta_{j.t}$ respectively, for $j = 1, \ldots, KM$ and $t = 1, \ldots, T$. $\gamma_{j,t}$ follows *Benoulli* distribution independently.

There are two differences in model 3, compared with model 2. The first is the indicator for each coefficient. The indicator γ now associate with not only identity j, but also with time t. That is, every coefficient in TVP-VAR now has its own indicator, unlike that $\beta_j^{1:T}$ share the same indicator γ_j in model 2; The second, error variacne covariance matrix Σ_t, displays time variant, i.e. stochastic volatility. The covariance matrix typically can be decompoed into the form of $\Sigma_t = A_t^{-1} D_t D_t A_t^{-1'}$,

where A_t is a lower triangular matrix with value of ones on the main diagonal

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{M1,t} & \cdots & a_{M(M-1),t} & 1 \end{bmatrix}$$

and D_t is the diagnoal matrix with elements $d_{i,t} = \exp\left(\frac{1}{2}h_{i,t}\right)$, for $i = 1,\ldots,M$. Here, the exponentialization makes values of diagonal elements always positive. Therefore D_t is the matrix

$$D_t = \begin{bmatrix} d_{1,t} & 0 & \cdots & 0 \\ 0 & d_{2,t} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_{M,t} \end{bmatrix} = \begin{bmatrix} e^{\frac{1}{2}h_{1,t}} & 0 & \cdots & 0 \\ 0 & e^{\frac{1}{2}h_{2,t}} & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & e^{\frac{1}{2}h_{M,t}} \end{bmatrix}$$

If we first stack the unrestricted elements of A_t below main diagonal by rows into a $\frac{M(M-1)}{2} \times 1$ vector as $a_t = \left[a_{21,t}, a_{31,t}, a_{32,t}, \ldots, a_{n(n-1)}\right]'$ for $n = 2,\ldots,M$, and $h_t = [h_{1,t},\ldots,h_{M,t}]'$, then β_t, a_t, and h_t follow independent random walks and the whole specification of the model in state space form is

$$y_t = Z_t\theta_t + A_t^{-1}D_t\upsilon_t \tag{1.13}$$

$$\beta_t = \beta_{t-1} + u_t$$

$$a_t = a_{t-1} + \varsigma_t$$

$$h_t = h_{t-1} + \eta_t$$

where $\upsilon_t \sim N(0, I_M)$ and $\theta_t = \Gamma_t\beta_t$.

The random walk setting presents the advantages of focusing on permanent shifts and reducing the number of parameters in the estimation procedure.

The innovations in the three state equations are

$$\begin{bmatrix} u_t \\ \varsigma_t \\ \eta_t \end{bmatrix} \sim i.i.dN \begin{bmatrix} Q & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & W \end{bmatrix}$$

where S can be block diagonal (Primiceri, 2005) or full matrix (De Wind and Gambetti, 2014).[4]

As for the sampling process for this model, details like Kalman filter and smoother are delegated to the technical appendix. Without loss of generality, I give the general expression for this model. It suffices to note two points. First , l_{0j} and l_{1j} in the above two models now change to l_{0jt} and l_{1jt} due to that the indicators $\gamma_{j,t}$ have been assigned to every coefficient in the TVP model over $j = 1,\ldots,KM$ and $t = 1,\ldots,T$. The conditional likelihood (approximate to) becomes:

$$l_{0js} = \pi_{0js} \exp\left\{ -\frac{1}{2}\left[\sum_{t\neq s}^{T} (y_t - Z_t\theta_t)' \Sigma_t^{-1} (y_t - Z_t\theta_t) + (y_s - Z_s\theta_s^{\star})' \Sigma_s^{-1} (y_s - Z_s\theta_s^{\star}) \right] \right\}$$

$$l_{1js} = \left(1 - \pi_{0js}\right) \exp\left\{ -\frac{1}{2}\left[\sum_{t\neq s}^{T} (y_t - Z_t\theta_t)' \Sigma_t^{-1} (y_t - Z_t\theta_t) + (y_s - Z_s\theta_s^{\star\star})' \Sigma_s^{-1} (y_s - Z_s\theta_s^{\star\star}) \right] \right\}$$

where $\theta_s^{\star}$ is defined to be equal to $\theta_s = \Gamma_s\beta_s$, but θ_{js} is equal to $\beta_{j,s}$, i.e. the corresponding indicator is $\gamma_{j,s} = 1$ for $s = 1,\ldots,T$ and for $j = 1,\ldots,KM$; Similarly, $\theta_s^{\star\star}$ is when $\theta_{j,s}$ is zero in the case of $\gamma_{j,s} = 0$ for $s = 1,\ldots,T$ and for $j = 1,\ldots,KM$. Then the probability for indicator $\gamma_{j,s} = 1$ in *Bernoulli* drawing is $\bar{\pi}_{j,s} = \frac{l_{0js}}{l_{0js}+l_{1js}}$. Note that the identity j and the time s are both randomly picked up.[5] Second, it is about the prior setting. Generally, there are two choices. Usually, the priors on the time-varying parameters are:

$$\beta \sim N\left(0, 4I_{n\beta}\right)$$

[4]De Wind and Gambetti (2014) prove that standard Kalman filter can still be used in the second state space model for a_t instead of equation by equation causing block diagonal in S as in Primiceri (2005).

[5]An efficient estimation is developed in appendix following Chan and Jeliazkov (2009).

$$l \sim N(0, 4I_{nl})$$

$$h \sim N(0, 4I_{nh})$$

The subscript presents dimensions of each parameter. The priors on their error covariacne are:

$$Q \sim IW\left(1+n_\beta, \left((k_Q)^2 (1+n_\beta) I_{n\beta}\right)\right)$$

$$S \sim IW\left(1+n_l, \left((k_S)^2 (1+n_l) I_{nl}\right)\right)$$

$$W \sim IW\left(1+n_h, \left((k_W)^2 (1+n_h) I_{nh}\right)\right)$$

where the hyperparamets are set to $k_Q = 0.01$, $k_S = 0.1$, $k_W = 0.01$. One can also construct the priors using a training sample (Primiceri, 2005). In particular, assume that $\hat{\theta}_{ols}$ and $V\left(\hat{\theta}_{ols}\right)$ are the mean and variance respectively of the OLS estimate of $\Theta = (\beta, l, h)$ based on a VAR with constant parameters using an initial training sample.[6] Then the priors can be written as

$$\beta \sim N(\beta_{ols}, 4V(\beta_{ols}))$$

$$l \sim N(l_{ols}, 4V(l_{ols}))$$

$$h \sim N(h_{ols}, 4V(h_{ols}))$$

and errors of covariance matrices are

[6]The priors can also be constructed via Bayesian estimation with noninformative priors.

$$Q \sim IW\left(1+n_{\beta}, \left(\left(k_Q\right)^2\left(1+n_{\beta}\right)V\left(\beta_{ols}\right)\right)\right)$$

$$S \sim IW\left(1+n_{l}, \left(\left(k_S\right)^2\left(1+n_{l}\right)V\left(l_{ols}\right)\right)\right)$$

$$W \sim IW\left(1+n_{h}, \left(\left(k_W\right)^2\left(1+n_{h}\right)V\left(h_{ols}\right)\right)\right)$$

Other notations are the same as above.

One thing should be mentioned again is that the model 3 presented here is the general case. This means that modification or restriction can be applied to the general case. If the investigation of parameter significance via $\gamma_{j,t}$ in model 3 is restricted to γ_j and without stochastic volatility, then model 3 is reduced to model 2. If further shutting off time variations on coefficients, it is model 1.

On the other hand, as we discussed stochastic search method is associated with mixture models, the possible choices will become proliferative when the target model involves many coefficients, especially in TVP models (over identity and over time). This will give rise to great computational burden. Therefore, in order to reduce the burden, an indicator that controls for a set of coefficients could be used instead of one for one. Actually, the model 2 has used the idea that each indicator γ_j corresponds to β_j over the whole period, namely $\beta_j^{1:T}$. I call it block checking. The property of indicators over each time and each identity in model 3 makes the investigation of any single coefficient (theoretically) possible if pursuing the best model, but causes dramatically increased computational cost in practice such that sometimes the estimation is infeasible. This is single checking. Hence, efficient estimation of stochastic variable selection model is required. I will discuss it in the next section.

1.3 Bayesian estimation

In this section, we focus on two points that are important for estimation and numerical computation. One is on stochastic volatility; The other is on stochastic variable selection.

1.3.1 Stochastic volatility

Stochastic volatility is a method modeling dynamic process for error term. Once stochastic volatility (SV afterwards) is used, generally, the econometric model involves a nonlinear or/and non-gaussian state space part. Jacquier, Polson and Rossi (1994) propose to, based on Carlin et al.(1992), use Metropolis step via a single move to draw stochastic volatility; here, we follows Primiceri (2005), using mixture normal distribution to approximate irregular distribution (specifically, $log\left(x^2\left(1\right)\right)$), suggested by Kim, Shephard and Chib (1998).

The method involves two steps. Firstly, we draw s^T from $f\left(s^T|y^T,\Xi,h^T\right)$, where s^T demonstrate which normal is chosen among mixture normal distributions each period; y^T is all the observation from y_1 to y_T, the same for log volatility states h^T and Ξ represents other parameters and hyperparameters. This conditional posterior is discrete distribution with seven points or ten points as support, see Kim et al. (1998) for seven normal distributons and Omori et al. (2007) further for ten, respectively; Secondly, draw h^T from $f\left(h^T|y^T,\Xi,s^T\right)$ given all s_1 to s_T from previous step. This step is very important, as it transforms non-gaussian to a given gaussian based on identity s_t each period. Hence, (nonlinear) non-gaussian state space model has been 'cut' into linear and gaussian state space model each period and naturally, multi-move Gibbs sampling of Carter and Kohn (1994) can be used, which makes chain mix well quickly and therefore reduce computational time, as opposed to single move of Jacquier et al. (1994).[7]

Details on this SV approximation method can be found in appendix. Some issues such as why draw identity s^T as first step and comparison with other methods dealing with SV, reader of interest are referred to Del Negro and Primiceri (2013), who confirm that performance of this method is

[7] In Chapter 2, the single move of Jacquier et al. (1994) is used for nonliner state space model.

quite good and quantitatively same as the results of methods exactly modeling SV.

1.3.2 Stochastic selection

In section 2, we have given the mechanism of stochastic selection (SS hereafter) in (TVP) VAR models and sketch the steps how to draw posteriors from full conditional distributions. In this subsection, we focus on implementation of this stochastic selection in computation, especially for model 3.

As we discussed in model 3, in section 2, compared with model 1 and model 2, all models share the essentially same rationale for SS, but computation burden dramatically increases from model 1 to model 3. Take model 3 for example. After incorporating SV and indicator for each and every coefficient (over identity and over time) in the TVP-VAR model, the model becomes very complicated though it allows for all possibilities. Considering a small TVP-VAR, typically with three endogenous variables and only two lags – usually used in monetary VAR and requiring more endogenous variables and lags such as oil analyses in Baumeiser and Peersman (2010) – and with sample period usually covering around $T = 150$ quarterly data set, we need to compute $2^{\wedge(21*150)}$ models only for stochastic selection part, not including SV, for just one iteration in the Bayesian estimation circle.

To make the implementation of SS feasible, two ways can be used. One is block checking, namely, the selection is based on a set of variables rather than single variable such as model 2. Apparently, this method can reduce the number of candidate models, but with the concern of missing potential better choices. Single checking is required if the best model is preferred in some context. This suggests the second way – an efficient computational method that is prepared for the general single checking. We use the efficient simulation method of Chan and Jeliazkov (2009). The basic idea is that we nest potential iterations in each Bayesian circle into a quiet large and sparse matrix to reduce number of iterations and therefore to speed up computation.

For example, we can construct large sparse matrices for a linear normal state space model without using typical steps of Kalman forward and backward recursion and directly obtain a draw of

$\beta^{1:T} = [\beta_1', \ldots, \beta_t', \cdots \beta_T']'$ at a time from a posterior conditional distribution. Thus it is the transformation from conventional estimation of TVP-VAR to estimation of constant parameter VAR via large and sparse matrices. On the other hand for SS portion, we also construct large sparse matrix for calculating conditional likelihood for model selection instead of iterating over whole time period. That is, corresponding to the draw of $\beta^{1:T}$, a large diagonal matrix $\Gamma = Blkdiag\left(\{\Gamma_t\}_{t=1}^{T}\right)$ is used where $\Gamma_t = diag\left([\gamma_{1,t}, \ldots, \gamma_{MK,t}]'\right)$. After this improvement of implementation in computation, we find we can cut off two thirds of time originally used for not doing this due to taking the advantage of matrix computation faster than iteration and large sparse matrix actually accounting for small space to save. For details, please refer to appendix for this chapter.

1.3.3 Exercise

Why we need to use stochastic variable selection models to do empirical analysis? Make the models parsimonious and close to reality. Here, I use an artificial data set to show this advantage of this approach by a simple exercise. For simplicity, I generate a 5 variable VAR with only $T = 50$ time series observations, one lag and no constant. The unrestricted OLS estimates perform poorly, while variable selection gives better estimates. This is because we restrict irrelevant variables, and then there are more degrees of freedom to estimate the parameters on the relevant variables.

Suppose that the coefficient matrix in the version of Kadiyala and Karlsson (1997) mentioned above is an identity matrix and that error covariance matrix is randomly chosen only positive definite for sure. Initial data y_1 is from a standard uniform distribution. Let's look at table 1.1. The first column contains posterior probatility in mean when $\gamma_j = 1$, namely, the probability of coefficient β_j in the VAR(1) model. Posterior mean of β_j is in column 2. Column 3 contains coefficients estimated by ordinary least square method (OLS). The true parameters are in the last column. The prior probability π_{0j} for each parameter β_j is set to 0.5. That's because one has no idea of including or excluding each parameter, say, perhaps in or out in the model half by half. Hence, Korobilis (2013b) regards it as 'noninformative prior' and criteria probability. Column 3, in contrast to last column – true values of coefficients – demonstrates that the estimation performance

is bad and many inefficient variables contained in. Column 1 gives posterior probability for each coefficient that match well with true value: when true value is 1, all posterior probability is 1; while true value is 0, almost all the corresponding posterior probabilities are extremely low, seldom above 5% contrast with 50% as criteria. The corresponding mean of each coefficient in column 2, compared with column 3, generally, is more close to true value because the indicators give more information to coefficients that exist. Obviously, the exercise shows great benefit on Bayesian VAR regression.

1.4 Monetary policy in stochastic selection models

1.4.1 Data description

I use a quaterly U.S. data set on the inflation rate, unemployment rate and the short interest rate. Inflation is the annual percentage change in a chain-weighted GDP price index; unemployment rate is seasonlly adjusted civilian unemployment, all workers over age 16; short rate is 3−month Treasury bill rate. They are denoted by $y_t = (\pi_t, u_t, r_t)'$ collected in a vector. The sample runs from 1953Q1 to 2006Q3. We choose this sample period due to two reasons. One is that this time span, covering Great Inflation, Monetary Targeting, and Great Moderation, is widely used in the monetary policy literature, especially for those focusing on issue of monetary policy regimes switching during this possible policy change time period. Therefore it is nature for us to choose the same span to compare our findings and results with previous ones. The other is that after recent global economic and financial crisis of 2008, though the Great Recession has gone but still in the process of slow recovery, the Fed has switched to and continued near zero rate policy until full recovery, which apparently an abrupt change in monetary instrument.

These three variables we choose in our analysis are commonly used in New Keynesian VAR literature. They are simple but representative. Examples of papers which use these, or similar variables, include Cogley and Sargent (2001, 2005), Koop, Leon-Gonzalez and Strachan (2009) and Primiceri (2005).

Restrict	Bayesian	OLS	True
1	0.984388	0.668605	1
0.025	-0.02115	-0.26653	0
0.0268	0.017219	-0.3257	0
0.348	0.063925	-0.3257	0
0.0288	0.01453	-0.1153	0
0.0925	-0.27718	-0.40878	0
1	0.791886	0.671789	1
0.028	0.015616	0.024751	0
0.0412	-0.03159	-0.03933	0
0.0326	0.042137	-0.07599	0
0.0158	0.05577	0.069901	0
0.017	0.073861	0.117344	0
1	0.787047	0.753396	1
0.212	-0.00324	-0.0383	0
0.0166	-0.03478	0.116155	0
0.163	-0.00202	0.09937	0
0.0474	0.033059	-0.02569	0
0.058	-0.02369	-0.00919	0
1	0.8962	0.92836	1
0.019	0.091339	-0.11235	0
0.0564	0.098384	0.350941	0
0.027	-0.02018	0.24663	0
0.0336	-0.01635	0.340174	0
0.0176	0.049076	0.187969	0
1	0.856714	0.90082	1

Table 1.1: Exercise result

Notes: The first column contains posterior probability in mean when $\gamma_j = 1$, namely, the probability of coefficient β_j in the VAR(1) model; The second column has posterior mean of β_j; The third column is OLS estimate of β_j; The fourth column is for true value of β_j that is used to construct the artificial data.

1.4.2 Lags and significance

In this subsection, I discuss this issue focusing on Model 1, namely static SVS-BVAR. Though its simplicity, it can specify this issue very well and shed light on the other two models.

Due to the property of stochastic selection models we discussed in section 2, this class of models itself is able to freely choose what coefficients enter or exit the system. In this way, if arbitrarily choose observables and lags, letting the data speak, in a Bayesian framework, one can finally find the best model associated with its coefficient significance and lags.

I first use model 1 (SVS-BVAR) to estimate the full range of the data set. Inflation, unemployment rate and 3-month treasury bill rate are typically used in small scale monetary VAR. Of course, more variables such as real GDP growth, money base, exchange rate and so on can also be nested into models and let the data decide if they are significantly interact each other and over time following the idea. Before this paper, to my best knowledge, Canova and Ganbetti (2009), Cogley and Sargent (2001, 2005), Koop et al. (2009) and Primiceri (2005) all directly give two lags without any check. The advantage of SVS model is that we can find the significant variables meanwhile the best lag. I first set the uninformative prior probability $\pi_{0j} = 0.5$, meaning the fair change of including or exluding the corresponding coefficient. The possible number of lags, say four lags, needs to be finally checked by the data.

In Table 1.2, the second column contains posterior probabilities of lagged coefficients for inflation. The third column is for unemployment, the last is for interest rate. The bold font probabilities are much higher than 50%, which implies the corresponding lagged variables should stay in this VAR model. We can also find after 3 lags, no significant dependents left. It says that the right lag may be 3. But a question arises from the prior choice. In table 1.2, I choose 0.5 as a prior for indicators, which perhaps lower the probability of including some variables, making some variables left the table that might stay in. To avoid this possibility, I gradually increase the prior from 0.5 to 0.75 by an increase of 5% step by step. Except that the short rate of lag 3 in policy equation some time not very significant but very close to criteria probability, other variables from table 1.3 to table 1.7 are very stable. Then, the second question comes out that I set prior for each variable

$\pi_{0j} = 0.5$	Lag = 4		
	π_t	u_t	r_t
c	0.171889	**0.9844**	0.090889
π_{t-1}	**1**	0.017244	**0.994911**
u_{t-1}	0.072044	**1**	**0.984111**
r_{t-1}	0.007156	0.016889	**1**
π_{t-2}	**1**	0.018111	**0.7271556**
u_{t-2}	0.034044	**1**	**0.975822**
r_{t-2}	0.006844	0.062489	0.032867
π_{t-3}	0.030333	0.015756	0.111333
u_{t-3}	0.042244	0.035956	0.083067
r_{t-3}	0.003689	**0.86116**	**0.692844**
π_{t-4}	0.014622	0.054178	0.173044
u_{t-4}	0.026467	0.0164	0.113711
r_{t-4}	0.005711	0.054644	0.035489

Table 1.2: Lag check1

the same every time. This may not be the case in reality. Therefore, I use standard uniform distribution for each prior. Table 1.8 shows the result is the same only with that interest rate of lag 3 for unemployment equation disappears. Actually, uniform distribution probably potentially can lower or raise prior for some variables and this may make the result controversial. Hence, relatively speaking prior of 50% is an appropriate choice. We also find that the third lag of interest rate is very isolated and this may be caused by more lags I choose. When lags are reduced to three, see table 1.9, significant variables are the same as previous results within four lags. When reduced to lag of two, the result is the same as those with more lags. Altogether, coefficients of significance found in table 1.2 is not prior sensitive and coefficients on fourth lag always keep insignificant. Now, I can say, lag of 3 is a good choice for the static VAR model using this data set.

Following the same spirit, we can also find the significance and therefore choose the best lags. Table 1.11 lists the significance of coefficients for model 2 (SVS-Partial-TVP-VAR).[8] Note that value in each cell corresponds to the significance of the coefficients over the whole sample period as a whole, i.e., $\beta_j^{1:T}$, while Fig 1.14 to Fig 1.16 for model 3 (SVS-Full-TVP-VAR) gives the significance of each coefficient in each equation, $\beta_{j,t}$, for $j = 1, \ldots, KM$ and for $t = 1, \ldots, T$, indi-

[8] Acutully for model 2, we find three lags are best choice. However, for convenience and consistence of comparison of results in the related literature, we choose two lags usually seen for quarterly monetary small VAR.

	$\pi_{0j} = 0.55$		
	π_t	u_t	r_t
c	0.273067	0.98576	0.120022
π_{t-1}	1	0.024378	0.99062
u_{t-1}	0.209089	1	0.78644
r_{t-1}	0.005044	0.0668	1
π_{t-2}	1	0.041689	0.78473
u_{t-2}	0.163756	1	0.78473
r_{t-2}	0.007133	0.040356	0.046889
π_{t-3}	0.0396	0.045089	0.180644
u_{t-3}	0.039133	0.055067	0.083978
r_{t-3}	0.004489	0.817	0.547133
π_{t-4}	0.014733	0.057511	0.152489
u_{t-4}	0.030556	0.0406	0.122422
r_{t-4}	0.004244	0.07311	0.037556

Table 1.3: Lag check2

	$\pi_{0j} = 0.6$		
	π_t	u_t	r_t
c	0.299644	0.99658	0.128133
π_{t-1}	1	0.017156	0.98989
u_{t-1}	0.202867	1	0.951
r_{t-1}	0.009511	0.049978	1
π_{t-2}	1	0.0162	0.7812
u_{t-2}	0.086553	0.9672	0.9286
r_{t-2}	0.015333	0.0338	0.046511
π_{t-3}	0.046911	0.067889	0.180244
u_{t-3}	0.089156	0.094422	0.112422
r_{t-3}	0.0096	0.83849	0.78979
π_{t-4}	0.02644	0.0746	0.217378
u_{t-4}	0.053822	0.030222	0.140511
r_{t-4}	0.011711	0.071578	0.040956

Table 1.4: Lag check3

	$\pi_{0j} = 0.65$		
	π_t	u_t	r_t
c	0.4738	0.99347	0.160111
π_{t-1}	1	0.022978	0.98916
u_{t-1}	0.417044	1	0.9962
r_{t-1}	0.010889	0.073869	1
π_{t-2}	1	0.034956	0.79469
u_{t-2}	0.241578	1	0.9668
r_{t-2}	0.013289	0.068956	0.058978
π_{t-3}	0.057911	0.043311	0.212667
u_{t-3}	0.098689	0.074711	0.198711
r_{t-3}	0.007244	0.8608	0.68487
π_{t-4}	0.027556	0.042622	0.204711
u_{t-4}	0.097222	0.037644	0.2132
r_{t-4}	0.008511	0.068667	0.073844

Table 1.5: Lag check4

	$\pi_{0j} = 0.7$		
	π_t	u_t	r_t
c	0.674244	0.98831	0.220867
π_{t-1}	1	0.03444	0.97342
u_{t-1}	*0.72776*	1	0.78644
r_{t-1}	0.010711	0.080044	1
π_{t-2}	1	0.038022	0.7376
u_{t-2}	0.504911	1	0.92064
r_{t-2}	0.010889	0.104978	0.07422
π_{t-3}	0.052822	0.041067	0.222467
u_{t-3}	0.149333	0.061311	0.210133
r_{t-3}	0.008222	0.88473	0.74669
π_{t-4}	0.035867	0.040178	0.230178
u_{t-4}	0.181156	0.030689	0.287578
r_{t-4}	0.009378	0.0838	0.060111

Table 1.6: Lag check5

	$\pi_{0j} = 0.75$		
	π_t	u_t	r_t
c	0.676533	0.99118	0.238756
π_{t-1}	1	0.0366	0.99551
u_{t-1}	0.6782	1	0.98373
r_{t-1}	0.018956	0.101644	1
π_{t-2}	1	0.54556	0.78229
u_{t-2}	0.344467	0.9658	0.9364
r_{t-2}	0.014333	0.085689	0.105933
π_{t-3}	0.067822	0.063178	0.344244
u_{t-3}	0.241533	0.126867	0.250044
r_{t-3}	0.0118	0.85351	*0.746289*
π_{t-4}	0.034489	0.0628	0.302933
u_{t-4}	0.159533	0.063089	0.346
r_{t-4}	0.009822	0.092244	0.096644

Table 1.7: Lag check6

	$\pi_{0j} \sim U(0,1)$		
	π_t	u_t	r_t
c	0.020733	0.99653	0.013222
π_{t-1}	1	0.009	0.96131
u_{t-1}	0.037489	1	0.96469
r_{t-1}	0.000956	0.023689	1
π_{t-2}	1	0.062667	0.63662
u_{t-2}	0.035689	0.94331	0.84693
r_{t-2}	0.014289	0.111756	0.000778
π_{t-3}	0.000956	0.036289	0.349956
u_{t-3}	0.001578	0.9794	0.174378
r_{t-3}	0.002978	0.151089	0.8954
π_{t-4}	0.025911	0.529983	0.424889
u_{t-4}	0.6608	0.279311	0.117289
r_{t-4}	0.001289	0.263667	0.059067

Table 1.8: Lag check7

	$\pi_{0j} = 0.5$		
	π_t	u_t	r_t
c	0.1501	0.999	0.0994
π_{t-1}	1	0.0099	0.9816
u_{t-1}	0.0094	1	1
r_{t-1}	0.0031	0.1093	1
π_{t-2}	1	0.0074	0.7273
u_{t-2}	0.0082	1	1
r_{t-2}	0.0042	0.116	0.0323
π_{t-3}	0.0314	0.0105	0.2294
u_{t-3}	0.0082	0.0145	0.0528
r_{t-3}	0.0022	0.8137	0.8235

Table 1.9: Lag check8

	$\pi_{0j} = 0.5$		
	π_t	u_t	r_t
c	0.1152	0.9789	0.0984
π_{t-1}	1	0.1765	0.9918
u_{t-1}	0.0124	1	1
r_{t-1}	0.0023	0.1378	1
π_{t-2}	1	0.149	0.6464
u_{t-2}	0.0096	1	1
r_{t-2}	0.0062	0.5454	0.0872

Table 1.10: Lag check9

	$\pi_{0j} = 0.5$		
	π_t	u_t	r_t
c_t	1	0.4620	0.9550
π_{t-1}	1	0.2170	0.9460
u_{t-1}	1	1	1
r_{t-1}	0.3250	0.0370	1
π_{t-2}	1	0.8810	0.7050
u_{t-2}	1	1	1
r_{t-2}	0.3430	0	1

Table 1.11: Lags in model 2

vidually. We shall come back to them in the next subsections when discussing possible structrual changes in agents' behavior and monetary policy implementation.

1.4.3 Impulse responses

Next, let's check impulse responses. First, we compare impulse responses of unrestriced Bayesian VAR with static SVS-BVAR of model 1. The unrestricted Bayesian VAR is equivalent to setting prior that all parameters are in the model with probability of one in terms of SVS-BVAR. All parameters in SVS-BVAR are set with prior probability of 50%.

Recursive identification of exogenous monetary policy shocks is used in which short rate is placed in the last order such that inflation and unemployment rate can impact policy immediately while policy rate affect them with one lag. The size of monetary shock is normalized to one percentage.

The result shows that in the first several periods, for unrestriced model, responses of unemployment rate get down, which contracts with economic theory and unsurprisingly, typically seen in small scale monetary VAR literature. However, when stochastic selection is imposed, it improves completely, getting rise directly after lags of 3 periods due to the inffluence of irrelevant variables alleviated by stochastic selection. Interest rate response to inflation in unrestricted model is modest, after around 5 periods starting above 1, then gradually get down to zero around the 21st period. Responses of interest rate in model with stochastic selection, are quite strong, until after 24 periods, it still stays close to 2 at median, which means in a long run respect Taylor rule is very

powerful that's different with the result for unrestricted model.[9]

Now let's look at impulse responses (IRs hereafter) of variables to monetary policy shocks in model 2 and model 3. Both models are imposed with multivariate stochastic volatility in order to jointly analyze systematic and non-systematic changes. Time variation of coefficients on regressors and stochastic volatility have extended the IRs with dynamic property over time as opposed to static model 1.

We randomly choose three period 1975 Q1, 1981 Q3 and 1996 Q1 to represent three chairmanships respectively. Figure 1.2, corresponding to model 3, shows the impulse response functions of inflation, unemployment and three month short rate to one percentage monetary policy shock, with solid line representing median, dashed lines 16^{th} and 84^{th} percentile respectively of posterior IR distribution. All the responses perfectly satisfy economic theory. There are no price puzzle typically seen in a small scale monetary VAR which can be weakened after adding forward looking prices or using large data set (Bernanke et al., 2005); Unemployment rates also increase quickly under contracting monetary policy. All the responses under recursive identification are in line with that under agnostic identification of sign restriction of Uhlig (2005). If comparing the same responses with that under model 2, in Figure 1.1, and unrestricted TVP-VAR, we find that there are 'price puzzle' and 'unemployment abnormal', though no longer significant in model 2 – the similar improvement in rich data set model such as factor augmented models, but seldom completely diminishing. This means that even though in a small scale VAR model, however, with coefficient restrictions, we can still obtain theory consistent results that are difficult without restriction. On the other hand, we also find error bands of IRs for model 3 in Figure 1.2 have substantially narrowed as opposed to those of IRs for model 2 in Figure 1.1. A reasonable explanation for this is that model 3 has restrictions both over identity and time, while model 2 not allowed for time dimension.

[9]Since the static typical small scale VAR is widely analysed, we do not provide figues for the save of space.

1.4.4 Systematic or exogenous change

An important thing we need to consider in monetary policy issue is that there might be structural changes in monetary policy with different chairmanships of Burns, Volcker and Greenspan, though some believe, some not. Related literature has given large number of evidence about this issue, though until now such dispute still exists there in theoretical and empirical study and this one sometimes more or less associated with other topics like sources or causes of the Great Inflation or the Great Moderation. One thing we need to confirm is that generally there are two periods suitable for monetary policy test. One is from 1963 Q1 to 1973 Q3 corresponding approximately to the period of rising inflation before the Volcker chairmanship. The period 1982 Q4 to 2006 Q3 corresponds to the Volcker and Greenspan chairmanships excluding the years of Monetary Targeting, for which the Taylor rule might not represent an appropriate description of systematic monetary policy (see, for instance, Hanson, 2006; Sims and Zha, 2006).

Recently, time varying parameter VAR with stochastic volatility (see, among many other, Cogley and Sargent, 2005; Cogley, Primiceri, and Sargent, 2010; Koop, Leon-Gonzalez, and Strachan, 2009; Koop and Korobilis, 2013) has been widely used in empirical analysis in business cycle, policy change, forecast and so on, because this kind of models can capture time variation properties of coefficient and volatility that probably reflect structural changes in a gradual manner rather than abrupt one like Markov regime switch model with probability or threshold regime switch model when threshold variable is above or below threshold value. On the other hand, SV can investigate some potential exogenous or non-systematic changes during sample period that is not capable for model without SV. However, these TVP models all associate with parameter proliferation problem due to the extension of parameters to time dimension even if in a small scale VAR, which might give incorrect properties and possible wrong inferences. Therefore, TVP-VAR with stochastic variable selection is a good tool to analyze such problems relevant with systematic changes via time varying coefficients and non-systematic changes via stochastic volatility.

1.4.4.1 Agents bahavior to monetary shocks

We investigate whether structure has changed in the agents' behavior to monetary policy shocks. The magnitude of monetary policy is standardized to one percentage in each period. We compare the difference among these three periods mentioned above.

Intuitively, in Figure 1.1 and Figure 1.2, IRs of inflation and unemployment did not seem change much. However, we need a way to precisely estimate the difference. Following Primiceri (2005) and others, we compute the difference of IRs between every two periods mutually among the three periods in every iteration after burin-in during Bayesian estimation and therefore obtain the posterior distribution of the IR difference .

Figure 1.3 and Figure 1.4, for model 3, show that there are no significant differences between these periods no matter for response of inflation or unemployment, because they are all insignificant with zero line. This means that economic agents have not altered their behavior and hence no structure changes in the agents in response to non systematic monetary policy shocks. The same results can also be found in model 2, see Figure 1.5 and Figure 1.6, and in unrestricted TVP-VAR, but with larger error band compared with model 3 for the reason we have discussed before.[10]

Since arbitrarily picking up three periods, we can not guarantee that the 'insignificant change in agent behavior' always hold during the whole sample period. The 3-D impulse responses of inflation and unemployment rate over the whole period respectively, in Figure 1.8 for model 3, are given, showing that it is reasonable to believe that economic agents keep the same way in decision making. Figure 1.7 for model 2 also support this conclusion with impuse responses appearing more smooth.

1.4.4.2 Systematic monetary policy

We check if the implementation of monetary policy has changed. Long run response of policy rate to inflation shocks and unemployment shocks is used to represent monetary policy stance.[11]

[10]For saving space, we do not give figures for unrestricted TVP-VAR from which they deliver the results in line with the literature.

[11]In standard New Keynesian DSGE models, the monetary policy rule follows this similar type:

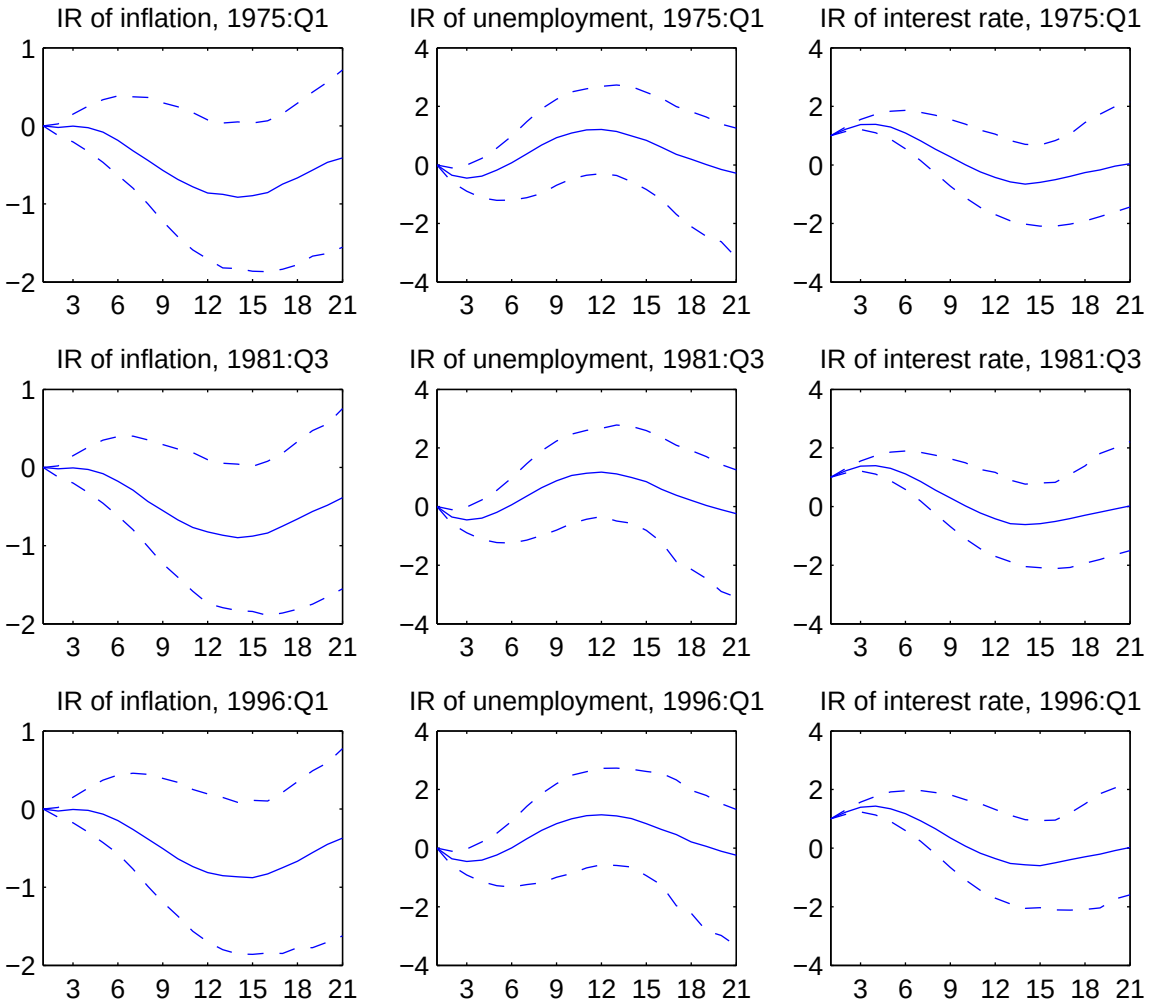


Figure 1.1: IRs for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The first column plots impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The second column plots impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The third column plots impulse responses of interst rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. Solid line represents posterior median of impluse response distribution with dashed lines of 16^{th} percentile and 84^{th} percentile, respectively.

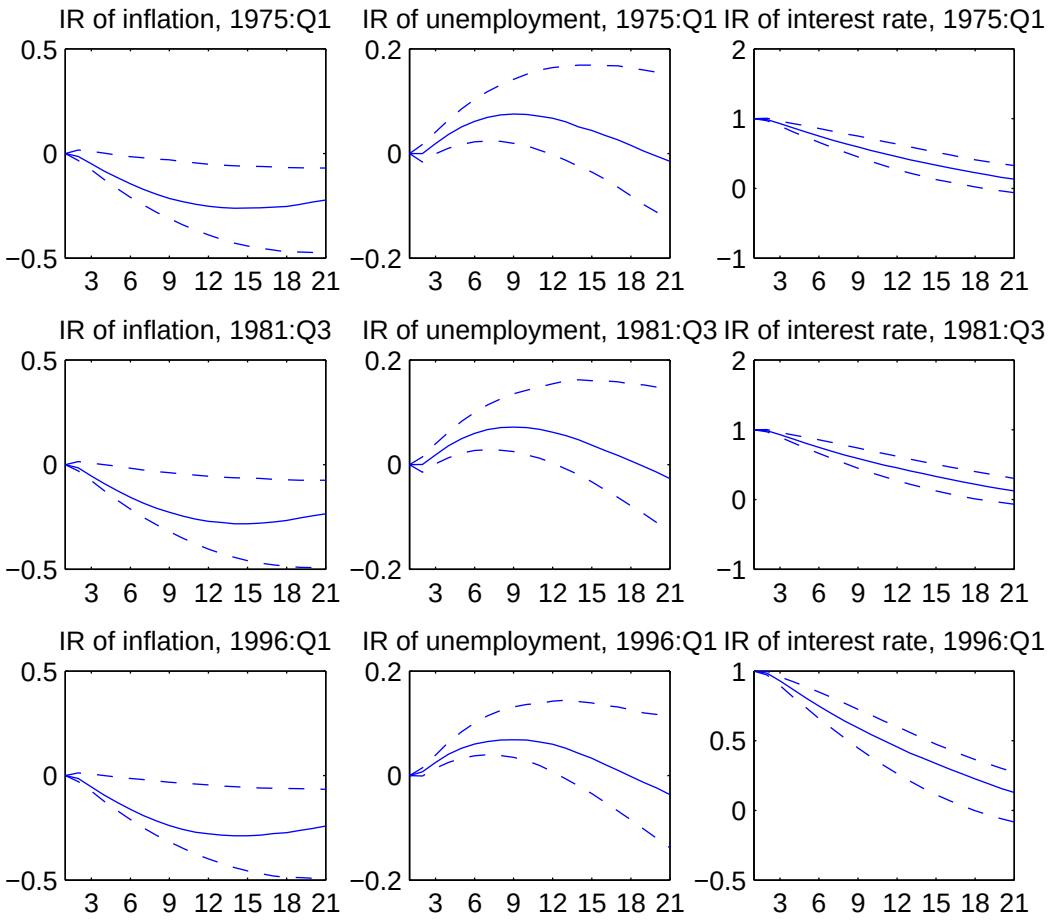


Figure 1.2: IRs for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The first column plots impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The second column plots impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The third column plots impulse responses of interst rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. Solid line represents posterior median of impluse response distribution with dashed lines of 16^{th} percentile and 84^{th} percentile, respectively.

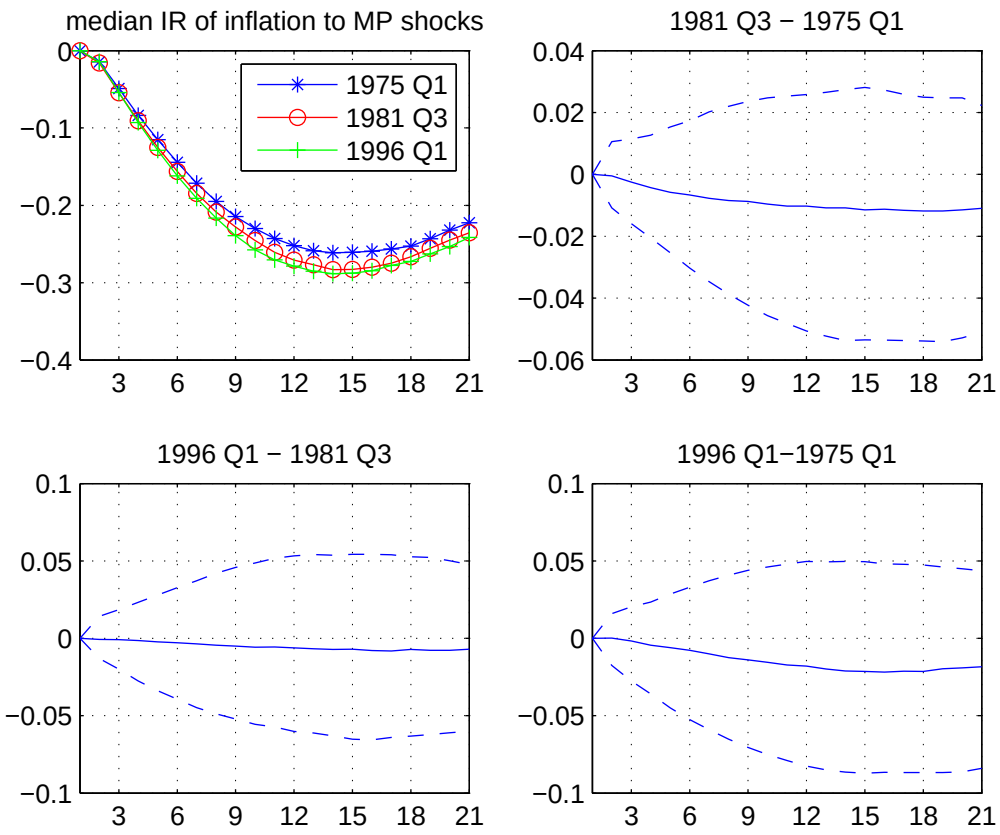


Figure 1.3: IR-inflation-comparison for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

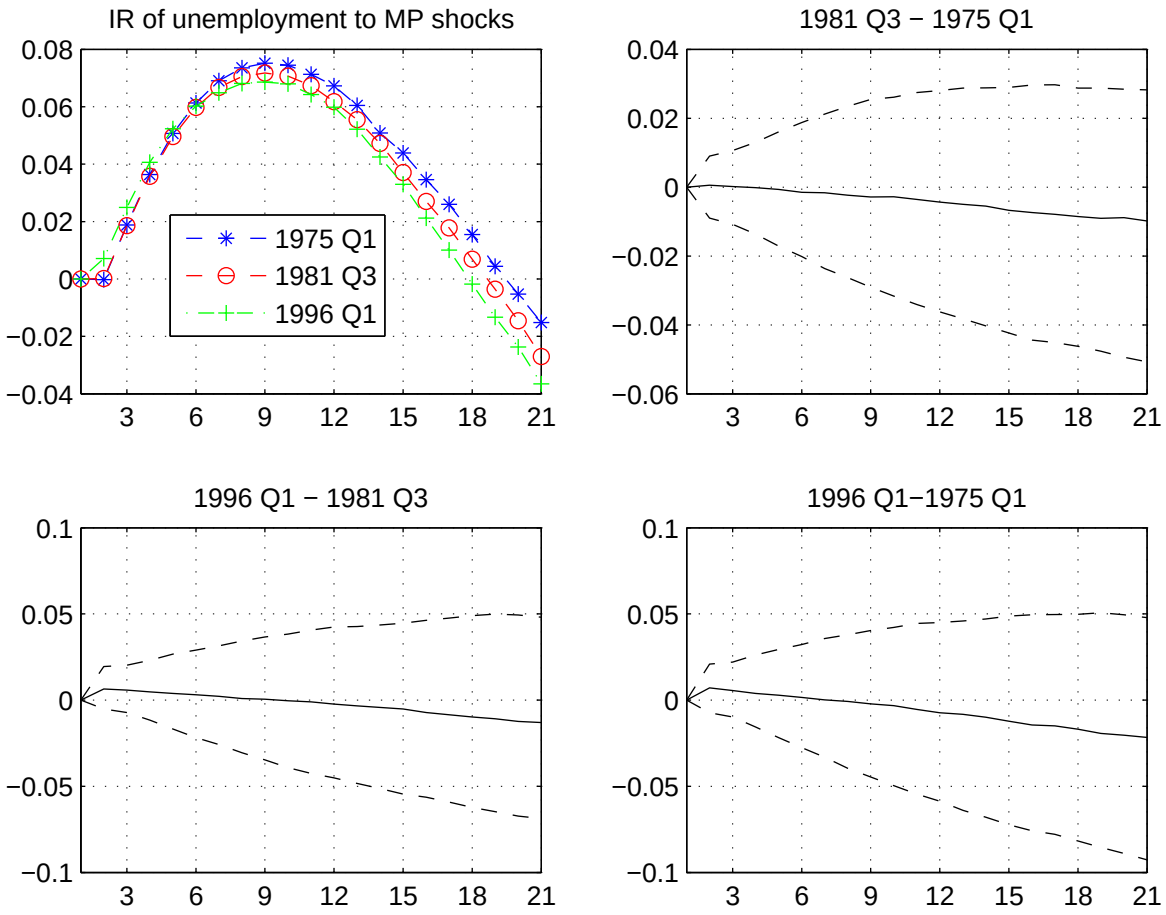


Figure 1.4: IRs-unemployment rate-comparison for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

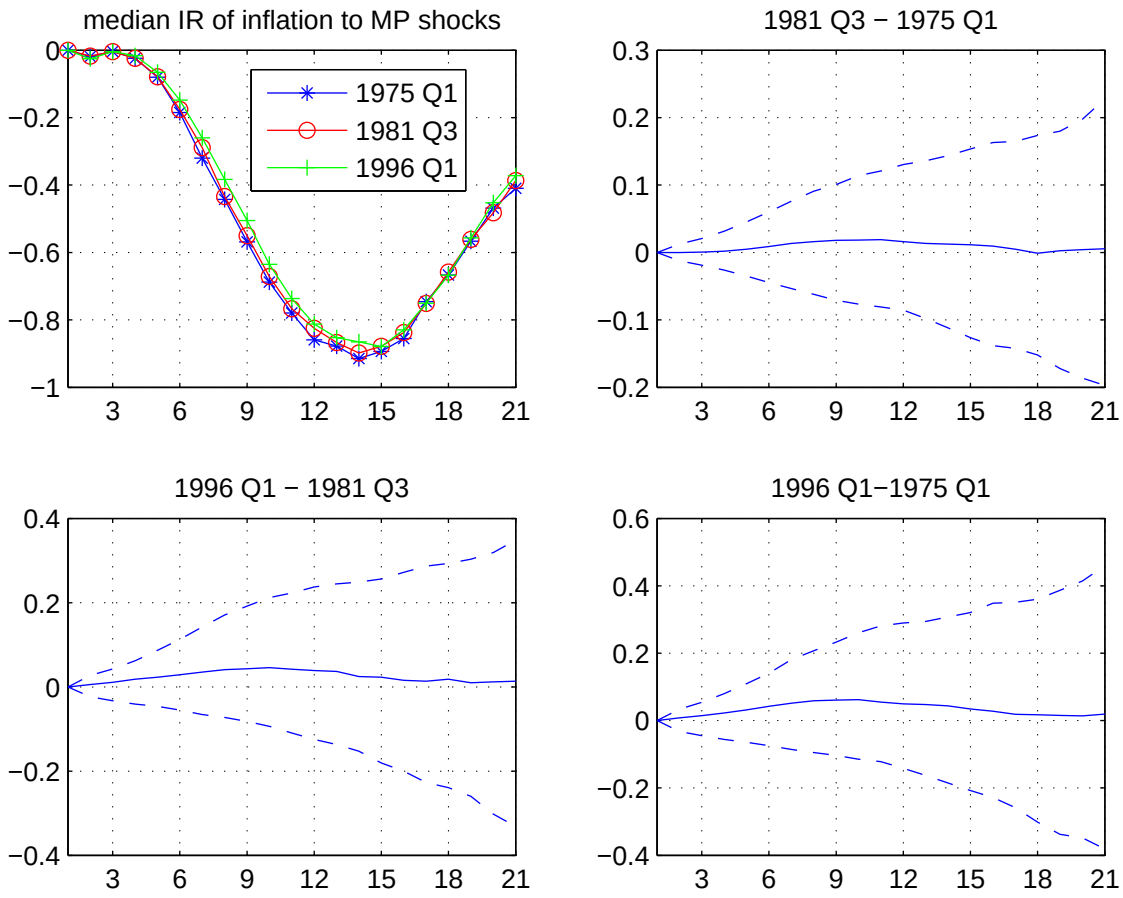


Figure 1.5: IRs-inflation-comparison for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

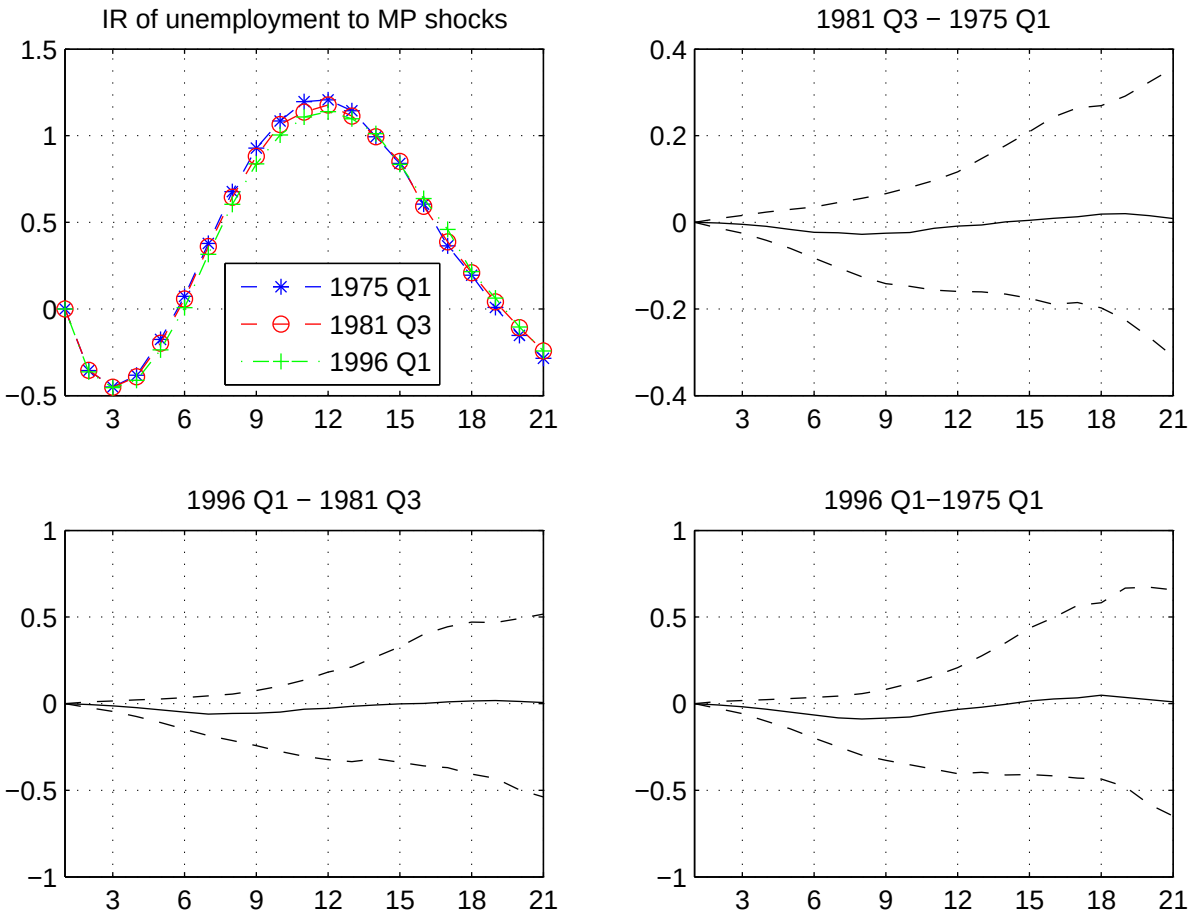


Figure 1.6: IRs-unemployment rate-comparison for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

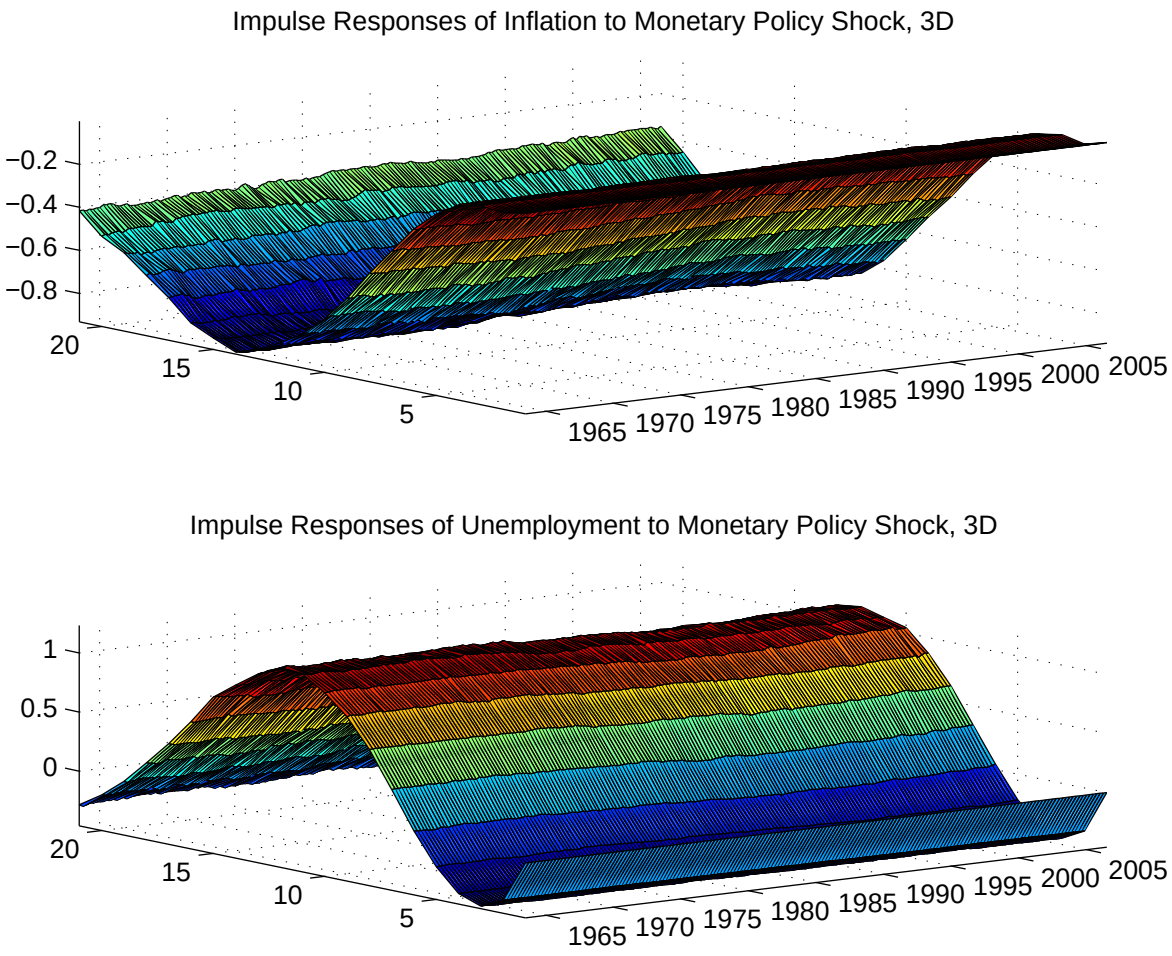


Figure 1.7: 3D-IRs for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The upper panel plots median impulse responses of inflation to monetary policy shocks over the whole sample period. The lower panel plots median impulse responses of unemployment rate to monetary policy shocks over the whole sample period.

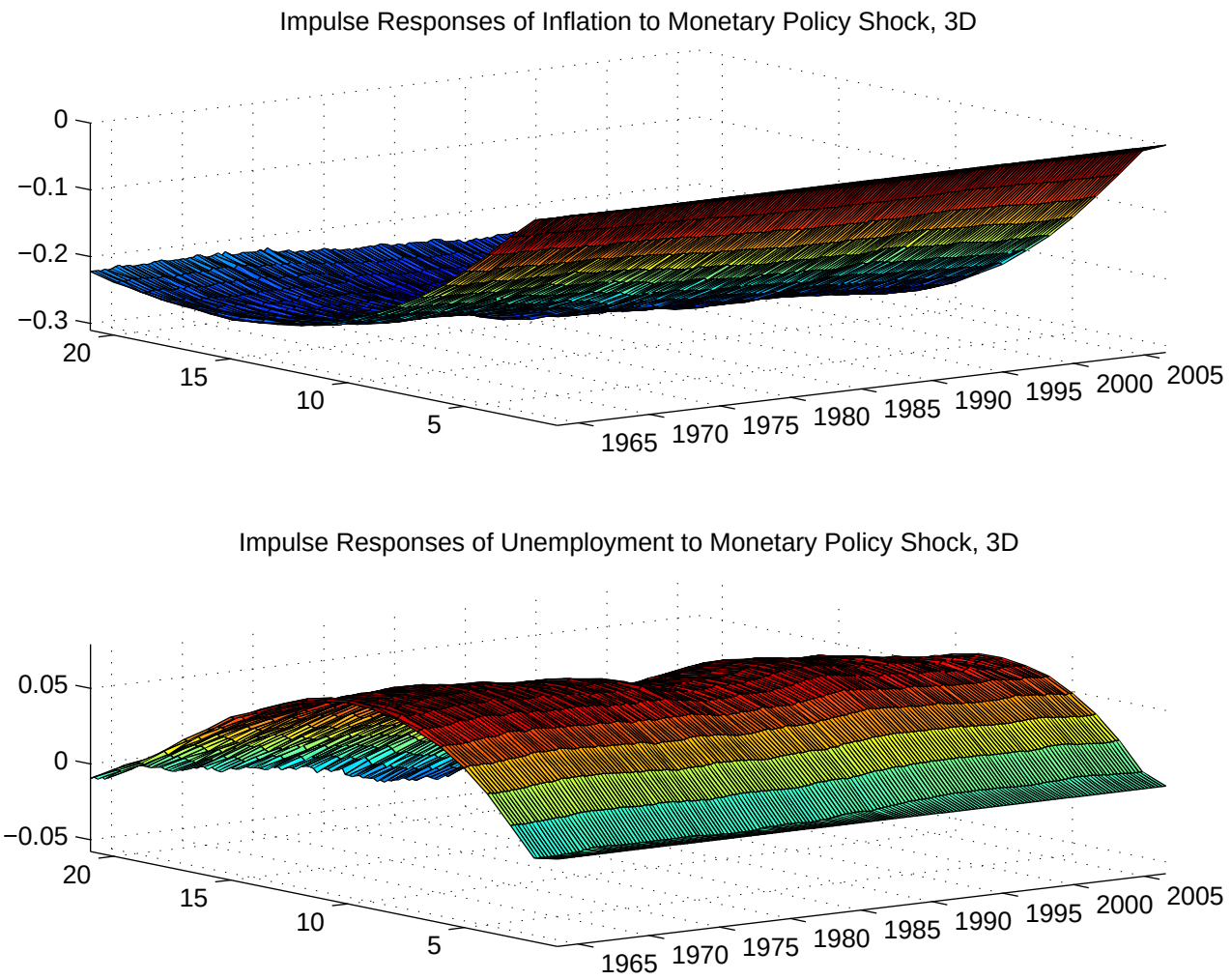


Figure 1.8: 3D-IRs for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The upper panel plots median impulse responses of inflation to monetary policy shocks over the whole sample period. The lower panel plots median impulse responses of unemployment rate to monetary policy shocks over the whole sample period.

Responses for 5 quarters, 10 quaters and 15 quarters are examined for policy strength.

We first look at policy response to inflation shock. In Figure 1.9, we find that around after middle 1970s, policy sensitivity to inflation has increased after 5 quaters, 10 quaters and 15 quaters, but not significant with a line always through the whole error band respectively. This line can be above one or below one. Figure 1.10 have the same property, however, with very broad confidence band. Therefore, we can only reach that there are no significant structural change in policy rule which is consistent with Primiceri (2005) and Sims and Zha (2006).

Monetary reaction to unemployment shocks has significantly enhanced after 5 quaters and 10 quaters, but become insignificant after 15 quaters in Figure 1.11 for model 3 over the sample period. This finding can not be captured in Figure 1.12 for model 2 as they are all not significant. Generally speaking, we can not conclude that policy stance to unemployment has significantly changed.

1.4.4.3 Non-systematic monetary policy shocks

The last panel of Figure 1.13 plots the dynamic process of stochastic volatility for nonsystematic policy with narrow band for model 3. This path is well confirmed in the literature that the volatility during era of the Great Inflation is much higher than the episode of the Great Moderation, and the flutuation in the period of monetary targeting of chairmanship of Volk is dramatically volatile. Together with the previous analysis, we only find volatility of exogenous shocks have changed substantially.

1.4.4.4 An interpretation

Altogether, even though with restricted TVP-VAR via stochastic variable selection in model 2 and model 3, We still have the almost the same conclusion that there are no significant switches in agents' behavior and monetary policy stance, as opposed to which, exogenous shocks have altered strongly.

$r_t = \rho_r r_{t-1} + (1-\rho_r)(\rho_\pi \pi_t + \rho_y y_t) + \varepsilon_t^M$ where long run response ρ_π and ρ_y represent monetary policy stance.

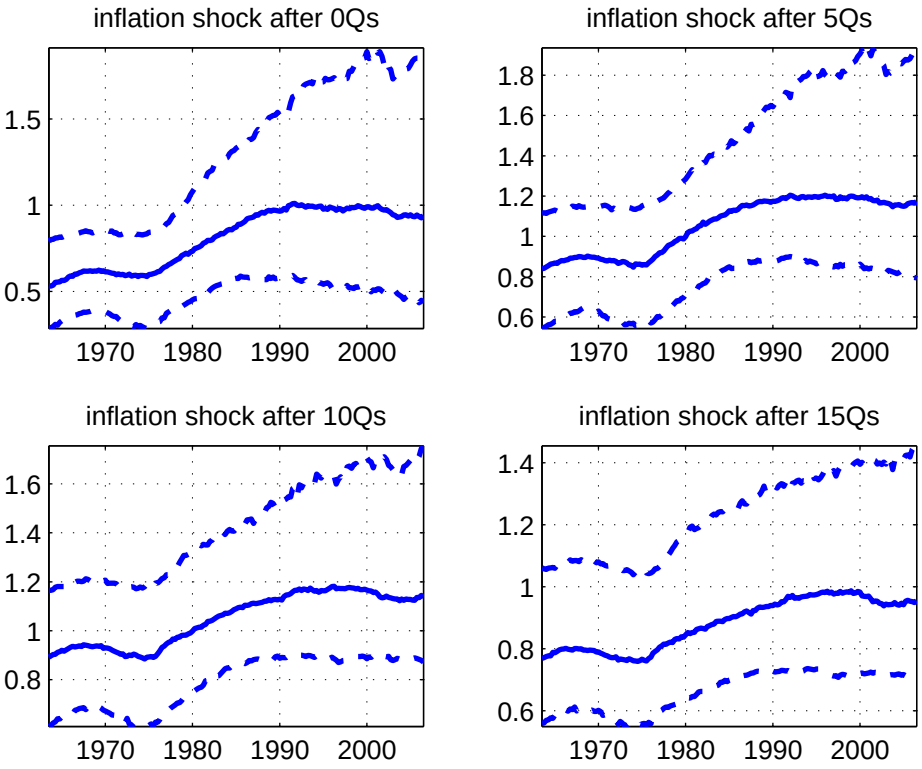


Figure 1.9: Policy response to inflation for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The upper left panel plots immediate impulse responses of policy rate to inflation shocks over the whole sample period. The remaining panels plot impulse responses of policy rate to inflation shocks in the 5^{th} , 10^{th}, and 15^{th} period respectively over the whole sample period.

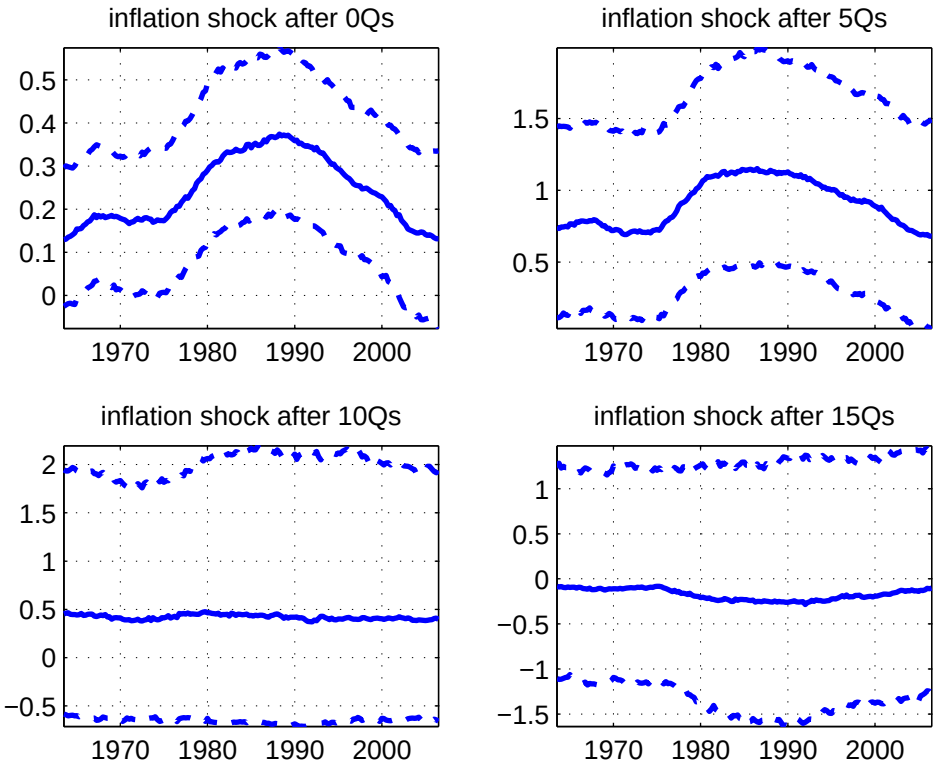


Figure 1.10: Policy response to inflation for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The upper left panel plots immediate impulse responses of policy rate to inflation shocks over the whole sample period. The remaining panels plot impulse responses of policy rate to inflation shocks in the 5^{th} , 10^{th}, and 15^{th} period respectively over the whole sample period.

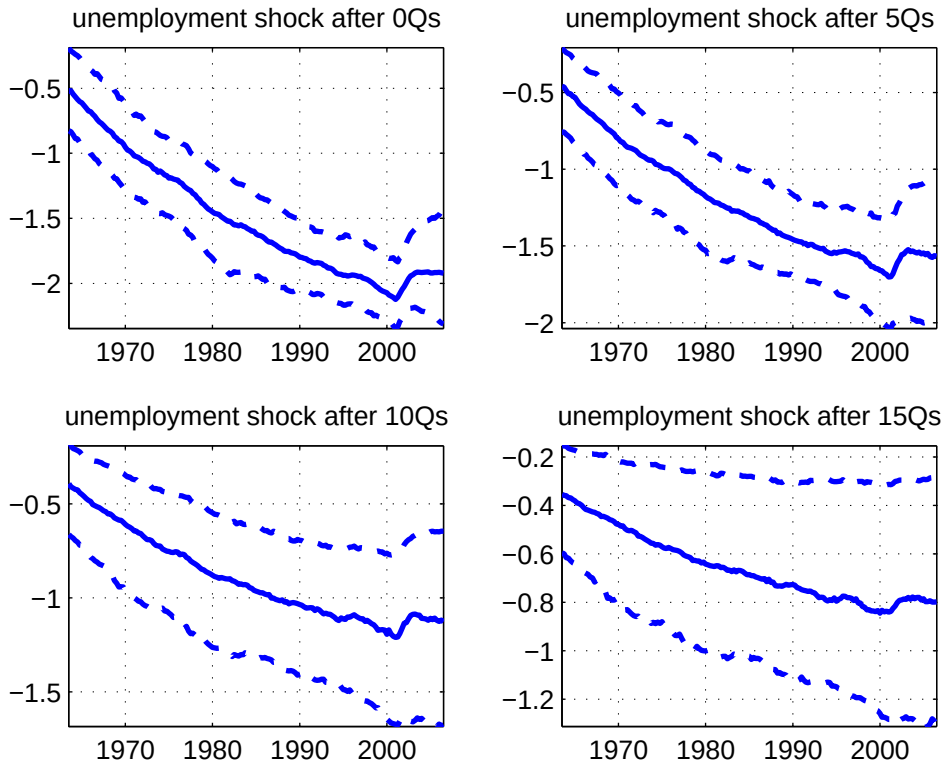


Figure 1.11: Policy response to unemployment rate for model 3

Notes: The model 3 is SVS-Full-TVP-VAR with SV. The upper left panel plots immediate impulse responses of policy rate to unemployment rate shocks over the whole sample period. The remaining panels plot impulse responses of policy rate to unemployment rate shocks in the 5^{th} , 10^{th}, and 15^{th} period respectively over the whole sample period.

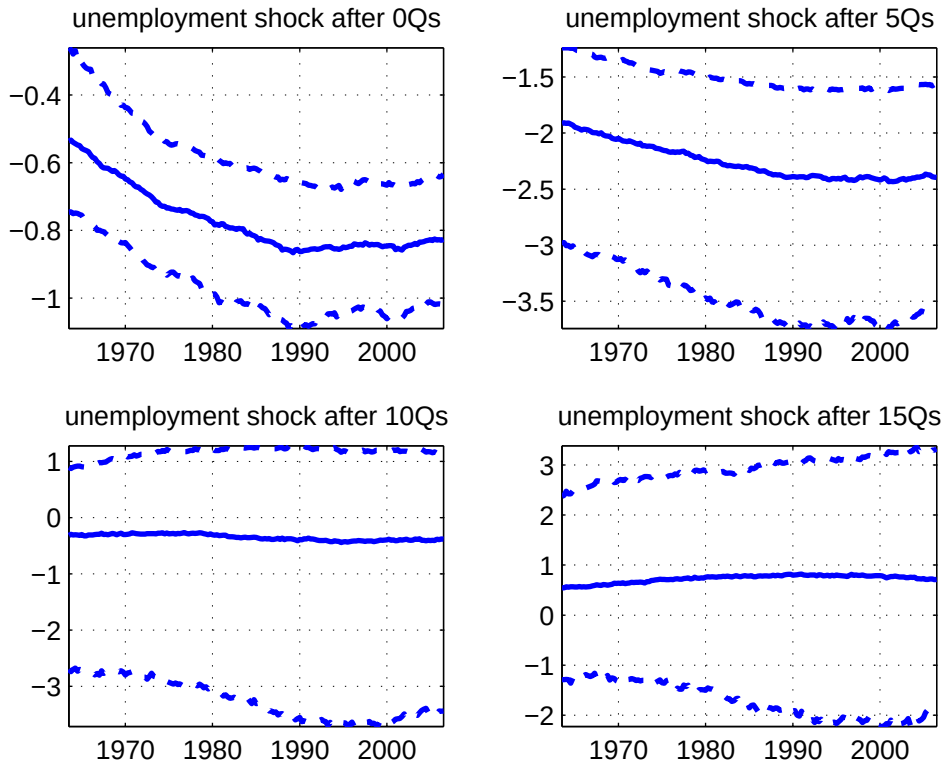


Figure 1.12: Policy response to unemployment rate for model 2

Notes: The model 2 is SVS-Partial-TVP-VAR with SV. The upper left panel plots immediate impulse responses of policy rate to unemplyment rate shocks over the whole sample period. The remaining panels plot impulse responses of policy rate to unemployment rate shocks in the 5^{th} , 10^{th}, and 15^{th} period respectively over the whole sample period.

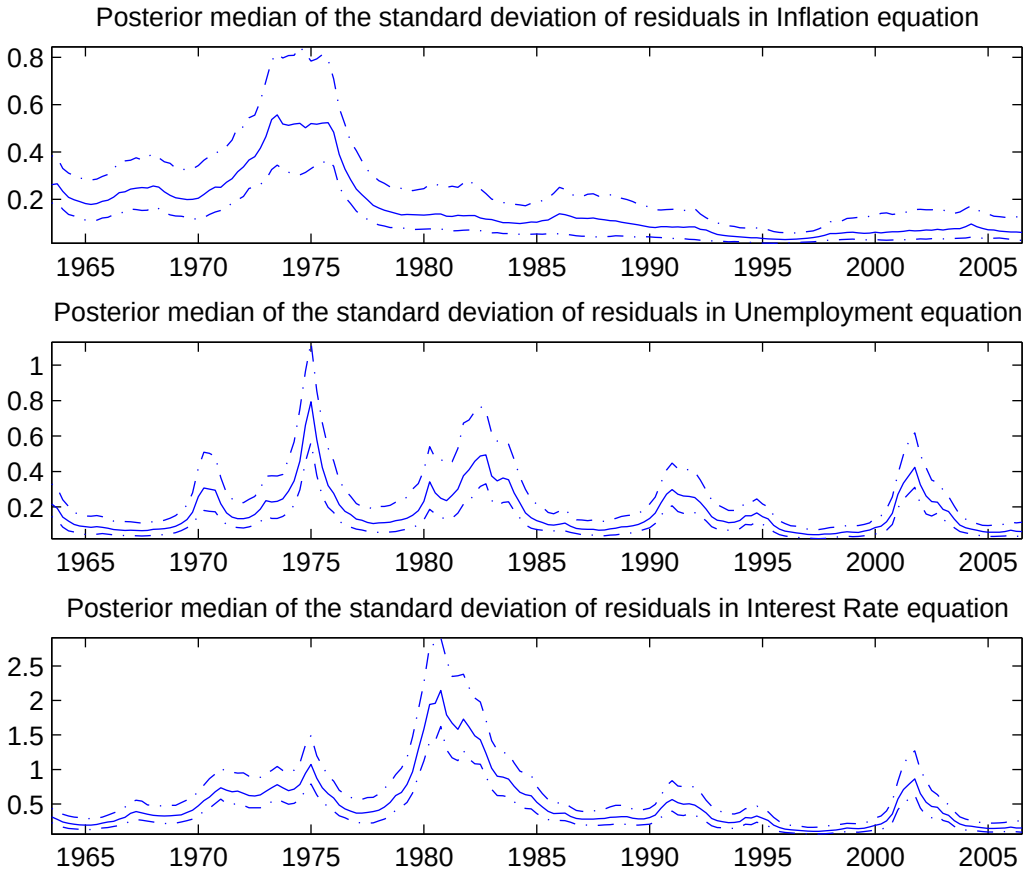


Figure 1.13: Volatility for each variable in model 3

A possible reason we believe is that although restricted TVP-VAR of model 2 and model 3, especially for model 3, have significantly improved the parameter prolification problem and reduced uncertainty, but this merit is not strong enough to overturn the view of stability of structures existing both in economic agents and policy authority. The prior probability for each coefficient in each equation is set to $\pi_{0jt} = 0.5$. Fig 1.14 to Fig 1.16 shows the posterior significance of each coefficient along the sample period. Two points can be found that first, variables always react to its own first lag with probability of one, and then become week around 50% with special case for inflation (equation 1) where it also strongly response to its own lag 2 during some periods of the Great Inflation and money base targeting; Second, other lags, no matter domestic or foreign, fluctuate with the mild and relatively the same magnitude around half percentage. These two points are consistent with empirical literature such as settings of shrinkage priors (for example, the Minnesota prior, see Doan, Litterman and Sims, 1984, Litterman, 1986 and its extension, see Banbura, Giannone and Reichlin, 2010, among others).

From these figures, there is no prominent jump of probability, except for their own lags, for each coefficient in each equation over sample period. That is, every coefficient always plays almost the same (relative) importance in this system. Structural change needs abrupt shift and keep it for a period or gradual increase or decease to over some critical threshold point. We can not find such style in these figures.

1.5 Robust check

Every TVP-VAR model has the nature that its latent state coefficients have to change or break every period due to its transition equation dynamics and this change has to be smoothed because of the backward recursion as described in Carter and Kohn (1994) in estimation after observing all the data. Hence there is a situation in which some state variables of small variations will more or less take over and narrow down the weight originally played by those of large variation in order to make the estimation smooth. Under such kind of operation, Primiceri (2005) discussed this

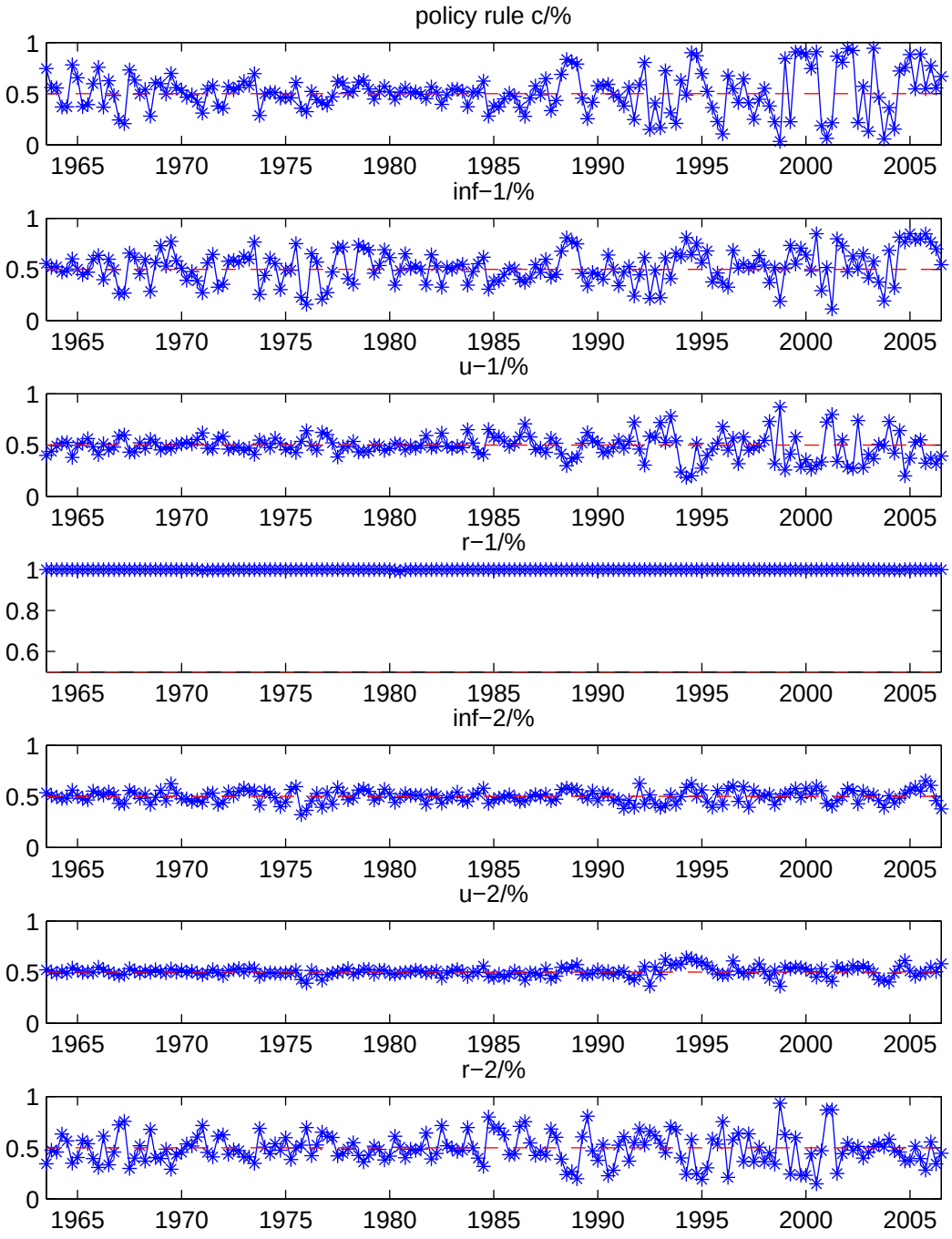


Figure 1.14: Posterior probability for each coefficient in policy rule

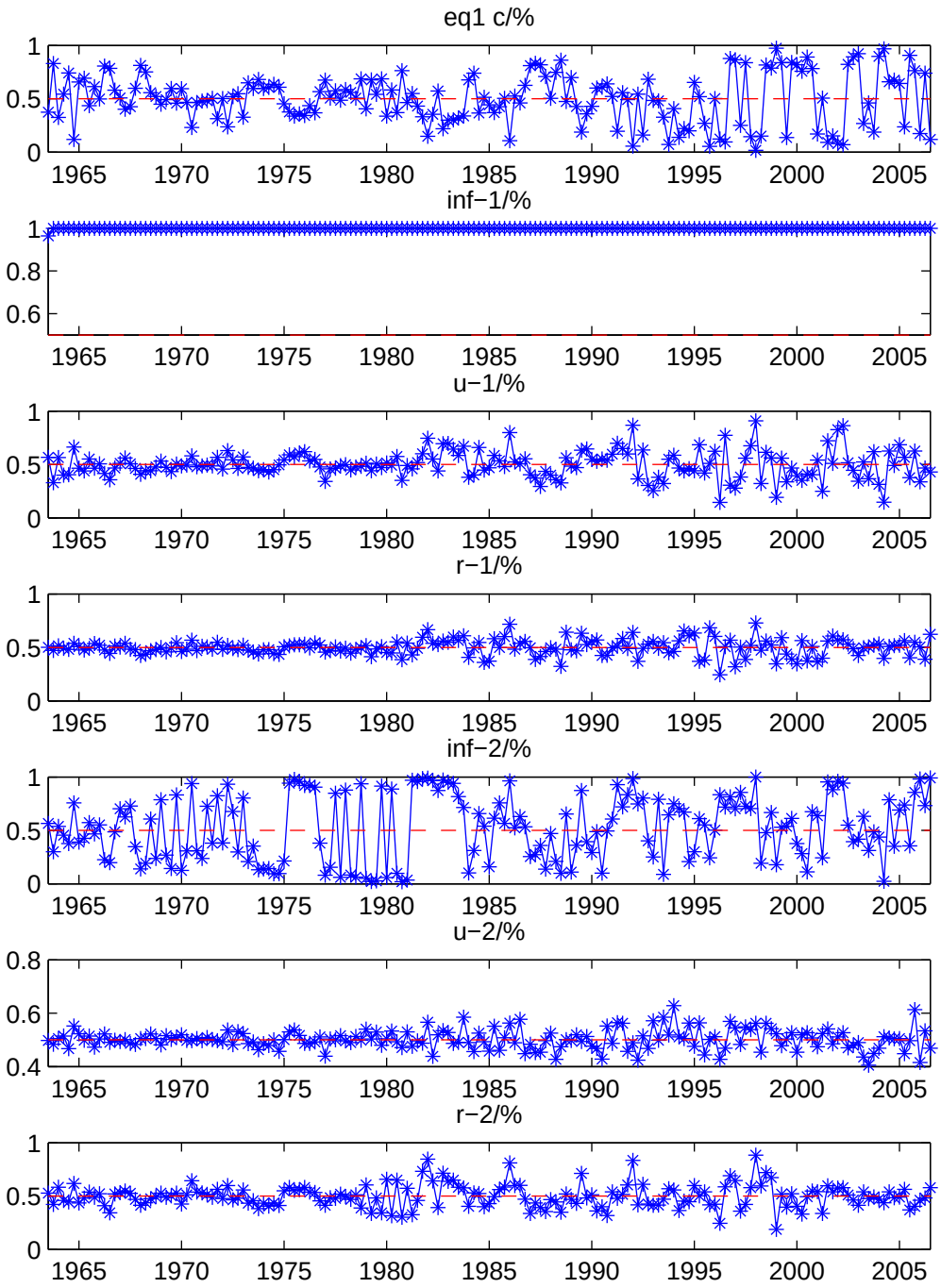


Figure 1.15: Posterior probability of each coefficient in inflation equation

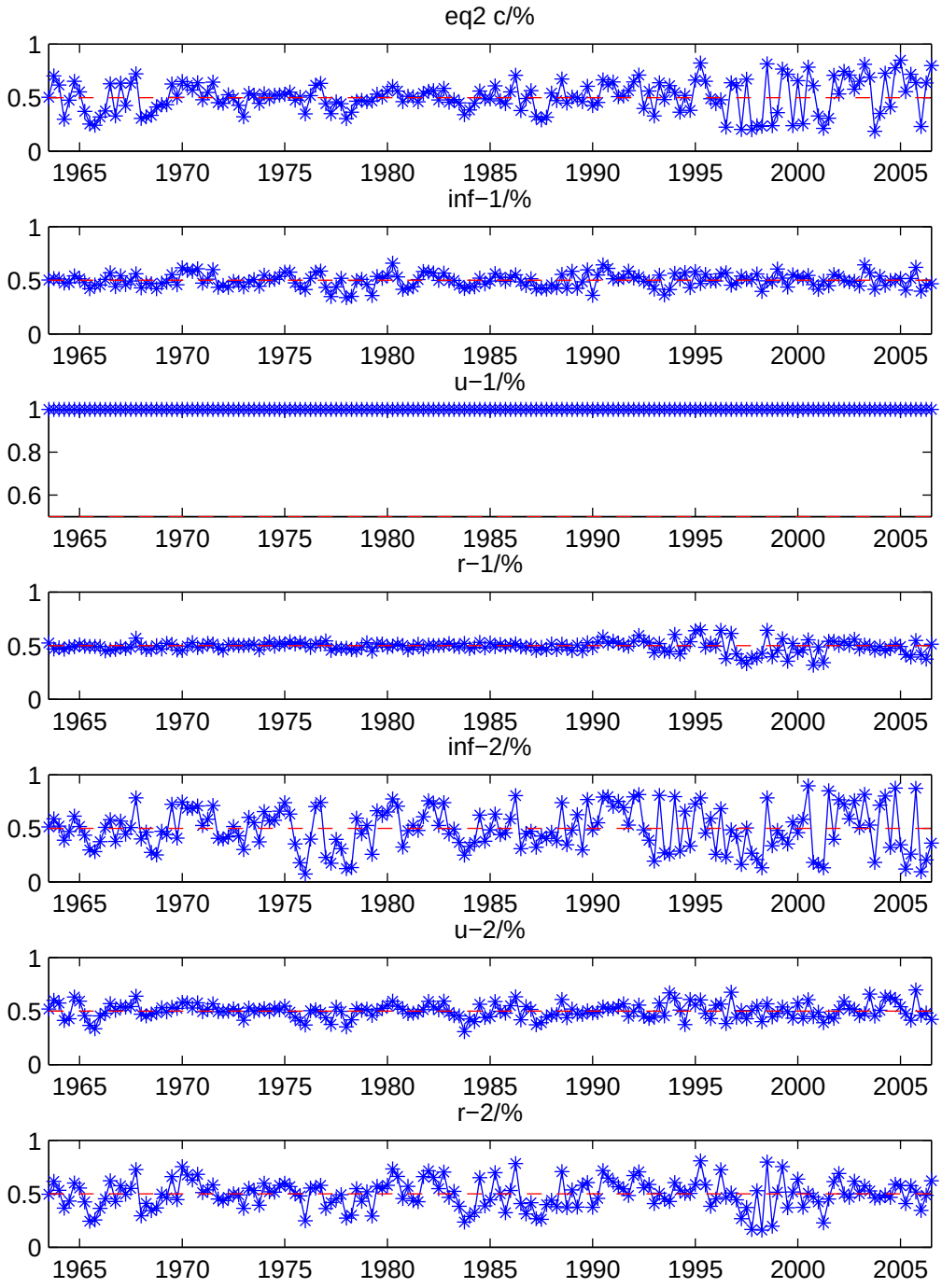


Figure 1.16: Poterior probability of coefficient in unemployment rate equation

concern. In this case, misleading could take place and this is potential missing in the proof via Fig 1.14 to Fig1.16 deriving from model 3 with also such kind of smoothness operation. Probably it is one reason for Sims and Zha (2006) to choose Markov regime switch models to detect structural changes occurring in an economy.

The above concern motivates the aim of the robust check in this section. Given that large scale breaks have high potential to be candidates of systematic changes while non-large-breaks not, then naturally, designing models good at being able to seize large breaks and meanwhile defense influence from not large variations is a necessary direction. Stepping further, if we still can not find systematic change under the large beak, it implies that it's more impossible to infer structural alterations in other non-large-break periods, because these non-large-breaks are more less qualified to be considered as possible structural switches. This deletes the above concern, complements and closes the proof that the TVP-VAR system is stable for the data set. Therefore there were no structural changes in monetary policy and economic behavior, and exogenous shocks naturally play the role of accounting for the empirical dynamic difference between the 'Great Inflation' and the 'Great Moderation' .

Recently, following Bauwens, Koop, Korobilis and Rombouts (2014), Korobilis (2013b) develop a TVP-VAR model involving discrete Markov process used for forecasting. We use this model to do our robust check.

The basic idea is that the latent coefficients in this TVP model are subject to several possible regime states which follows Markov process, meaning the regime – the corresponding coefficient – is not necessary to move every period. The model is formulated as follows modified on the previous TVP models:

$$y_t = Z_t \theta_{s_t} + \varepsilon_t \tag{1.14}$$

$$\theta_{s_t} = \Gamma \beta_{s_t} \tag{1.15}$$

$$\beta_{s_t} = \beta_{s_{t-1}} + u_t \tag{1.16}$$

where $\varepsilon_t \sim N(0,R)$, $u_t \sim N(0,Q)$ and $s_t \in [1,\ldots,M+1]$ is Markov process of order one with block diagonal transition matrix of the form

$$\begin{bmatrix} p_{11} & p_{12} & 0 & \cdots & 0 \\ 0 & p_{22} & p_{23} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ & & 0 & p_{MM} & p_{M,M+1} \\ 0 & \cdots & 0 & 0 & p_{M+1} \end{bmatrix}$$

This model specifies there is a break between t and $t+1$, namely $\beta_{s_{t+1}} \neq \beta_{s_t}$ due to $u_t \neq 0$ if only if $s_t \neq s_{t+1}$, otherwise $\beta_{s_{t+1}} = \beta_{s_t}$ and the state process s_t only goes forward and never comes back without memory.[12] This defines the new transition equation. Not every β_{s_t} is time varying, only when regime switch happens that does deserve. In other words, not always, only the limited variations in β_{s_t} uniquely correspond to several large breaks (herein M breaks) that have relatively more importance than other non-large-breaks in the data set.

The assumption of regime switching 'without memory' (the block diagonal transition matrix) seems very strong, probably not in line with empirical evidence, but it is very suitable for us to detect large breaks – the candidates of structural changes – in posterior respect meanwhile filtering out small variation not qualified as 'breaks' that however always involved in TVP models causing potential contamination of inference. In a word, the model only paying attention to large breaks has been already sufficient for us to find large breaks.

We priorly set two breaks, namely three regimes for U.S. sample set.[13] Fig 1.17 plots the posterior probability for each regime. It is evident that only two regimes, regime 1 and regime 2

[12]The 'never comes back without memory' is based on the idea that it is reasonable to regard it as a 'new break' if u_t is large enough and also on compatibility of estimation after incorporating discrete Markov (regime switching) process into TVP models for smoother part, which can be found in the appendix of Bayesian estimation of the model.

[13]We also tried three breaks but denied by the data.

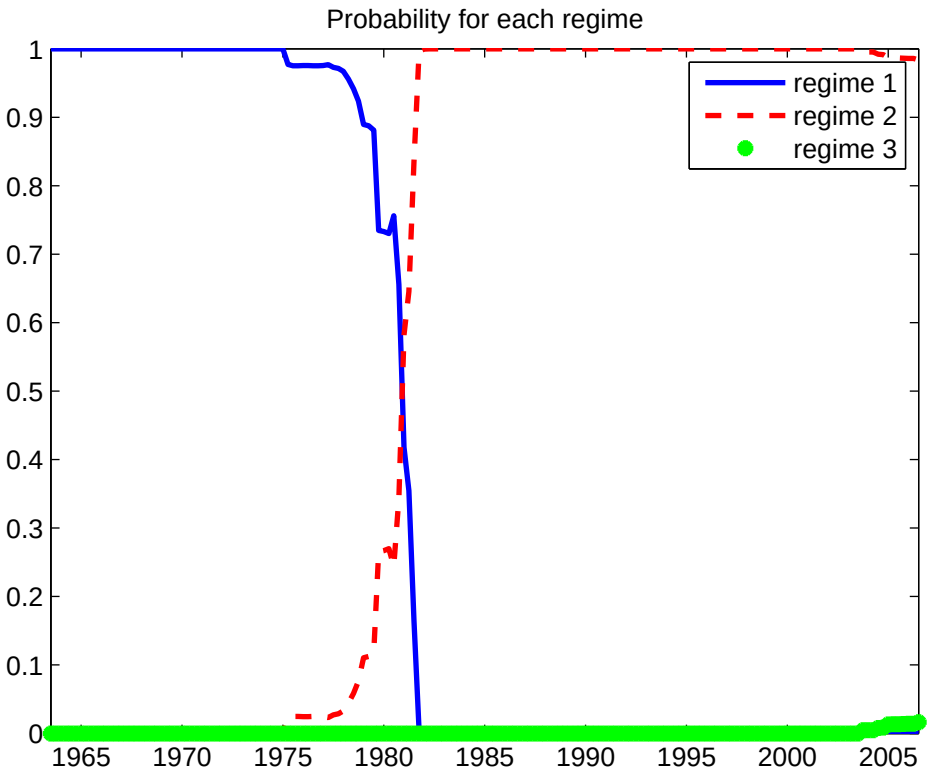


Figure 1.17: Posterior probability of each regime

are strongly supported by the data. For regime 3, it is zero percentage before around 2003, after that 2% significantly denied by the data. Now let's look at the dynamic relationship between regime 1 and regime 2. Regime 1 strongly dominated with posterior probability of value one before the end of 1980s, then after cross point around 1981, the regime 2 took the place of regime 1 dramatically climbing up to probability near one till the end of 2006 Q2 with the trivial role of regime 3 of extremely low level support. The information showing up in Fig 1.17 is not surprising, conforming with the literature that the year of 1981, among monetary targeting project in the chairmanship of Volcker, is the threshold point that divides the sample into two regimes. The property of TVP-VAR with regime switching that is able to pick up large breaks clearly catch the significant switch of monetary policy.[14]

[14]Note that the TVP-VAR with regime switch in this paper does not belong to the category of conventional state space models with regime switching in which regime switch occurs on parameters. See chapter 10 of Kim and Nelson (1999).

Has the economy changed before and after monetary targeting policy in the aim to attacking high inflation? Fig 1.18 and Fig 1.19 give the 'NO' answer. Take regime 3 into account, though it is trivial, to give a full piture. Differences among these three regimes for reactions of inflation and unemployment to monetary policy shocks are not significant with zero, which is in line with findings from Model 2 and Model 3. Monetary policy stance is also investigated. In Fig 1.20, in response to inflation shocks, the Fed's reaction had no significant differences during every two regimes, no matter in short run or long run. The same result is also found in response to unemployment shocks in Fig 1.21. Altogether, there were no systematic changes in Monetary policy and agents' behavior.

Since in large breaks systematic changes are not found, the concern in model 2 and model 3 about possible influence of small variation on large variation is neglectable.

1.6 Concluding remarks

In this paper, we present models for dealing with the problem of parameter proliferation – associated with potential incorrect inferences – in VAR models (under usual limited data access). They are static SVS-BVAR (Model 1), SVS-Partial-TVP-VAR (Model 2) and SVS-Full-TVP-VAR (Model 3). For the latter two models, I incorporate multivariate SV so as to investigate systematic and non-systematic changes jointly. Actually, the three models are consistent each other; Model 1 and Model 2 are special cases of Model 3 after relaxing some restrictions, as we discussed in section 2.

With these models and U.S. quarterly data, we analyzed whether there were systematic switches in U.S monetary stance and economic agents' behavior. After investigating long run responses of policy rate to inflation and unemployment shocks, respectively, we find that there were insignificant changes in systematic monetary policy along the whole sample period. We also evaluated agents' behavior examined by monetary policy shocks in three arbitrary periods, then extended to all the sample period and still no significant responses were found. We payed more attention on

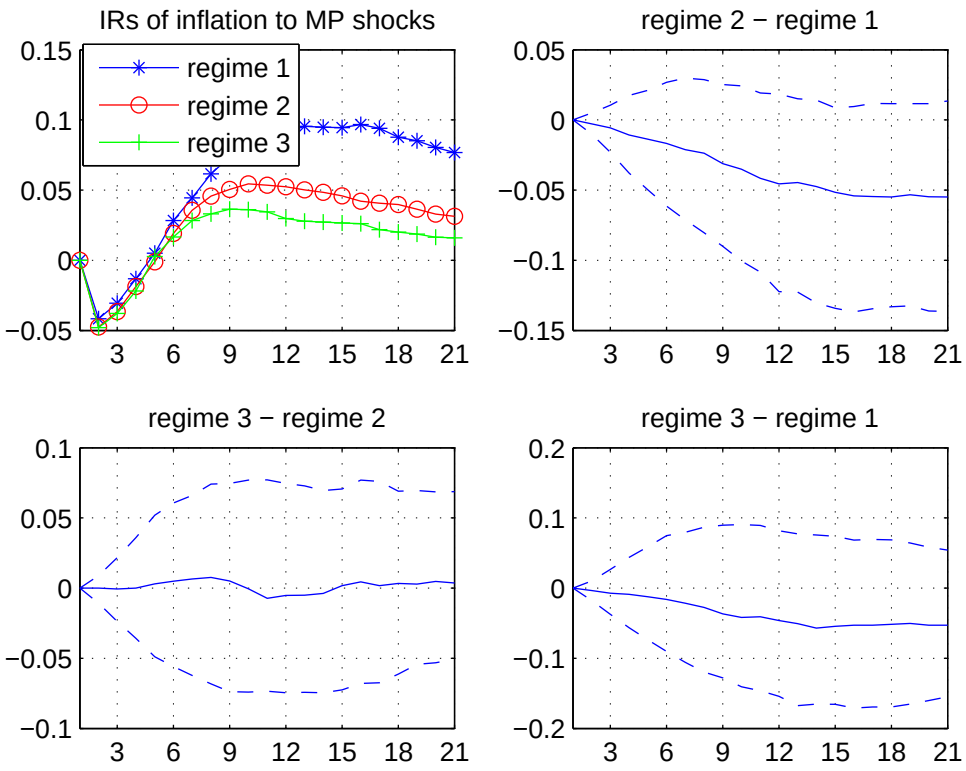


Figure 1.18: IRs of inflation and their difference between different regimes

Notes: The upper left panel plots median impulse responses of inflation to monetary policy shocks in regime 1, regime 2, and regime 3. The remaining three panels plot difference of responses between every two regimes mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

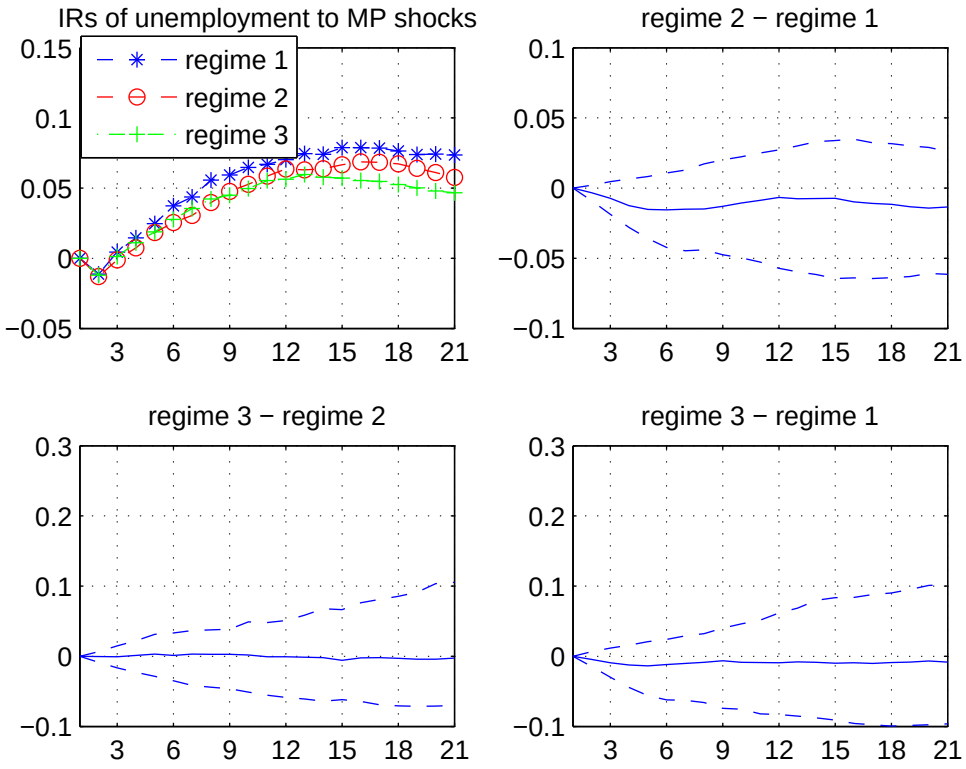


Figure 1.19: IRs of unemployment rate and their difference between different regimes

Notes: The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in regime 1, regime 2, and regime 3. The remaining three panels plot difference of responses between every two regimes mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

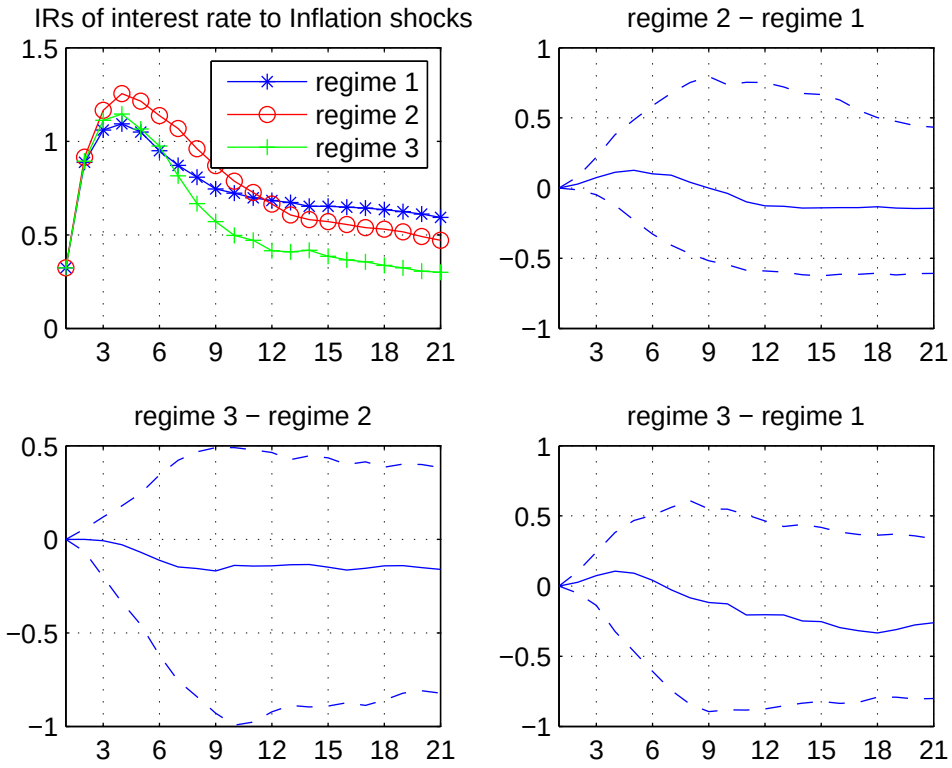


Figure 1.20: IRs of interest rate to inflation shocks

Notes: The upper left panel plots median impulse responses of interest rate to inflation shocks in regime 1, regime 2, and regime 3. The remaining three panels plot difference of responses between every two regimes mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

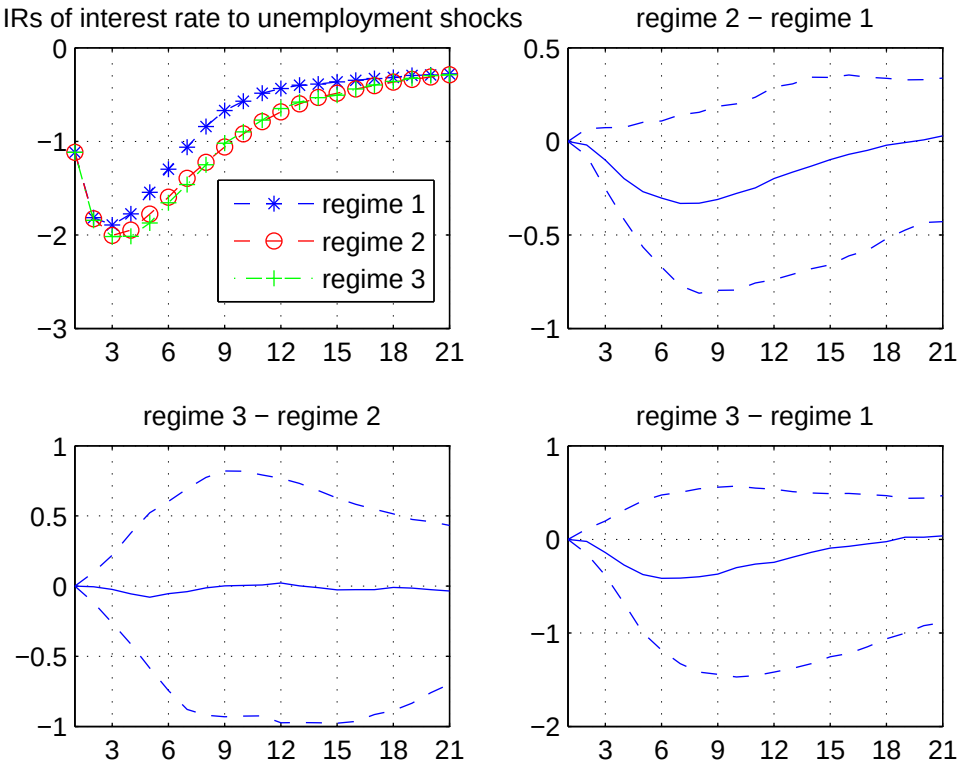


Figure 1.21: IRs of interest rate to unemployment shocks

Notes: The upper left panel plots median impulse responses of interest rate to unemployment shocks in regime 1, regime 2, and regime 3. The remaining three panels plot difference of responses between every two regimes mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

model 2 and model 3 due to their time variation property in tracking dynamic process. With the model 2 of block checking and model 3 of single checking together, in case of possible missing and misleading in model 2, We find insignificance alterations in behavior of both economic agents and policy authority.

One thing worth noting is that figures for model 3 indeed have seized some changes in the policy implementation with obviously narrow error band as opposed to model 2, however, these benefits resulting from stochastic selection controlling significance of each variable along time period have not overturn the findings in model 2. Posterior probability for every coefficient over time dimension delivers the plausible reason that except for its own lag of one, relative importance of interaction among variables over time is generally stable that can be found from Fig 1.14 to Fig 1.16.

The concern that large variation can be weakened by small variation due to smoothness operation in TVP-VAR estimation gives rise to sensitivity check via a modified TVP-VAR model involved with discrete Markov process. We find the concern can be ignored. Hence it is believable to conclude that since the system is indeed stable by statistical proof, the empirical dynamic change before and after Volcker monetary targeting can only be attributed to exogenous shocks, not determined by the system. This conclusion consistent with Primiceri (2005) and Sims and Zha (2006). On the other hand, the stability, namely the equality of relative importance of each variable in the typical small scale monetary TVP-VAR system is not sensitive to the choice of models of parsimonious restrictions, making the concern about over parameterization and smoothness typically seen in TVP-VAR system neglectable.

In this sense, the value of the paper lies not only in examining and confirming one of views in the literature using other different models, but also giving the underlying reason as well as statistical proof why their view can still hold even though with more enriched models such as stochastic variable selection models presented in this paper.

These models can apply to other topics like fiscal policy change, financial market fluctuation and so on that are left for our future research.

Appendix

A. The basic Kalman filter and state smoother

Consider a typical state space model, ignoring the influence of exogenous variables[15]:

$$y_t = X_t \beta_t + u_t \tag{1.17}$$

$$\beta_t = F \beta_{t-1} + \varepsilon_t \tag{1.18}$$

where y_t is an $n \times 1$ observables, X_t is a $n \times k$ matrix of regressors, β_t is a $k \times 1$ unobserable vector state following the dynamic process in (1.18); $u_t \sim N(0, R_t)$ and $\varepsilon_t \sim N(0, Q)$ as well as $cov(u_t, \varepsilon_s) = 0$ for $\forall t$ and $\forall s$. Usually, equation (1.17) is called measurement or observation equation and (1.18) the transition or evolution or state equation.

Denote $\beta_{t|s}$ and $V_{t|s}$ conditional mean and conditional variance of β_t based on information set up to and including t, respectively. The forward filter, with initial moments $\beta_{0|0}$ and $V_{0|0}$, consists of two steps:

Prediction

$$\beta_{t|t-1} = F \beta_{t-1|t-1}$$

$$V_{t|t-1} = F V_{t-1|t-1} F^{'} + Q$$

Updating

$$K_t = V_{t|t-1} X_t^{'} \left(X_t V_{t|t-1} X_t^{'} + R_t \right)^{-1}$$

$$\beta_{t|t} = \beta_{t|t-1} + K_t \left(y_t - X_t \beta_{t|t-1} \right)$$

$$V_{t|t} = V_{t|t-1} - K_t X_t V_{t|t-1}$$

[15] Including exogenous variables in measurement equation and/or transition equation does not affect the derivation of Kalman filter and state smoother.

this process proceeds from $t = 0$ to $t = T$. At the conclusion of the forward recursion, draw β_T from $N\left(\beta_{T|T}, V_{T|T}\right)$.

With β_T regarded as observation, backward smoother starts from $t = T-1$ to $t = 1$ with

$$\beta_{t|t+1} = \beta_{t|t} + V_{t|t}F'V_{t+1|t}^{-1}\left(\beta_{t+1} - F\beta_{t|t}\right)$$

$$V_{t|t+1} = V_{t|t} - V_{t|t}F'V_{t|t+1}^{-1}FV_{t|t}$$

then β_t is drawn from posterior $N\left(\beta_{t|t+1}, V_{t|t+1}\right)$ after observing all the data.

B. Modified Kalman filter and state smoother with breaks

Consider a modified state space model

$$y_t = X_t \beta_{s_t} + u_t$$

$$\beta_{s_t} = \beta_{s_{t-1}} + \varepsilon_t$$

in which setting, the time varying coefficients β_t depend on latent state s_t that follows a discrete Markov process. For details of the model specification, please refer to section 5 in this chapter.

For the forward filter part,

$$\beta_{t|t-1} = \beta_{t-1|t-1}$$

$$V_{t|t-1} = \begin{cases} V_{t-1|t-1} + Q & \text{if } s_{t-1} \neq s_t \\ V_{t-1|t-1} & \text{otherwise} \end{cases}$$

$$K_t = V_{t|t-1} X_t' \left(X_t V_{t|t-1} X_t' + R_t \right)^{-1}$$

$$\beta_{t|t} = \beta_{t|t-1} + K_t \left(y_t - X_t \beta_{t|t-1} \right)$$

$$V_{t|t} = V_{t|t-1} - K_t X_t V_{t|t-1}$$

For the backward smoother part,

$$\beta_{t|t+1} = \begin{cases} \beta_{t|t} + V_{t|t} V_{t+1|t}^{-1} \left(\beta_{t+1} - \beta_{t|t} \right) & \text{if } s_{t+1} \neq s_t \\ \beta_{t|t} & \text{otherwise} \end{cases}$$

$$V_{t|t+1} = \begin{cases} V_{t|t} - V_{t|t} V_{t|t+1}^{-1} V_{t|t} & \text{if } s_{t+1} \neq s_t \\ V_{t|t} & \text{otherwise} \end{cases}$$

C. Estimation of states without Kalman filter and state smoother

Consider the general state space model in (1.17) and (1.18) in A1. Stack the two equations period by period and collect all the data and states together, one obtains

$$y = X\beta + u \tag{1.19}$$

where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, X = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_T \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_T \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix},$
with $u \sim N(0,R)$ and $R = Blkdiag\left(\{R_t\}_{t=1}^{T}\right)$.

The transition equation becomes

$$H\beta = \varepsilon \tag{1.20}$$

where

$$H = \begin{bmatrix} I_q & & & & \\ F & I_q & & & \\ & F & I_q & & \\ & & & \ddots & \\ & & & F & I_q \end{bmatrix}$$

and $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix} \sim N(0,S)$ with

$$S = \begin{bmatrix} D & & & & \\ & Q & & & \\ & & Q & & \\ & & & \ddots & \\ & & & & Q \end{bmatrix}$$

and initial state $\beta_1 \sim N(0, D)$.

Equation (1.20) can be further written as

$$\beta \sim N\left(0, K^{-1}\right) \tag{1.21}$$

with precision matrix $K = H'S^{-1}H$. (1.21) is regarded as the prior for β constructed from the structure of transition equation (1.18). Thus all the state varibles nested in β can be drawn at a time without forward filtering and backward smoothing:

$$\beta \sim N(\bar{\beta}, \bar{P}_{\beta}^{-1})$$

with

$$\bar{P}_{\beta} = X'R^{-1}X + K$$

$$\bar{\beta} = \bar{P}_{\beta}^{-1}\left(X'R^{-1}y\right)$$

Note that K and R are both block banded and sparse matrices, therefore so is $\bar{P}_{\beta}$.

D. Efficient estimation of SVS-Full-TVP-VAR

Consider a standard TVP-VAR with SV:

$$y_t = c_t + A_{1,t} y_{t-1} + \cdots + A_{p,t} y_{t-p} + u_t \tag{1.22}$$

where y_t is an $n \times 1$ vector of observed endogenous variables, c_t is an $n \times 1$ vector of time varying constants, $\{A_{i,t}\}_{i=1}^{p}$ are $n \times n$ matrices of time varying autoregressive parameters, and u_t is an $n \times 1$ vector of shocks following normal distribution $u_t \sim N(0, R_t)$, for $t = 1, \ldots, T$. The R_t with SV has the same structure as in the section 2.3 this chapter.

Let $\beta_t = vec([c_t, A_{1,t}, \ldots A_{p,t}]')$ denote the vector of time varying parameters each period of dimension $k \times 1$, with $k = n(1+np)$ and *vec* column stacking operator. The law of motion of β_t follows random walk process

$$\beta_t = \beta_{t-1} + \varepsilon_t \tag{1.23}$$

which is assumed that $\varepsilon_t \sim N(0, Q)$.

Rewrite (1.22) into the form of

$$y_t = X_t \beta_t + u_t \tag{1.24}$$

where $X_t = I_n \otimes [1, y_{t-1}', \ldots y_{t-p}']$, with $\otimes$ Kronecker product. Then stacking (1.24) over time to pool all the data together, one obtains

$$y = X\beta + u$$

where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, X = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_T \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_T \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix},$

with $u \sim N(0, R)$ and $R = Blkdiag\left(\{R_t\}_{t=1}^{T}\right)$.

The dynamic process of β_t is also stacked into

$$H\beta = \varepsilon \tag{1.25}$$

where

$$H = \begin{bmatrix} I_q & & & & \\ I_q & I_q & & & \\ & I_q & I_q & & \\ & & & \ddots & \\ & & & I_q & I_q \end{bmatrix}$$

and $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix} \sim N(0,S)$ with

$$S = \begin{bmatrix} D & & & & \\ & Q & & & \\ & & Q & & \\ & & & \ddots & \\ & & & & Q \end{bmatrix}$$

and initial state $\beta_1 \sim N(0,D)$.

From (1.25), the prior of whole states is distributed as

$$\beta \sim N_{kT}\left(0, K^{-1}\right)$$

with precision matrix $K = H'S^{-1}H$. The prior is derived from the random walk structure seen in matrix H.

After above several steps of transformation, a stochastic variable selection TVP-VAR with SV is finally expressed as

$$y = X\theta + u \tag{1.26}$$

$$\theta = \Gamma\beta$$

with prior $\beta \sim N\left(0, K^{-1}\right)$ and $\Gamma = diag\left(\{\Gamma_t\}_{t=1}^{T}\right)$ in which $\Gamma_t = diag\left(\gamma_{1,t}, \ldots, \gamma_{k,t}\right)$. Apparently,

the large square matrix Γ contains all the indicators $\gamma_{j,t}$ over identity and time corresponding to the large column vectorβ stacking all the states β_t.

The Bayesian estimation procedure is similar to the benchmark model of SVS-BVAR due to the exclusion of Kalman forward filtering and backward smoothing. Here, the steps associated with stochastic selection are given:

1. Draw β from the posterior density

$$\beta \sim N_{kT}(\bar{\beta}, \bar{P}_{\beta}^{-1})$$

with

$$\bar{P}_{\beta} = X^{\star\prime} R^{-1} X^{\star} + K$$

$$\bar{\beta} = \bar{P}_{\beta}^{-1} \left(X^{\star\prime} R^{-1} y \right)$$

and $X^{\star} = X\Gamma$.

2. Sample γ_l from the posterior density

$$\gamma_l \sim Bernoulli(1, \bar{\pi}_l)$$

where $l = 1, \ldots, kT$, $\bar{\pi}_l = \frac{l_{0l}}{l_{0l}+l_{1l}}$, with

$$l_{0l} = p(y|\theta, \gamma_{-l}, \gamma_l = 1)\pi_{0l}$$

and

$$l_{0l} = p(y|\theta, \gamma_{-l}, \gamma_l = 0)(1 - \pi_{0l})$$

The expression $p(y|\theta, \gamma_{-l}, \gamma_l = 1)$ and $p(y|\theta, \gamma_{-l}, \gamma_l = 0)$ are conditional likelihood expressions. Here we define $\theta^{\star}$ to be equal to θ but with the l^{th} element $\theta_l = \beta_l$ in the case of $\gamma_l = 1$. Similarly, we define $\theta^{\star\star}$ to be equal to θ but with the l^{th} element $\theta_l = 0$ when $\gamma_l = 0$.

Then in terms of likelihood of (1.26), we can write l_{0l}, l_{1l} analytically as

$$l_{0l} = \exp(-\frac{1}{2}(y - X\theta^{\star})'R^{-1}(y - X\theta^{\star}))\pi_l \tag{1.27}$$

$$l_{1l} = \exp(-\frac{1}{2}(y - X\theta^{\star\star})'R^{-1}(y - X\theta^{\star\star}))(1 - \pi_l) \tag{1.28}$$

Note that the order l^{th} is randomly picked up, which is equivalent to assigning randomness to both identity j and time t, for $j = 1, \ldots, k$ and $t = 1, \ldots, T$, respectively.

3. Draw R_t with stochastic volatility that contains the same blocks and steps as in Primiceri (2005) and Del Negro and Primiceri (2013) or De Wind and Gambetti (2014).

4. Go back to setp 1 again, start the next iteration.

Chapter 2

A General Parsimonious Estimation of Time-varying Vector Autoregressions

2.1 Introduction

Since the pioneering work of Sims (1980), the vector autoregressive (VAR) models have become popular and widely used in economic and policy analysis. The most importance of vector autoregressive models is that it provides a tool to analyse the dynamic relationship among multiple macroeconomic variables by allowing all variables in a vector to response to all variables at all lags, as in a economy practioners and economists not only care about the intertemporal relations, but also focus on dynamic processes among macroeconomic variables which can lead them to conduct economic analyses and forecasting. Simplicity, tractability and properties of VARs such as impulse response functions and variance decompotion make the original complicated problem of exploring inter-relationship among different macro observations straightforward. Therefore, VAR models gradually become more and more popular and dominate the empirical analysis in macroeconomics. For instance, researchers typically use VARs to find some empirical realities, then relying on these findings to establish dynamic stochastic general equilibrium models (DSGE), trying to interpret the underlying rationales.

Econometric models are always evolutionary with the requirement to converge as close as possible to economic reality. The U.S. economy has experienced the Great Depression – the 1930s, the Great Inflation - 1970s and early 1980s, the Great Moderation from middle 1980s to 2006 that most industrialized economies also have the similar experience, and recent the Great Recession since the global financial crisis in 2008 and still on the way to the recovery. Except for the characteristics of business cycles, the U.S. economy also has experienced four chairmanships of Federal Reserve, Burns (1970-1978), Volker (1979-1987), Greenspan (1987-2006) and Bernanke (2006-2014). Did they conduct the same monetary policy or not? What's the relationship between their policies and the alternations of the above mentioned business cycles? Besides monetary policy, are there other factors influencing the business cycle and have these factors changed over different economic stages? To answer such questions and issues require that the conventional assumption of the constant coefficients in VARs might be poor and have to be relaxed to be time varying. The great decline in volatility – the property of the Great Moderation – in most macro-variables in the

U.S. and in most industrialized economies led to an increasing focus on appropiate modeling of the error covariance matrix in VARs and this led to the incorporation of multivariate stochastic volatility in many recent empirical papers. Hence, the time varying coefficients on regressors and stoachstic volatility on covariance matrix of forecast errors become the standard analysing tool in applied macroeonomics. In model settings, from Cogley and Sargent (2001) with only time variance on coefficients on regressors, based on which, Cogley and Sargent (2005) extended the early model to stochastic volatility on covariance matrix of forecast errors with some restrictions, finally Primiceri (2005) developed a model with sufficient flexibility on all the parameters that can be considered as the today's standard TVP-VAR workhouse.

The time verying parameter VARs with stochastic volatility is not a parsimonious model. This kind of model increase paramters dramatically. Suppose a VAR with n dimention and p lags and a constant. The number of coefficients on regressors follows $k = n(1 + n \cdot p)$ and the number of parameters on covariance matrix of errors is $m = \frac{n(n+1)}{2}$. When n and p increase, the total number of paramters of this model will increase nonlinearly and causes the so called problem of parameter proliferation, given that we usually have limited length of macro data sets. In addtion, satisfying the recent research requirement of extending the constant parameters to time varying ones, there will be T, the net sample periods in this model, time paths for the $k+m$ parameters and the newly created, associated covariance matrices of innovations to the dynamic process of above parameters. In all, this TVP-VAR has total parameters of $K_p(n,p,T) = T(k+m) + \frac{k(k+1)}{2} + \frac{n(n+1)}{2} + \frac{r(r+1)}{2}$ where $r = \frac{n(n-1)}{2}$. With limited sample span, over-parameterization, more or less, is inevitable, especially in medium and large scale TVP-VAR models. This explains why in practice, small scale models with short lags are often seen such as moetary policy analysis typically with three variables and two lags.

Over parameterization problem in TVP-VARs with stochastic volatiliy strongly limits the number of variables and lags that can be incorporated in the model. Nevertheless, for many applications a large set of variables and more lags are necessary. In a modern economy, a large number of variables work together and react each other. A variation of one variable will cause fluctua-

tions of other variables and these dynamics will typically continue for a period. If some important varibles are missing in the model, it probably will give rise to the corresponding missing of transition channels or shocks. A typical case is 'price puzzle' that will cause misunderstanding and misleading in economic and policy analysis that usually seen in a small scale monetary analysis. Banbura, Giannone, and Reichlin (2010), Carriero, Kapetanios, and Marcellino (2011) and Koop (2013) demonstrate that a system of 15-20 variables performs better than small systems in point forecasting and structual analysis.

As for the number of lags, for a quarterly data, Blanchard and Perotti (2002) argue that at least 4 lags can catch the dynamic interaction of economy and fiscal policy. For a monthly data, Uhlig (2005) use 13 lags to analyse effects of monetary policy shock. Alessandri and Mumtaz (2014) investigate the effects of uncertainty shock under different financial regimes, also use 13 lags for a monthly data set. Generally, more variables and lags can capture potentially possible inter-reactions among different variables.

Two conflicts arise. The first conflict is between preferred more variables and more lags and the parameter proliferation, which becomes even stronger under time varying parameter framwork that is often desired and required in current empirical time series analysis. As discussed above, the problem of over parameterization is alway accompanying and become serious with the increase of dimension of observations, number of lags and sample periods. The other conflict is between estimation, computational burden and tractability. The TVP-VAR with stochastic volatility model essentially is a combination of different state space models. The time varying parameters on regressors and covariance matrix actully are state or latent variables in state equations that drive dynamic process in measurement equations given other parameters and data. Note that the incorporation of stochastic volatility typically involves nonlinear and non-gaussian state space blocks which absolutely require high estimation skill and increase computational burden. Large data set, long lags as well as time varying parameters make the estimation and computation much complicated and sometimes hard to deal with.

The challenge facing the economists is how to build models or modify conventional models,

that are flexible enough to capture the main dynamic process in the data set, but not to cause serious over-parameterization, thus making the estimation tractable and result believable. There are two branches in the literature. The first one is on shrinkage.[1] The starting point is to set priors that impose restrictions on the parameters, say shrinking to zero. The Minnesota Prior (see Doan, Litterman and Sims, 1984 and Litterman, 1986) probably is the most typical one among them. The basic idea of the prior is to make the VARs shrink to random walk, with stronger shrinkage for coefficients on longer lags and across variables. Further development of this prior includes imposing restrictions on sum of coefficients and cointegration, see Banbura et.al. (2010). Note that the dimension of data set n has not changed, only using some particular priors to shrink to the desired values of parameters. The second branch is based on factor idea that has been applied to different directions. Since large data set is desirable and available in empirical analysis, several factors that extracted from large data set could significally decrease the model dimension n and alleviate over parameterization problem replying on the assumption that the bulk of dynamic inter relations within a large data set can be explained by several common factors (Forni et al., 2002 and Stock and Watson, 2002b). The factors that reduce the dimension and VARs that explore dynamic interrelationship motivate the combination of the two methods. Bernanke, Boivin and Eliasz (2005) and Stock and Watson (2005) have combined the two models, the so called factor augmented vector autoregressive models (FAVARs). Del Negro and Otrok (2008) and Korobilis (2013a) extended the FAVAR models to have time varying features on parameters, hence the TVP-FAVARs. Using also the factor idea, De Wind and Gambetti (2014) focus on the factors that drive time varying parameters instead of those that drive large data set in factor models and factor augmented VAR models. They find that a q dimensional factors can sufficiently capture the bulk of time variation in the k dimensional latent parameters in the full rank TVP-VAR model and sometimes $q \ll k$. That is the covariance matrix of the innovations to the time varying parameters is reduced rank rather than full rank. Simply speaking, this model tranforms the original k sources of variations, the same dimentison as the number of time varying parameters, to q sources, the number of factors. Kim

[1] Using shrinkage prior in large Bayesian VAR will be discussed in Chapter 3.

and Yamamoto (2012) proposed a new approach to apply the factor idea. They argue that since more lags sometimes are necessary to capture the dynamic process among variables, especially in monthly data, recent lags could be qualified to drive longer lags due to that current variables are more affected by recent lags in a standard VAR framwork. That is the constant vector and several coefficient matrices on the recent lags are the dynamic sources driving the variation of the standard TVP-VAR model. Factor idea can also be used in the time variation, namely the stochastic volatility, on covariance matrix of forecast errors in VAR. Carriero, Clark and Marcellino (2012) proposed a computationally effective way to model stochastic volatility to greatly speed up computations for smaller VAR models and make estimation tractable for larger models. The newly presented method links to the observation that the pattern of estimatied volatilities in empirical analysis is often very similar across variables. They used a common unobserved factor – common volatility – to drive the individual volatility in standard TVP model with stachastic volatility. They find that common volatility model significantly improves model fit and forecasting accuracy compared to constant volatility. Alessandri and Mumtaz (2014) by a VAR model and Mumtaz and Theodoridis (2014) via a dynamic factor model also use common volatility to study effects of 'endogenous' uncertainty (generated from the model itself) on the U.S. economy.

As we discussed above, TVP-VAR with SV is not a parsimonious model which almost always company over parameterization concern that could lead to misunderstanding and misleading in economic inferences. A natural solution to this estimation is to make the model parsimonious and tractable, meanwhile to be able to catch main dynamic characteristics among these variables.

In this chapter we propose a general parsimonious estimation method based on factor idea that collects all the advantages of the above mentioned models. This model has the ability to solve all the possible sources of parameter proliferation such as number of lags, dimension of latent coefficients on regressors and complicated stochastic volatility, and therefore can reduce the computation burden, making the estimation tractable whatever small or large models. That's why we call it 'general parsimonious estimation'. This model is also flexible enough that could pay different attention to different sources of possible over parameterization. For instance, if the

number of lags is small enough and suitable for some economic analysis, the model only need to focus on the dimension of latent coefficients and stochastic volatility.

Three papers are relevant to our method. Kim and Yamamoto (2012) is only on reduced lags without possible factors driving latent coefficients and stochastic volatility; De Wind and Gambetti (2014) applies latent factor both on latent coefficients and stochastic volatiliy in estimation procedure, but the complication of estimation of stochastic volatiliy still there; Carriero, Clark and Marcellino (2012) use a latent factor as common volatility under constant coefficients.

We apply the 'general parsimonious estimation' presented in this chapter to small scale monetary VAR. Specifically, we increase the number of lags from one to four to create possible over parameterization enviroment. We implement principal component analysis on the covariance matrix of innovations to latent coefficients. The results show that even though with only one lag, the VAR model is still not parsimonious; as the number of lags increases, the problem of over fitting become serious and several factors are enough to capture the amount of variation that is present in full rank model. The common volatility from the model with one lag to the model with four lags works very well that coincides with the feature of U.S. business cycle. We also checked economic agents' reaction to monetary policy shocks under the parsimonious estimation and find that there are no significant changes in the responses to non-systematic monetary policy in line with Primiceri (2005) and Sims and Zha (2006). These evidences suggest that parsimonious estiamtion is not only good at alleviating over fitting, but also suitable for structural analysis.

The Chapter 2 is organized as follows. In section 2, we present the model specification that is based on factor idea to conduct a parsimonious estimation. In section 3, Bayesian estiamtion procedure is given step by step and we also point out that the free combination of different blocks in estimation is equivalent to the simple verison of the presented general model seen in the literature. Comparison with conventional model or estimation is also conducted to demonstrate the advantage of reducing over-parameterization and at the same time also increase the estimation efficiency. Section 4 in an empirical analysis, gives strong evidences of over-fitting even in a small scale TVP-VAR and evaluates the performance of the factor driven model under the setting when the

extent of over parameterization becomes more and more serious. Finally, Section 5 conludes.

2.2 Model specification

In this section, We present the general model that imposes all the possible factors that drive lags, latent coefficients and stochastic volatility. Of course, when it applies to an empirical analysis, appropriate settings should be chosen by researchers depending on the sepcific objects.

Suppose y_t is an $n \times 1$ vector that follows VAR(p) process

$$y_t = c_t + B_{1,t} y_{t-1} + \cdots + B_{p,t} y_{t-p} + u_t \tag{2.1}$$

where c_t is an $n \times 1$ time varying constant, $\{B_{i,t}\}_{i=1}^{p}$ are $n \times n$ matrices of time varying autoregressive parameters, and u_t is an $n \times 1$ vector of forecast errors. The errors are assumed to independently and identically follow normal distribution $u_t \sim N(0, \Sigma_t)$.

Let's first look at the stochastic volatility part of Σ_t. Following Carriero et. al. (2012), we assume that

$$\Sigma_t = A^{-1} D_t A^{-1'} \tag{2.2}$$

where A is a lower triangluar matrix with values of ones on the main diagonal. The volatility process is defined as

$$D_t = \lambda_t S \tag{2.3}$$

$$S = diag\left([1, s_2, \ldots, s_n]'\right) \tag{2.4}$$

$$\log(\lambda_t) = F \cdot \log \lambda_{t-1} + \eta_t \tag{2.5}$$

where $\eta_t \sim iidN(0, Q_h)$. $h = \log(\lambda_t)$ follows AR(1) process that is common to all variables and drives the time variation in the entire covariance matrix of the VAR errors. The first element of diagonal variance matrix S – the loading – is normalized to one for identification of common

volatility λ_t.

Next we can rewrite equation (2.1) using the above volatility structure in the form of

$$y_t = B_t X_t + \lambda_t^{\frac{1}{2}} A^{-1} S^{\frac{1}{2}} \varepsilon_t \tag{2.6}$$

where $B_t = [c_t, B_{1,t}, \ldots, B_{p,t}]$ and $X_t = \left[1, y'_{t-1}, \ldots, y'_{t-p}\right]'$. Following Kim and Yamamoto (2012), the time varying coefficients on regressors can be decomposed into

$$B_t = B + \bar{B}_t G \tag{2.7}$$

where $\bar{B}_t$ and G are $n \times r$ and $r \times (np+1)$ respectively. For the model to be properly identified, we assume that

$$\bar{B}_1 = 0 \text{ and } G = [I_r \; G_1] \tag{2.8}$$

we as usual assume that $b_t = \text{vec}\left(\bar{B}'_t\right)$ follows random walk

$$b_t = b_{t-1} + u_{bt} \tag{2.9}$$

where $u_{bt} \sim iidN(0, Q_b)$. Note that from $\bar{B}_t G = [\bar{B}_t \; \bar{B}_t G_1]$, $\bar{B}_t$ should have the similar time variation to the elements in the first r columns of B_t matrix of the unrestricted model.

Finally, we further assume that the $nr \times 1$ time varying paramter b_t is driven by q factors where $q \leq nr$ (De Wind and Gambetti, 2014). This means that the covariance matrix Q_b is of less than full rank. Decompose the covariance matrix $Q_b = \Lambda_b \Lambda_b'$ where Λ_b is a $nr \times q$ matrix as factor loadings implying that $rank(Q_b) = q$. The transition equation (2.9) can be written as

$$b_t = b_{t-1} + \Lambda_b \upsilon_t \tag{2.10}$$

where correspondingly υ_t is a $q \times 1$ shocks that follows $\upsilon_t \sim N\left(0, I_q\right)$. The above equation (2.10) implies that $\triangle b_t$ is on the column space of Λ_b but b_t is not necessarily in the column space of

Λ_b. In other words, changes in time varying parameters are driven by factors while levels are not necessary since there are more forces determining the economy than forces changing the economy. Thus the law of motion of (2.10) can be further written as

$$b_t = P_b b_t + M_b b_t \tag{2.11}$$

where $P_b = \Lambda_b \left(\Lambda_b^{'} \Lambda_b\right)^{-1} \Lambda_b^{'}$ is a projection matrix onto the column space of Λ_b that contains the changing part of the b_t; $M_b = I_{nr} - P_b$ is the projection matrix onto the left null space of Λ_b that sizes the time invariant part of b_t. Defining $\tilde{b}_t = \left(\Lambda_b^{'} \Lambda_b\right)^{-1} \Lambda_b^{'} b_t$ as driving factors, then

$$b_t = \Lambda_b \tilde{b}_t + M_b b_0 \tag{2.12}$$

and the law of motion in terms of the underlying factors follows

$$\tilde{b}_t = \tilde{b}_{t-1} + \upsilon_t \tag{2.13}$$

via premultiplying equation (2.10) by $\left(\Lambda_b^{'} \Lambda_b\right)^{-1} \Lambda_b^{'}$ both sides.

The model specification consists of three parts. Equations from (2.1) to (2.5) describe a common factor that drive the volatilities of all the variables in y_t. This factor application is due to two reasons. One is from the observation that the pattern of estimated volatilities in empirical analysis is often very similar across variables. As discussed above, the U.S. economy has experienced two episodes of the 'Great Inflation' and the 'Great Moderation', respectively. In the former period, most macroeconomic variables had very high volatility; while in the latter period, modest volatiliy was shared by most macro-variables. The other is that it greatly reduces the computational budern in which nonlinear and nongaussian state space involoved in full stochastic volatility part has to be transformed to linear gaussian state space via seven (Kim, Shephard and Chib, 1998) or ten (Omori, Chib, Shephard and Nakajima, 2007) mixture normals. In addition, the length of history of common volatility is always fixed at T, independent of n, while the full volatility is determined

by both T and n, i.e. $\frac{n(n+1)}{2} \cdot T$.

Equations (2.6) - (2.9) give the factor idea of using early lags to drive whole lag coefficients. Many empirical analysis, especially in field of forecasting, have evidenced that recent lags are more relevant than long lags affecting current variables which justfies our model setting. The last four equations (2.10) - (2.13) present another interpretation of dynamic process of latent time varying coefficients that only limited factors drive the the bulk of variation in time varying coefficients. De Wind and Gambetti (2014) give the emipical evidence and theoretical support for this setting.

One thing worth mention is that the three portions of the general-setting model is not necessarily connected together. They can be freely combined or independently exist with respect to the research object ones are conducting. For instance, if relaxing one of factor restrictions, the corresponding part becomes the standard setting. This can be seen clearly in Bayesian estimation procedure in the next section.

2.3 Bayesian estimation procedure

In this section, we describe the Bayesian estimation procedure step by step. In each step or, precisely speaking, in each block, after some appropriate transformation, the estimation finally reduces to standard Bayesian estimation such as linear regression models and state space models either linear or nonlinear.

2.3.1 Draw a history of $\{b_t\}_{t=1}^{T}$

We need some transformation to construct a state space model for b_t. Substituting the decomposition equation (2.7) into measurement equation (2.6), one obtains

$$y_t = BX_t + \bar{B}_t GX_t + u_t \tag{2.14}$$

It can be further written as via column stacking operator both sides

$$y_t^\star = Z_{b,t} b_t + u_t \tag{2.15}$$

where $y_t^\star = y_t - BX_t$, $Z_{b,t} = \left(I_n \otimes (GX_t)'\right)$, $b_t = vec\left(\bar{B}_t'\right)$ and $u_t \sim N(0, \Sigma_t)$; $\otimes$ denotes Kronecker product and *vec* is column stacking operator.[2]The composition of $b_t = \Lambda_b \tilde{b}_t + M_b b_0$ in equation (2.12) then is plugged into (2.15) to obtain

$$y_t^\star = Z_{b,t}\Lambda_b \tilde{b}_t + Z_{b,t} M_b b_0 + u_t \tag{2.16}$$

Clearly, We find that drawing the history of b_t is divided into three steps: i) draws of $\tilde{b}_t$, ii) draws of $M_b b_0$ and iii) draws of b_t based on i) and ii).

2.3.1.1 Draw a history of $\left\{\tilde{b}_t\right\}_{t=1}^T$

The state space model for $\tilde{b}_t$ based on equation (2.16) is orgonized as

$$y_t^{\star\star} = Z_{b,t}^\star \tilde{b}_t + u_t \tag{2.17}$$

with the law of motion in (2.13)

$$\tilde{b}_t = \tilde{b}_{t-1} + \upsilon_t$$

where $y_t^{\star\star} = y_t^\star - Z_{b,t} M_b b_0$, $Z_{b,t}^\star = Z_{b,t}\Lambda_b$, $u_t \sim N(0, \Sigma_t)$ and $\upsilon_t \sim N\left(0, I_q\right)$ due to the decomposition of Q_b. Since the above two equations form the standard linear and gaussian state space model, posterior draws of $\{b_t\}_{t=1}^T$ can be sampled by the algorithm of Carter and Kohn (1994).

Standard Kalman filter and a smoother apply to the linear and gaussian state space model for $\tilde{b}_t$. We herein give the basic description. The filter goes forward until T and obtain a draw from $\tilde{b}_T \sim N\left(\tilde{b}_{T|T}, V_{T|T}\right)$ in the last period; Then based on the draw of $\tilde{b}_T$ as an obervation, the filter goes backward into a smooth process until the first period. That is $\tilde{b}_t \sim N\left(\tilde{b}_{t|t+1}, V_{t|t+1}\right)$ based

[2] $vec(ABC) = \left(A \otimes C'\right) vec\left(B'\right)$

on previous draw as a new observation successively for $t = T-1,\ldots,1$. The forward recursive formulae are given by

$$V_{t|t-1} = V_{t-1|t-1} + I_q$$

$$K_t = V_{t|t-1} Z_{b,t}^{\star'} \left(Z_{b,t}^{\star} V_{t|t-1} Z_{b,t}^{\star'} + \Sigma_t \right)^{-1}$$

$$\tilde{b}_{t|t} = \tilde{b}_{t-1|t-1} + K_t \left(y_t^{\star\star} - Z_{b,t}^{\star} \tilde{b}_{t-1|t-1} \right)$$

$$V_{t|t} = V_{t|t-1} - K_t \left(Z_{b,t}^{\star} V_{t|t-1} \right)$$

where the notation $x_{\cdot|t}$ is used to condition on the information set up to and including time t. Note that the initialization of the recursion follows from the prior distribution on $b_{0|0} \sim N(b_0, V_0)$. By the definition of $\tilde{b}_t = R_b b_t$ where $R_b = \left(\Lambda_b' \Lambda_b \right)^{-1} \Lambda_b'$, then prior distribution becomes $\tilde{b}_{0|0} \sim N\left(\tilde{b}_0, \tilde{V}_0\right)$ where correspondingly $\tilde{b}_0 = R_b b_0$ and $\tilde{V}_0 = R_b V_0 R_b'$. The backward recursion, namely the smoother has

$$\tilde{b}_{t|t+1} = \tilde{b}_{t|t} + V_{t|t} V_{t|t+1}^{-1} \left(\tilde{b}_{t+1} - \tilde{b}_{t|t} \right)$$

$$V_{t|t+1} = V_{t|t} - V_{t|t} V_{t|t+1}^{-1} V_{t|t}$$

Finally, $\tilde{b}_t$ premultiplying by Λ_b obtain the time varying part of b_t following equation (2.12).

2.3.1.2 Draw $M_b b_0$ as a whole

After drawing of time varying part, now we consider the time invariant part of b_t. Equation (2.16) can be organized as

$$y_t^{\star\star} = Z_{b,t} \delta + u_t \tag{2.18}$$

where $y_t^{\star\star} = y_t^{\star} - Z_{b,t} \Lambda_b \tilde{b}_t$ and $\delta = M_b b_0$. The above equation can be interpreted as a restricted linear regression, as the vector coefficient δ from its definition is on the column space of M_b

naturally with the restriction $P_b\delta = 0$ or equivalently $R_b\delta = 0$.

Stacking (2.18) with the above restriction, it displays as

$$y^{\star\star} = Z_b\delta + u \tag{2.19}$$

$$R_b\delta = 0$$

where $y^{\star\star} = \left[y_1^{'}, \ldots, y_T^{'}\right]^{'}$, $Z_b = \left[Z_{b,1}^{'}, \ldots, Z_{b,T}^{'}\right]^{'}$, $u = \left[u_1^{'}, \ldots, u_T^{'}\right]^{'}$ as well as $u \sim N(0, \Sigma^{\star})$ with $\Sigma^{\star} = Blkdiag([\Sigma_1, \ldots, \Sigma_T])$ in which covariance matrix each period is placed on the main diagonal. Details on the estimation of the restricted linear regression is delegated to appendix in this chapter. The basic idea to dealing with the Bayesian estimation with restrctions on the parameters is to think of the restrictons as another prior information and incoporate it into the posterior. Since $b_{0|0} \sim N(b_0, V_0)$, the prior for δ becomes $\delta \sim N(\underline{\delta}, \underline{V}_{\delta})$ where $\underline{\delta} = M_b b_0$ and $\underline{V}_{\delta} = M_b V_0 M_b^{'}$. The posterior of δ also follows normal distribution

$$\delta \sim N\left(\bar{\delta}, \bar{V}_{\delta}\right)$$

with

$$\bar{\delta} = \left(I_{nr} - \tilde{V}_{\delta} R_b^{'} \left(R_b \tilde{V}_{\delta} R_b^{'}\right)^{-1} R_b\right) \tilde{\delta}$$

$$\bar{V}_{\delta} = \left(I_{nr} - \tilde{V}_{\delta} R_b^{'} \left(R_{\delta} \tilde{V}_{\delta} R_b\right)^{-1} R_b\right) \tilde{V}_{\delta}$$

where the posterior mean of $\tilde{\delta}$ and posterior variance of $\tilde{V}_{\delta}$ are from the standard Bayesian estimation of linear unrestricted regression

$$\tilde{\delta} = \left(Z_b^{'} \Sigma^{\star -1} Z_b + \underline{V}_{\delta}^{-1}\right)^{-1}$$

$$\tilde{V}_{\delta} = \tilde{\delta}\left(Z_b^{'} \Sigma^{\star -1} y^{\star\star} + \underline{V}_{\delta}^{-1} \underline{\delta}\right)$$

2.3.1.3 Draw a history of $\{b_t\}_{t=1}^T$

The final step is straightforward to sum the time varying part and time invariant part together, i.e. $b_t = \Lambda_b \tilde{b}_t + M_b b_0$ which yields b_t draw for $t = 1, \ldots, T$. This completes the drawing from the posterior distribution of $\{b_t\}_{t=1}^T$ via a two separate gibbs sampling.

2.3.2 Draw reduced rank covariance matrix Q_b

The posterior distribution of Q_b only depends on the history of all b_ts. Sampling is based on the following state equation (2.9)

$$b_t = b_{t-1} + u_{b,t}$$

The posterior distribution of Q_b follows inverse wishart distribution given the prior distribution of the same type

$$Q_b \sim SIW\left(\bar{Q}_b, \bar{\nu}\right)$$

with

$$\bar{Q}_b = \sum_{t=2}^{T} \left(b_t - b_{t-1}\right)\left(b_t - b_{t-1}\right)' + \underline{Q}_b$$

$$\bar{\nu} = \underline{\nu} + T - 1$$

where $\underline{Q}_b$ and $\underline{\nu}$ are scale matrix and degree of freedom respectively for prior inverse wishart distribution of reduced rank Q_b.

2.3.3 Draw constant matrix B

Go back to equation (2.14), place unrelated term to the left hand side and obtain the linear regression associated with matrix B

$$y_t^\star = BX_t + u_t \tag{2.20}$$

where $y_t^\star = y_t - \bar{B}_t G X_t$ and $u_t \sim N(0, \Sigma_t)$ with $u_t = \lambda_t^{\frac{1}{2}} A^{-1} S^{\frac{1}{2}} \varepsilon_t$. Transpose the above equation and divide by $\sqrt{\lambda_t}$ both sides, we obtain

$$y_t^{\star'} / \sqrt{\lambda_t} = \left(X_t / \sqrt{\lambda_t} \right) B^{'} + u_t^{'} / \sqrt{\lambda_t} \tag{2.21}$$

Stacking the above equation row by row

$$Y^\star = X B^{'} + U \tag{2.22}$$

then stacking the equation column by column, we obtain the final equation form for our Bayesian estimation

$$y^\star = \Xi b + u \tag{2.23}$$

where $\Xi = I_n \otimes X$, $b = vec\left(B^{'}\right)$ and $u = vec(U)$ with $u \sim N(0, \Sigma \otimes I_T)$.[3]

The aim of the above steps is to transform the heterogeneous linear regression to homogeneous linear regression model. With respective to (2.23), standard normal posterior distribution of b can be found with also the normal prior distribution

$$b \sim N\left(\bar{b}, \bar{V}_b\right)$$

with

$$\bar{b} = \bar{V}_b \left(\Xi^{'} (\Sigma \otimes I_T)^{-1} y^\star + \underline{V}_b^{-1} \underline{b} \right)$$

$$\bar{V}_b = \left(\Xi^{'} (\Sigma \otimes I_T)^{-1} \Xi + \underline{V}_b^{-1} \right)^{-1}$$

where prior follows $b \sim N(\underline{b}, \underline{V}_b)$. Finally transform column vector b back to matrix B.

[3] $vec(ABC) = \left(C^{'} \otimes A \right) vec(B)$

2.3.4 Draw constant matrix G_1

We still focus on equation (2.14). Place the term BX_t to the left hand side based on previous draw of B and obtain

$$y_t^\star = \bar{B}_t X_{1,t} + \bar{B}_t G_1 X_{2,t} + u_t$$

where $y_t^\star = y_t - BX_t$ and the decomposition of X_t into $X_{1,t}$ and $X_{2,t}$ is due to the structure $G = [I_r \; G_1]$. We further get the linear regression associated with $G_1$by column stacking operator towards the above equation both sides

$$y_t^{\star\star} = W_t g_1 + u_t \tag{2.24}$$

where $y_t^{\star\star} = y_t^\star - \bar{B}_t X_{1,t}$, $W_t = \left(\bar{B}_t \otimes X_{2,t}^{'}\right)$, $g_1 = vec\left(G_1^{'}\right)$ and $u_t \sim N(0, \Sigma_t)$. The same as (2.21), heteroscedasticity can be removed through dividing both sides of (2.24) by $\sqrt{\lambda_t}$. Here we skip this step, directly transform the (2.24) into a large matrix form

$$y^{\star\star} = W g_1 + u \tag{2.25}$$

where $y^{\star\star} = \left[y_1^{\star\star'}, \ldots, y_T^{\star\star'}\right]^{'}$, $W = \left[W_1^{'}, \ldots, W_T^{'}\right]^{'}$, $u_t = \left[u_1^{'}, \ldots, u_T^{'}\right]^{'}$ and $u \sim N(0, \Sigma^\star)$ with $\Sigma^\star = Blkdiag([\Sigma_1, \ldots, \Sigma_T])$.

The posterior of g_1 follows normal with variance and mean given by $\bar{V}_{g_1}$ and $\bar{g}_1$ respectively

$$g_1 \sim N(\bar{g}_1, \bar{V}_{g_1})$$

$$\bar{g}_1 = \bar{V}_{g_1}\left(W^{'}\Sigma^{\star -1} y^{\star\star} + \underline{V}_{g_1}^{-1}\underline{g}_1\right)$$

$$\bar{V}_{g_1} = \left(W^{'}\Sigma^{\star -1} W + \underline{V}_{g_1}^{-1}\right)^{-1}$$

and prior follows $g_1 \sim N\left(\underline{g}_1, \underline{V}_{g_1}\right)$.

2.3.5 Draw structural impact matrix A

To draw impact matrix A, we concentrate on equation (2.6). Take unrelated term to the left hand side, obtaining

$$A\hat{y}_t = \lambda_t^{\frac{1}{2}} S^{\frac{1}{2}} \varepsilon_t \tag{2.26}$$

where $\hat{y}_t = y_t - B_t X_t$. Since the recursive structure of A – lower triangular matrix with value of ones on the main diagonal, we can estimate (2.26) individually

$$\hat{y}_{i,t} = -\hat{y}'_{-i,t}\alpha_i + \lambda_t^{\frac{1}{2}} s_i^{\frac{1}{2}} \varepsilon_t \tag{2.27}$$

for $i = 2, \ldots, n$ and for $t = 1, \ldots, T$. $\hat{y}_{-i,t}$ collect elements from $\hat{y}_{1,t}$ to $\hat{y}_{i-1,t}$ and α_i nest corresponding row elements in matrix A. To estimate the above equation, divide by $\sqrt{\lambda_t s_t}$ both sides to remove the error heteroscedasticity

$$\hat{y}_{i,t}/\sqrt{\lambda_t s_t} = \left(-y_{-i,t}/\sqrt{\lambda_t s_t}\right)\alpha_i + \varepsilon_t$$

Stack them row by row to obtain

$$\hat{y}_i = X_{-i}\alpha_i + \varepsilon \tag{2.28}$$

where $\varepsilon \sim N(0, I_T)$.

The posterior of α_i follows normal with mean and variance

$$\alpha_i \sim N(\bar{\alpha}_i, \bar{V}_\alpha)$$

with

$$\bar{\alpha}_i = \bar{V}_\alpha \left(X'_{-i}\hat{y}_i + \underline{V}_\alpha^{-1}\underline{\alpha}_i\right)$$

$$\bar{V}_\alpha = \left(X'_{-i}X_{-i} + \underline{V}_\alpha^{-1}\right)^{-1}$$

and prior $\alpha_i \sim N(\underline{\alpha}_i, \underline{V}_\alpha)$.

2.3.6 Draw diagonal elements of S

Using equation (2.27), divide both sides by $\sqrt{\lambda_t}$ and obtain

$$\hat{y}_{i,t}/\sqrt{\lambda_t} = \left(-y_{-i,t}/\sqrt{\lambda_t}\right)\alpha + s_i^{\frac{1}{2}}\varepsilon_t$$

again stack row by row

$$\hat{y}_i = X_{-i}\alpha_i + \varepsilon \tag{2.29}$$

where $\varepsilon \sim N(0, s_i I_T)$ for $i = 2, \ldots, n$.[4]

Given the prior $s_i \sim IG(\underline{a}, \underline{b})$ where IG denotes inverse gamma distribution, the posterior follows

$$s_i \sim IG\left(\bar{a}, \bar{b}\right)$$

with

$$\bar{a} = \underline{a} + \frac{T}{2}$$

$$\bar{b} = \underline{b} + \frac{\varepsilon'\varepsilon}{2}$$

Note that the step 5 of drawing α_i can be merged into step 6 only focusing on equation (2.29).

2.3.7 Draw common stochastic volatility $\{\lambda_t\}_{t=1}^T$

Unlike b_t, the latent factor λ_t is not in a linear and gaussian state space model given other parameters and hyperparamters. Thus typical way of multiple drawing of latent variables of Carter and Kohn (1994) algorithm is no longer suitable for λ_t. Following Jacquier et.al. (1994), single drawing date by date is used for the nonlinear model. Carlin et.al (1992) show that conditional distribution of state variables in a general state space model can be written as product of three

[4] i starting from 2 is for the identification of common (latent factor) volatility. See section 2 on the model specification.

terms

$$f\left(h_t|\Theta,y^T\right) \propto f\left(h_t|h_{t-1}\right)\cdot f\left(h_{t+1}|h_t\right)\cdot f\left(y_t|h_t,\Theta\right) \tag{2.30}$$

which is the starting point of sampling of the latent factor $h_t = log\left(\lambda_t\right)$. The first two terms on the right hand side can be further written as $f\left(h_t|h_{t+t},h_{t-1}\right) \propto f\left(h_t|h_{t-1}\right)\cdot f\left(h_{t+1}|h_t\right)$.[5] Hence conditional posterior of h_t is now a product of two terms

$$f\left(h_t|\Theta,y^T\right) \propto f\left(h_t|h_{t-1},h_{t+1}\right)\cdot f\left(y_t|h_t,\Theta\right) \tag{2.31}$$

That is the target distribution from which draws of h_t come.

Nonlinearity of the state space model on measurement equation makes the posterior target does not have an analytical form, therefore metropolis algorithm is required. We choose $f\left(h_t|h_{t-1},h_{t+1}\right)$ as proposal since it is part of target distribution with the same support and the most importance is that the proposal is analytical due to that the law of motion of the common factor is linear and gaussian. The basic idea is that one can regard h_t as paramters to be estimated, h_{t+1} as the obserbation driven by h_t and h_{t-1} as prior information about h_t. This can be seen explicitly in successive two periods

$$h_t = F\cdot h_{t-1} + \eta_t$$

$$h_{t+1} = F\cdot h_t + \eta_{t+1}$$

from the first equation one can set prior for $h_t \sim N\left(F\cdot h_{t-1}, Q_h\right)$ while in the second equation h_t is coefficient to be estimated and $h_{t+1}$is an observation. From the standard Bayesian estiamtion of linear regression model, one is familiar with

- For $t = 1,\ldots,T-1$,

$$h_t \sim N\left(Fh_{t-1}, Q_h\right)$$

[5] $f\left(h_t|h_{t-1},h_{t+1}\right) = \frac{f(h_{t-1},h_t,h_{t+1})}{f(h_{t-1},h_{t+1})} = \frac{f(h_{t+1}|h_t,h_{t-1})\cdot f(h_t|h_{t+1})}{f(h_{t+1}|h_{t-1})} \propto f\left(h_{t+1}|h_t\right)\cdot f\left(h_t|h_{t-1}\right)$

$$h_t|h_{t-1},h_{t+1} \sim N(u_h, Q_h) \tag{2.32}$$

with

$$u_h = V_h\left(F^{'}Q_h^{-1}h_{t+1} + Q_h^{-1}Fh_{t-1}\right)$$

$$V_h = \left(F'Q_h^{-1}F + Q_h^{-1}\right)^{-1}$$

- For $t = 0$,

$$h_0 \sim N(\underline{u}_h, \underline{V}_h)$$

since

$$h_1 = F \cdot h_0 + \eta_1$$

then

$$h_0 \sim N(\bar{u}_h, \bar{V}_h) \tag{2.33}$$

$$\bar{u}_h = \bar{V}_h\left(F^{'}Q_h^{-1}h_1 + \underline{V}_h^{-1}\underline{u}_h\right)$$

$$\bar{V}_h = \left(F^{'}Q_h^{-1}F + \underline{V}_h^{-1}\right)^{-1}$$

- For $t = T$,

$$h_T|h_{T-1} \sim N\left(u_{T,h}, V_{T,h}\right) \tag{2.34}$$

$$u_{T,h} = V_{T,h}\left(Q_h^{-1}Fh_{T-1}\right)$$

$$V_{T,h} = \left(F^{'}Q_h^{-1}F + Q_h^{-1}\right)^{-1}$$

With the above proposals at hand (2.32) - (2.34), the date by date independence metropolis is implemented as follows

1. When $t = 0$, draw h_0 from (2.33) given prior for h_0.

2. When $t = 1,\ldots,T-1$, a) draw a candidate $h_t^{\star}$ from (2.32); b) compute the acceptance ratio

(probability) $r = \min\left(\frac{f(y_t|h_t^\star,\Xi)}{f\left(y_t|h_t^{old},\Xi\right)},\ 1\right)$ where $f(y_t|h_t,\Xi)$ is the likelihood of the observation t; c) for each $h_t^\star$, draw a value u from the Uniform (0,1) distribution. If $u \leq r$, accept $h_t^\star$ as h_t^{new}. Otherwise, still keep h_t^{old} in period t.

3. When $t = T$, a) draw a candidate $h_T^\star$ from (2.34); b) compute the acceptance ratio (probability) $r = \min\left(\frac{f(y_T|h_T^\star,\Xi)}{f\left(y_T|h_T^{old},\Xi\right)},\ 1\right)$ where $f(y_T|h_T,\Xi)$ is the likelihood of the observation T; c) for each $h_T^\star$, draw a value u from the Uniform (0,1) distribution. If $u \leq r$, accept $h_T^\star$ as h_T^{new}. Otherwise, still keep h_T^{old} in period T.

4. Repeat step 1 to step 3 date by date in each iteration.

We complete the drawing of h_t from $t = 1$ to $t = T$. Note that though there is no explicit backward smoother procedure as in the algorithm of Carter and Kohn (1994) for linear gaussian model, the draws of h_t are stll based on the information set of all the observations as they directly come from posterior condition $f\left(h_1,\ldots,h_T|y^T,\Xi\right)$ instead of two seperate steps of forward Kalman filter and a backforward smoother with the analytical expression in linear gaussian model.

2.3.8 Draw coefficient F

Give the previous draws of $\{h_t\}_{t=1}^T$with equation (2.5)

$$h_t = F \cdot h_{t-1} + \eta_t$$

where $\eta_t \sim N(0, Q_h)$. By typical column stacking implementation

$$y_h = F \cdot x_h + \eta \tag{2.35}$$

the posterior of F has

$$F \sim N\left(\bar{u}_F, \bar{Q}_F\right)$$

with

$$\bar{u}_F = \bar{Q}_F \left(Q_h^{-1} x_h^{'} x_h + \underline{Q}_F^{-1} \underline{u}_F \right)$$

$$\bar{Q}_F = \left(Q_h^{-1} x_h^{'} y_h + \underline{Q}_F^{-1} \right)^{-1}$$

where prior $F \sim N\left(\underline{u}_F, \underline{Q}_F\right)$.

2.3.9 Draw variance Q_h

Still focusing on equation (2.35) with prior distribution of inverse gamma $Q_h \sim IG(\underline{c}, \underline{d})$, the posterior becomes

$$Q_h \sim IG\left(\bar{c}, \bar{d}\right)$$

with

$$\bar{c} = \underline{c} + \frac{T}{2}$$

$$\bar{d} = \underline{d} + \frac{\eta^{'} \eta}{2}$$

The above nine steps or nine blocks consist of one iteration in Bayesian estimation via Markov chain Monte Carlo (MCMC). The MCMC simulation involves Gibbs sampling and Metropolis Hasting algorithm. After discarding some burn-in iterations, the draw of parameters and hyperparameters from each conditional posterior is equivalent to one from joint posterior from which bayesain inference can be conducted.

The characteristic of the model is that we use factor idea on lags, coefficients and volatility to make the TVP-VAR model parsimonious. We use first several lags to drive long lags that are identified by G matrix, see (2.8); We use a latent factor, namely the common volatility to represent the volatilities of the whole observations that is identified by S matrix, see (2.4) and finally several factors could drive the latent coefficients by decomposing covariance matrix Q_b into reduced rank that can be seen in (2.10) and (2.11). The difference of our model with conventional factor models or factor augmented models is that we focus on reducing dimension of parameters rather than on

dimension of observations in a rich data set.

In the next section, we apply this model to monetary policy analysis to evaluate the model performance and make some inferences.

2.4 Empirical analysis

In this section, we first analyse whether there are enough evidences for factor driving of the dynamic process of the time varying coefficients in the TVP-VAR model. We apply the general factor-driven model to a typical small scale monetary VAR. We however still find strong evidences in supporting factor driving. Then we turn to structrual analysis on the agents' response to monetary policy shock. No significant difffences are found.

The same data is used as in Chapter 1. It contains three variables, namely inflation rate, unemployment rate and short rate of 3-month treasury bill rate which cover the period from 1953 Q1 to 2006 Q3 before the 2008 financial crisis after that unconventional monetary policy was conducted. The reason we choose this data set is threefold. First, it is a good description of workings of the economy in real activity, nominal variable and monetary policy that explicitly correspond to IS curve, Philips Curve and policy rule in a typical small scale DSGE model. Thus this data set is widely used in policy and business cycle analysis and also a subset of medium or large scale data set for such analysis; Second, due to the well known problem of parameter proliferation in TVP-VAR with or without SV, researchers usually tend to use such kind of small data set to weeken this concern in order to make their findings or/and conclusions belivable. Based on this arrangement, we try to ask is there still over-parameterization in this small scale model they specially choose and if it is, is it strong enough; Lastly, since the data is widely used, it is convenient for us to conduct comparison with extant literature.

2.4.1 Are factors important in time varying parameters?

In this subsetion, we test whether there is enough evidence supporting over fitting in the small scale TVP-VAR. That is the precondition and starting point for our factor-based model. The test is based on two models that all impose no factor restriction on time varying coefficients on regressors, namely let the coefficients freely fluctuate. The only difference is that one with full stochastic volatility (Primiceri, 2005) and the other common volatility (Carriero et al., 2012). Including stochastic volatility in the model is because shutting off the volatility channel probably will cause misunderstanding of the dynamic process.[6] If the test is implemented in a model that can potentially cause the amount of time variation originating from the part of volatiliy incorrectly transferring to the part of time varying coefficients on regressors – that is the time varying coefficients now have more variation than they should have, even though we find evidence on over parameterization, the result is still questionable. Testing results from both models with SV are very similar, here we only present the results from the model with common volatility for the save of space.[7]

The TVP-VAR with common volatility (TVP-VAR-CV for short hereafter) is implemented with one lag to four lags with common volatility respectively. We conduct pincipal component analysis of the covariance matrix of innovations to time varying coefficients Q_b in each model. Table 2.1 and Table 2.2 give the contribution of each factor for each model with lag from one to four. They are given in terms of mean and median in percentage in descending order. Some figures do not list in the tables because contributions are already too small.

In the model with only 1 lag, it has only 12 time varying coefficients each period and therefore the most less over-fitting model. Principal component analysis shows that factor driving is still very obvious and can not be ignored. From Table 2.1 under the Lag 1 title, we find that the first factor

[6] Consider a AR(1) process $y_t = \rho y_{t-1} + u_t$ where $u_t \sim N\left(0, \sigma^2\right)$. The unconditional variance of y_t is $\text{var}\left(y_t\right) = \frac{\sigma^2}{1-\rho^2}$. If the volatility σ_t is increasing with time, but the model is estimated on time varying coefficient ρ_t with constant σ, we will find the persistent ρ_t of the model is increasing that contradicts the reality.

[7] The empirical test on the TVP-VAR with full SV is also conducted. The results are almost exactly the same as common SV which confirms the finding in Carrero et al. (2012) that for the typical U.S. data set, common SV is a good representation for full SV.

TVP-VAR-CV	Lag 1		Lag 2	
Principal component	Mean %	Median %	Mean %	Median %
1	63.8684	63.5164	96.0390	96.1510
2	18.2649	18.9576	1.6741	1.0047
3	11.4011	11.0741	0.5968	0.7234
4	3.0299	2.9678	0.4953	0.4322
5	1.2880	1.3223	0.3128	0.2725
6	0.9465	0.9347	0.2257	0.2507
7	0.5503	0.5536	0.1662	0.2011
8	0.3539	0.3742	0.1353	0.1691
9	0.2276	0.2251	0.0903	0.1464
10	0.0463	0.0457		
11	0.0146	0.0213		
12	0.0085	0.0071		

Table 2.1: Contribution for each factor via principal component analysis on covariance matrix Q_b for TVP-VAR-CV from lag one to lag two

TVP-VAR-CV	Lag 3		Lag 4	
Principal component	Mean %	Median %	Mean %	Median %
1	95.1109	93.5693	93.7272	91.1517
2	1.8860	1.7230	3.1857	2.3241
3	0.8257	1.0010	0.9423	1.1121
4	0.5718	0.5718	0.4749	1.0458
5	0.3397	0.5330	0.4012	0.6961
6	0.2658	0.4804	0.2641	0.6427
7	0.1683	0.3620	0.2270	0.4846
8	0.1612	0.3288		
9				
10				
11			0.0712	0.2264
12	0.0728	0.1361		

Table 2.2: Contribution for each factor via principal component analysis on covariance matrix Q_b for TVP-VAR-CV from lag three to lag four

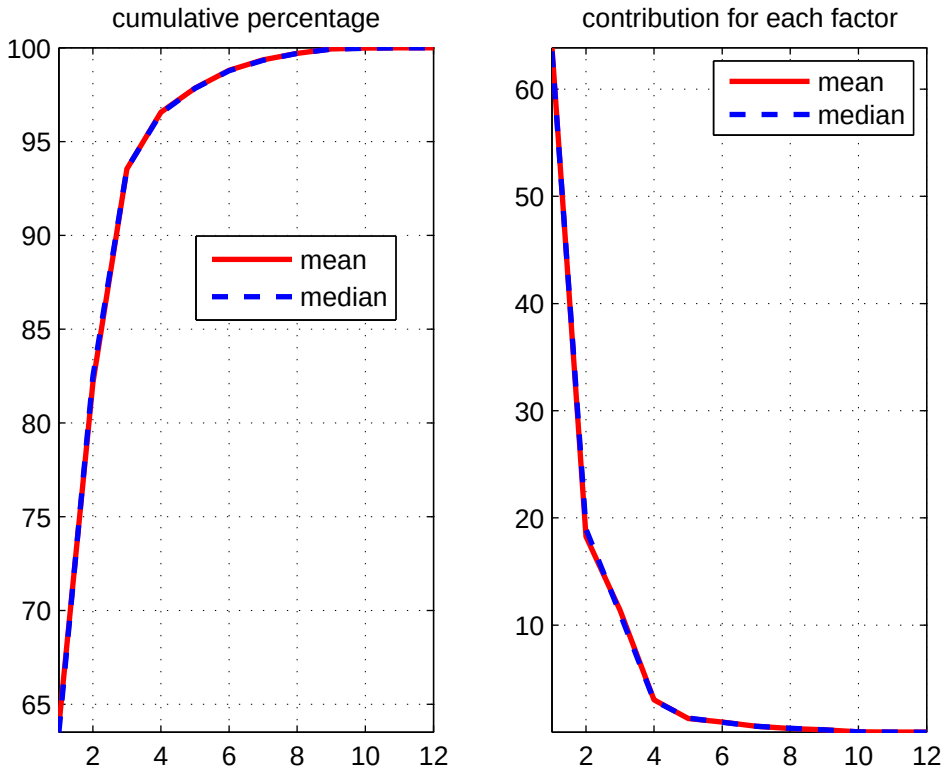


Figure 2.1: Contribution and cumulative contribution for lag 1

Notes: This is TVP-VAR-CV with 1 lag and 12 coefficients on regressors each period. The left panel plots cumulative contribution for all the factors via principal component analysis of covariance matrix Q_b. The right panel plots contribution for each factor in descending order via principal component analysis of covariance matrix Q_b. The solid line represents posterior mean, while dashed line posterior median.

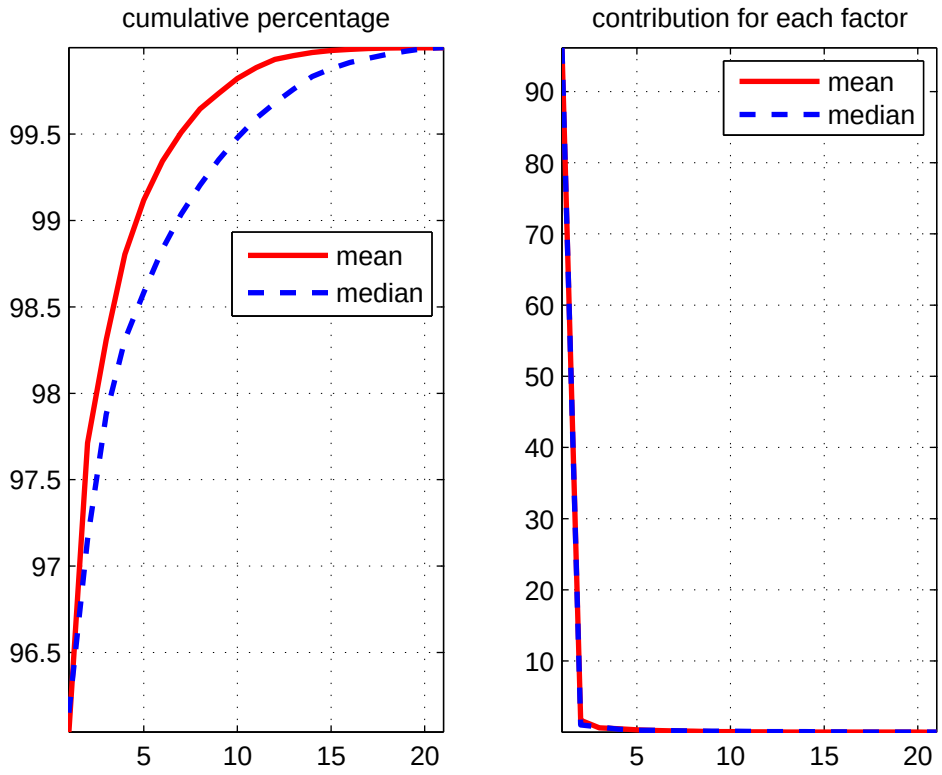


Figure 2.2: Contribution and cumulative contribution for lag 2

Notes: This is TVP-VAR-CV with 2 lag and 21 coefficients on regressors each period. The left panel plots cumulative contribution for all the factors via principal component analysis of covariance matrix Q_b. The right panel plots contribution for each factor in descending order via principal component analysis of covariance matrix Q_b. The solid line represents posterior mean, while dashed line posterior median.

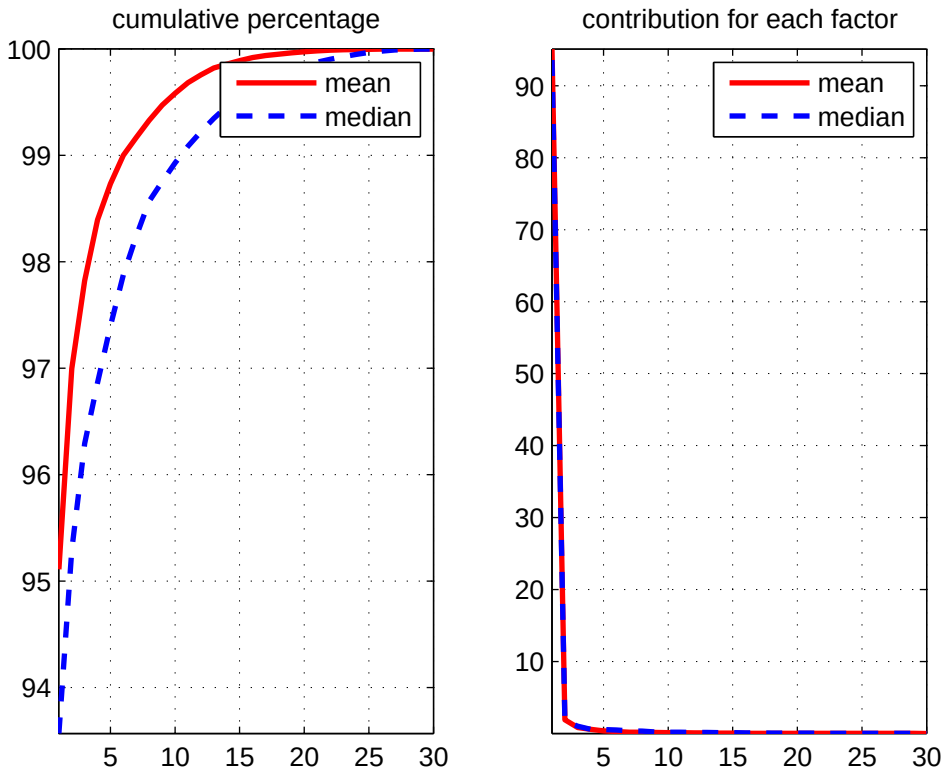


Figure 2.3: Contribution and cumulative contribution for lag 3

Notes: This is TVP-VAR-CV with 3 lag and 30 coefficients on regressors each period. The left panel plots cumulative contribution for all the factors via principal component analysis of covariance matrix Q_b. The right panel plots contribution for each factor in descending order via principal component analysis of covariance matrix Q_b. The solid line represents posterior mean, while dashed line posterior median.

accounts for 64% contribution of the variation of all coefficients. The second takes 18% much less than the first and the third 11% with a moderate amount of interpretation. This dynamic process can be found in Fig 2.1 in which the left panel gives the cumulative contribution and the right panel shows the contribution for each factor where solid line denotes mean and dashed line median. It is easy to observe that the first three factors, all above 10%, together account for near 95% and the first six factors together almost 100%. After the sixth factor, all the remaining factors with very little explanation can be ignored which is evidenced by the right panel of Fig 2.1. Generally speaking, the samll scale model with the shortest lag indicates that it is still very over fitting, at least half of the principals should be discarded.

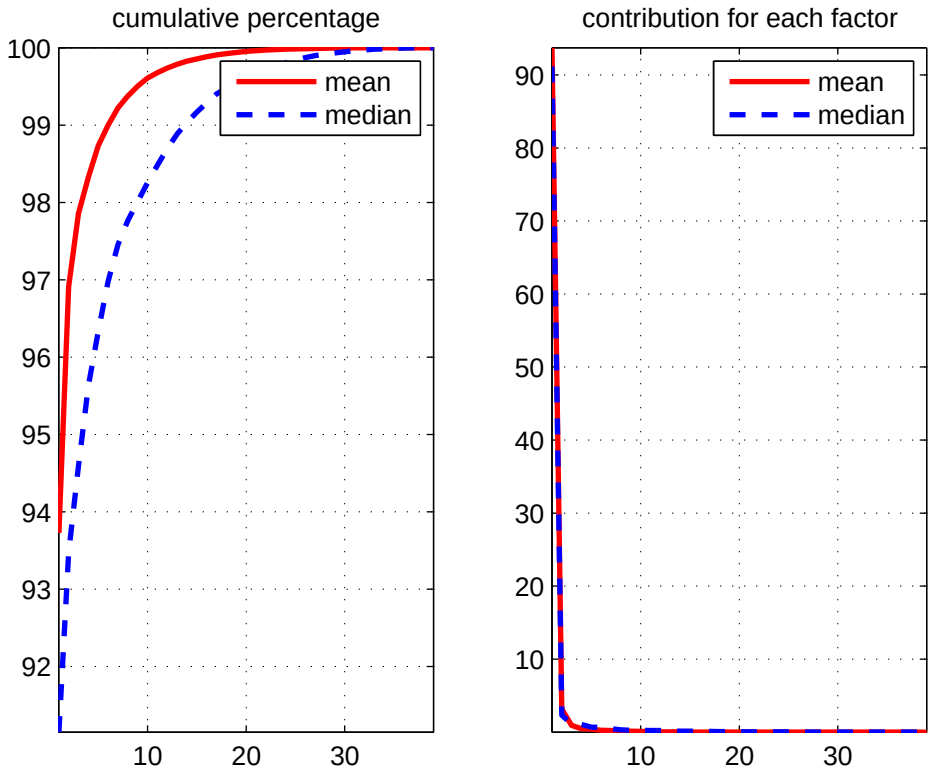


Figure 2.4: Contribution and cumulative contribution for lag 4

Notes: This is TVP-VAR-CV with 4 lag and 39 coefficients on regressors each period. The left panel plots cumulative contribution for all the factors via principal component analysis of covariance matrix Q_b. The right panel plots contribution for each factor in descending order via principal component analysis of covariance matrix Q_b. The solid line represents posterior mean, while dashed line posterior median.

Let's go to two lags. The two columns under the title of Lag 2 strongly suggest that, in Table2.1, the first factor is large enough with 96% to interpret almost all the variation in the part of time varying coefficients, while the second factor, in contrast to the model of one lag, dramatically decline to less than 2%. This can be evidenced by the right panel of Fig 2.2. After the second factor, all the remaining factors have contributions close to zero. When the model is set to three lags, the result is very similar that the first accounts for more than 95%, the second less than 2% and from the third onward, all factors are very near zero, see Fig 2.3 intuitively. The model of four lags is also the same case no matter in mean or median even though the number of factor candidates raising to 39, which can be directly observed in Fig 2.4. Models of two, three and four lags share the same property that only the first factor account for most and two or three together for almost all the variation, while the model with one lag needs more factors where the tangent of cumulative percentage curve in Fig 2.1 is less steeper than others in Fig 2.2 to Fig 2.4.

At least three findings and implications can be derived. The first, there are strong evidence of factor driving of dynamic process in time varying coefficients no matter in the model of one lag or of four lags. This indicates that in a typical model settings of TVP-VARs, the number of dynamic sources, also the same number of coefficients should be dramatically reduced.

The second, following the first, it further confirms researchers' concern that TVP-VAR is not a parsimonious model even it is set in a small scale with short lags which are in line with Cogley and Sargent (2005) and De Wind and Gambetti (2014). Small scale TVP-VAR models still need factor-driven estimation.

The third, we find significant difference in the style of factor driving in model of one lag and model of more lags. In the context of the typical three observations, one-lag model needs more factors than more-lag models and with the increase of the number of lags, the style of (almost) one factor leading is consistent from two-lag model to four-lag model.[8] The cosistency of one factor driving means that the sources of variation in the model have been already fully identified, implying that more than enough lags will cause serious over fitting. A reasonable explanation is

[8]We also tried lags more than four, the results are quantitatively the same.

that if a model does not have enough lags, its dynamic inter relationship among variables could be messed up and the factors extracted from it distribute on a broad range (see again the left panel of Fig 2.1) such that economic implication is difficult to give, displaying merely some purely statistic properties. Nevertheless, if with enough lags, two lags above, the inter relationship can be fully released and expressed, therefore the factor driving analysis could imply some important structural interpretation.[9] If more than enough, since important factors have been fully extracted, naturally serious over fitting will arise, comparing Fig 2.4 with Fig 2.2 on the right panel.

The above three variable empirical test consistently demonstrates that almost only one important force influence the dynamics of the economy, confirming again that there are much more forces determining the economy than those changing the economy. The result of empirical test is in line with dynamic stochastic general equilibrium model (DSGE). If a DSGE model can be transfomed into a VAR form under some conditions (Fernandez-Villaverde et al. 2006, Morris, 2012 and Ravenna, 2007), the coefficients on the regressors must be the functions of 'deep parameters' of preference, technoloy and policy rule in the DSGE model and therefore the coefficients in the transformed VAR are cross equation restricted.[10] Changes in one deep parameter will cause changes in almost all the coefficents in the VAR. This provides theoretical support for the results of the above empirical test and recommend a factor-driven VAR model setting.

Since we find strong factor leading evidence, even in a small scale TVP-VAR, empirically and theoretically, another question will be naturally asked can the model with factor driven specified in section 2 be used in structrual economic analysis and is it consistent or inconsistent with the literature. We answer these questions in the next subsection.

[9]The 'structural interpretation' does not mean structural identification associated with structural shocks in VARs. It only mean several important factors already sufficiently drive the dynamic process of the economy.

[10]A general representation of a log-linearized DSGE model has a state space form. A state space corresponds to a VARMA form (see, e.g., Aoki, 1990). But the VARMA does not necessarily can be transformed to a VAR form. Fernandez-Villaverde et al. (2006) give the VAR(∞) expression for a DSGE under some conditions; Ravenna (2007) for VAR(p) and Morris (2012) for VAR(1)under required conditions. Giacomini (2013) give a good literature review on the relationship between DSGE and VAR models.

2.4.2 Structural analysis and model evaluation

The above empirical tests for the three variables model - inflation rate, unemployment rate and 3-month treasury rate - suggest that factors should be imposed on the time varying coefficients and two lags are enough for fully capturing the dynamic process in the data and therefore the time variation in the latent coeffficients. Since the empirical tests have already found that the factor leading style for the first lag is quite different with the style for from two to four lags, we can not use early lags – the first lag and constant here – to drive the remaining long lags. Hence the factor settings on lags should be skipped under this empirical context.[11]

We implement the analysis in a model setting factors only on time varying coefficients and stochastic volatility. That is, we estimate the model only involving factors on coefficients and volatility, imposing no restriction on lags. As discussed in section 3, though the model gives a general factor treatment to every part of TVP-VARs, where the factor restriction should be used depends on the specific research object.

In this section we have two purposes. One is to investigate whether agents' responses to monetary policy shocks have changed or not over sample period and the second is to evaluate the model performance under different lag settings. The first is to check after factor extracting whether the model is still capable of structural analysis and the second is to check whether the model is flexible enough to deal with the case when the extent of over fitting goes strong as the number of lags incease.

These two are jointly conducted together. Our analysis is based on above empirical tests. Dates chosen for comparison are 1975 Q1, 1981 Q3 and 1996 Q1. They are somewhat representative of the typical economic conditions of the chairmanships of Burns, Volcker and Greenspan, but apart from that, they are choosen arbitrarily. Enough number of factors is chosen for each model according to Table 2.1 and Table 2.2 so as to avoid the potential concern that the bulk of variation not fully captured by less than enough factors could affect inference and might lead wrong understanding.

[11] We have estimated the model of the factor settings on lags, namely the general factor-driven model presented in section 2. The result is messy to conduct instructive analysis. A reasonable interpretation is that the dynamic inter relationship among the variables has been distorted by the the constant and the first lag.

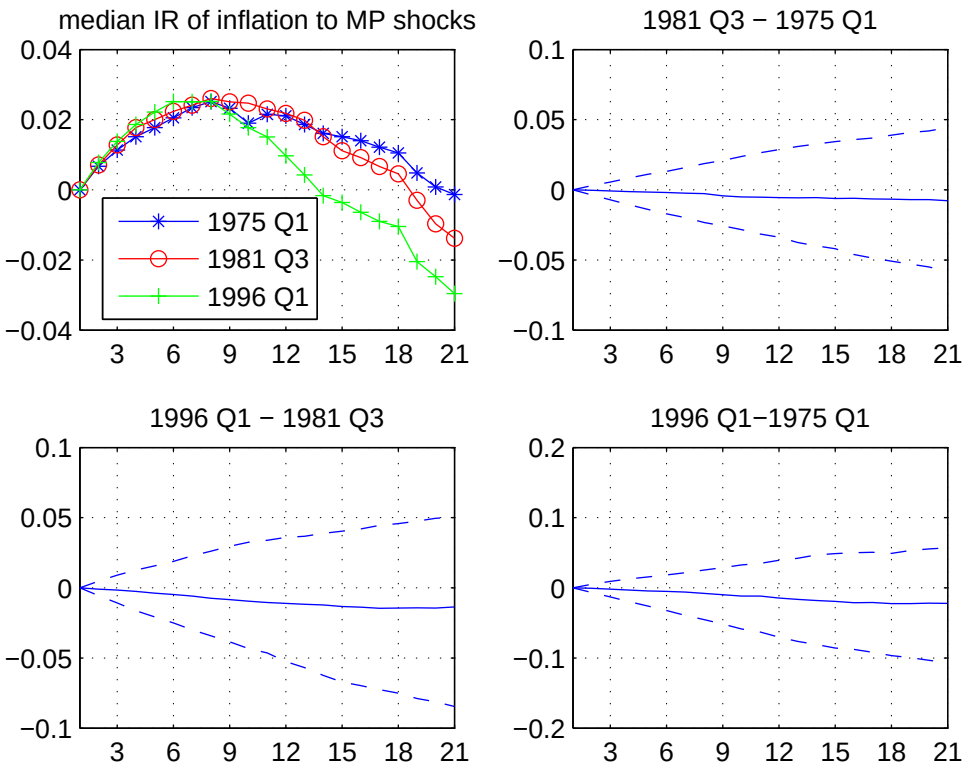


Figure 2.5: TVP-VAR-CV-lag1-factor7-IR-inflation

Notes: This is TVP-VAR with 1 lag, 7 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

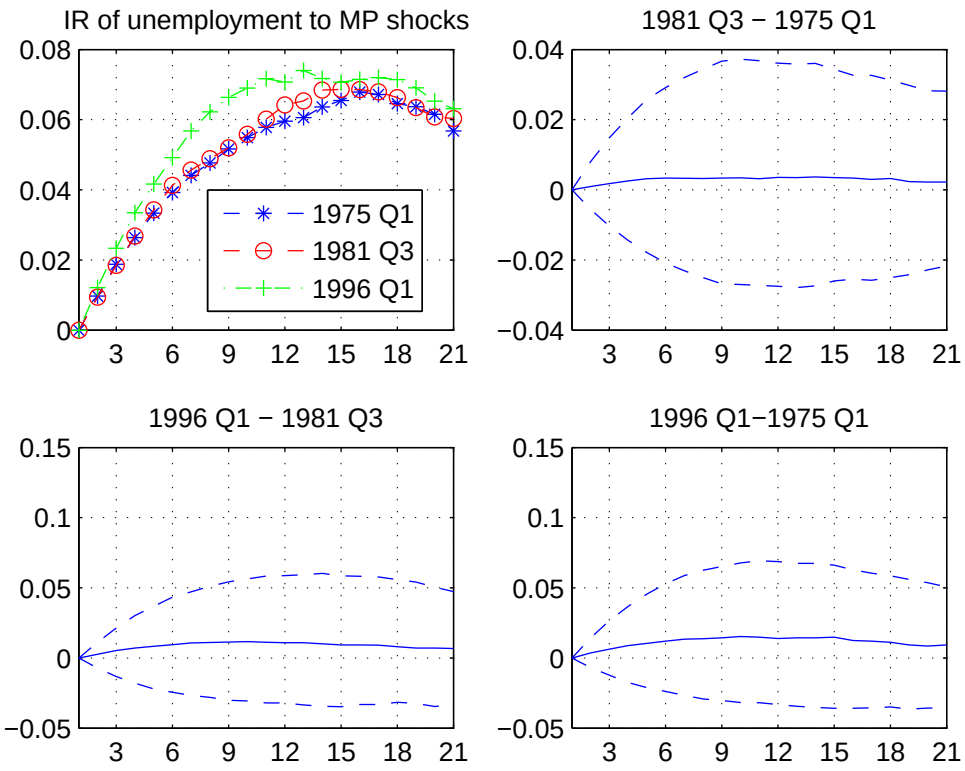


Figure 2.6: TVP-VAR-CV-lag1-factor7-IR-unemployment rate

Notes: This is TVP-VAR with 1 lag, 7 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

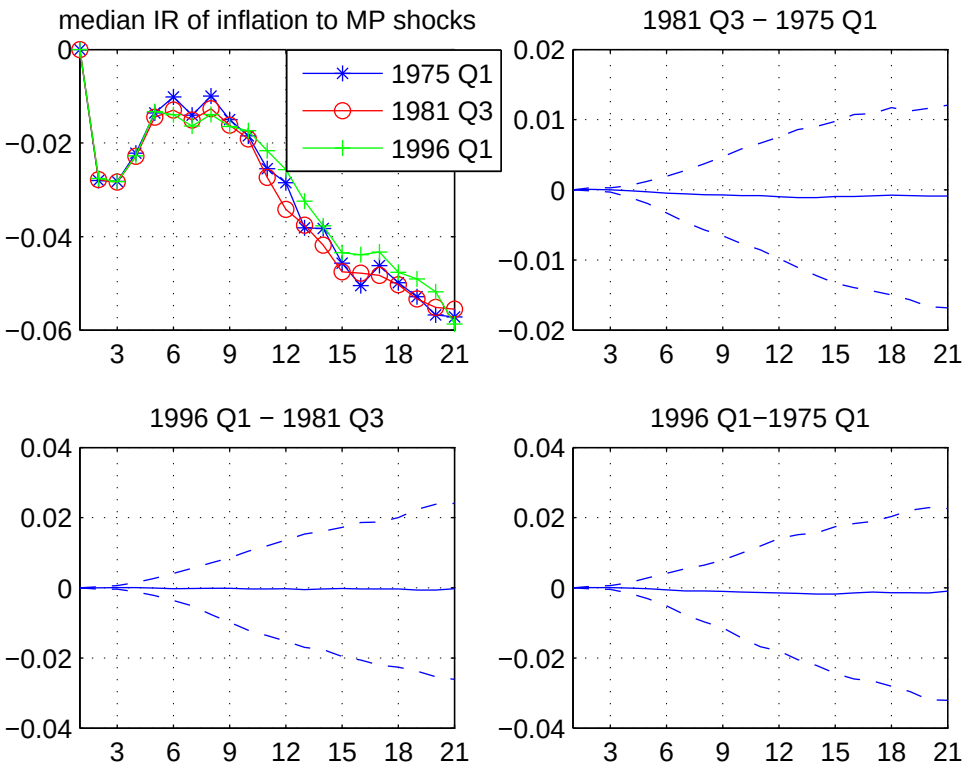


Figure 2.7: TVP-VAR-CV-lag2-factor3-IR-inflation

Notes: This is TVP-VAR with 2 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

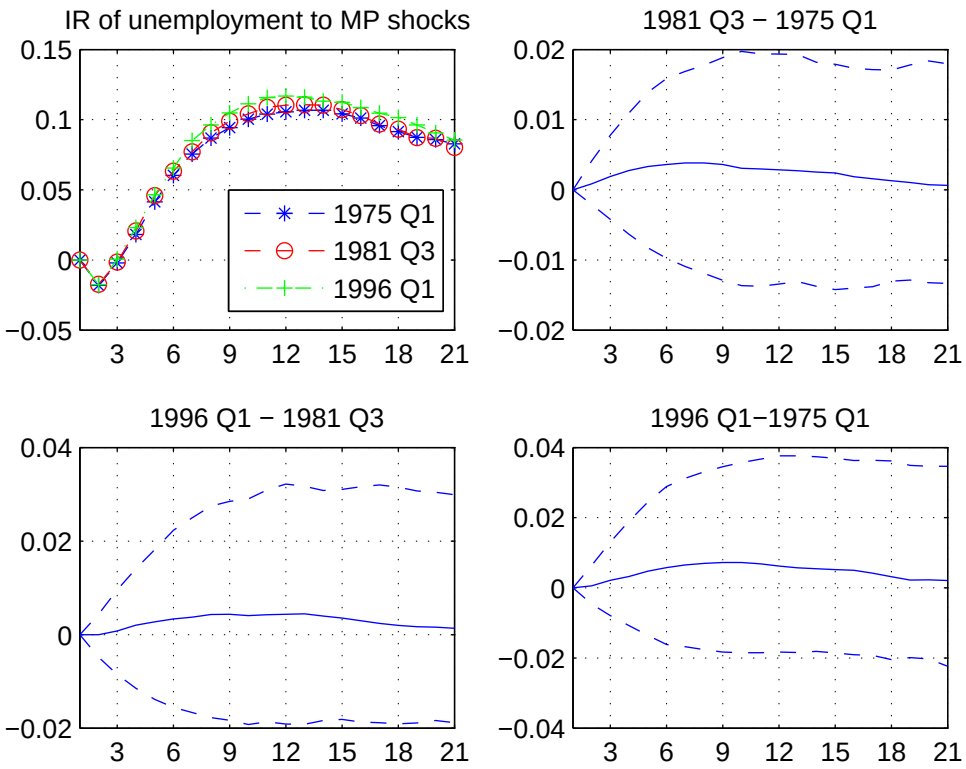


Figure 2.8: TVP-VAR-CV-lag2-factor3-IR-unemployment rate

Notes: This is TVP-VAR with 2 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

For model of one lag, we choose 7 factors since 7 factors can cover all the important factors that together accounting for near 100% evidenced by Table 2.1 and Fig 2.1. In Fig 2.5 upper left panel, we find strong 'price puzzle' in each period respectively, a typical finding in small scale monetary VAR. The remaining three graphs show the median of the distribution of the difference of inflation responses between every two periods mutually with lower 16^{th} percentile and upper 84^{th} percentile confidence band. There are no significant difference among three periods in that every cofidence band contains a zero line. The response of unemployment rate increase, see Fig 2.6, in each period consistent with economic theory. The difference of unemployment responses between every two periods are not significant with zero. As for the model with two lags, we choose 3 factors in order to cover as much time variation as possible even though the first factor has already taken 96% seen in Table 2.1. The price puzzle disappears, in Fig 2.7, for each period; Inflation decline after positive monetary policy shock. The difference of inflation responses every two periods is not significant in the remaining panel. Unemployment rate increase in each period and no significant difference mutually in Fig 2.8. When the model goes from 2 lags to 3 lags with the same 3 factors as they have the same factor driving style, we find that, in both Fig 2.9 and Fig 2.10, the results are almost the same: inflation went down and unemployment went up each period, and the dynamic process in each response function is significantly no difference. Though the model is of more over-parameterization, the parsimonious estimation of factor driving gives the qualitatively and quantitatively the same results. How about the model of four lags when it become further over-fitting with 39 coefficients? We assign again 3 factors and still find the qualitatively and quantitatively the same results in Fig 2.11 and Fig 2.12.[12]

The common volatility for each model in Fig 2.13 illustrates almost the same dynamic property of the U.S. business cycle. The level of commom volatility was climbing during the 1970s up to the peak in early 1980s during which Fed Chairman Volcker implemented monetary targeting policy, after that from middle 1980s the volatility declined and stayed on a low level in a whole 1990s and

[12]We assign 3 factors to the portion of time varying coefficients for models of 3 lags and 4 lags in order to make sure most amount of variation can be covered. Actually, one factor for 2 lags, 3 lags and 4 lags already works well respectively.

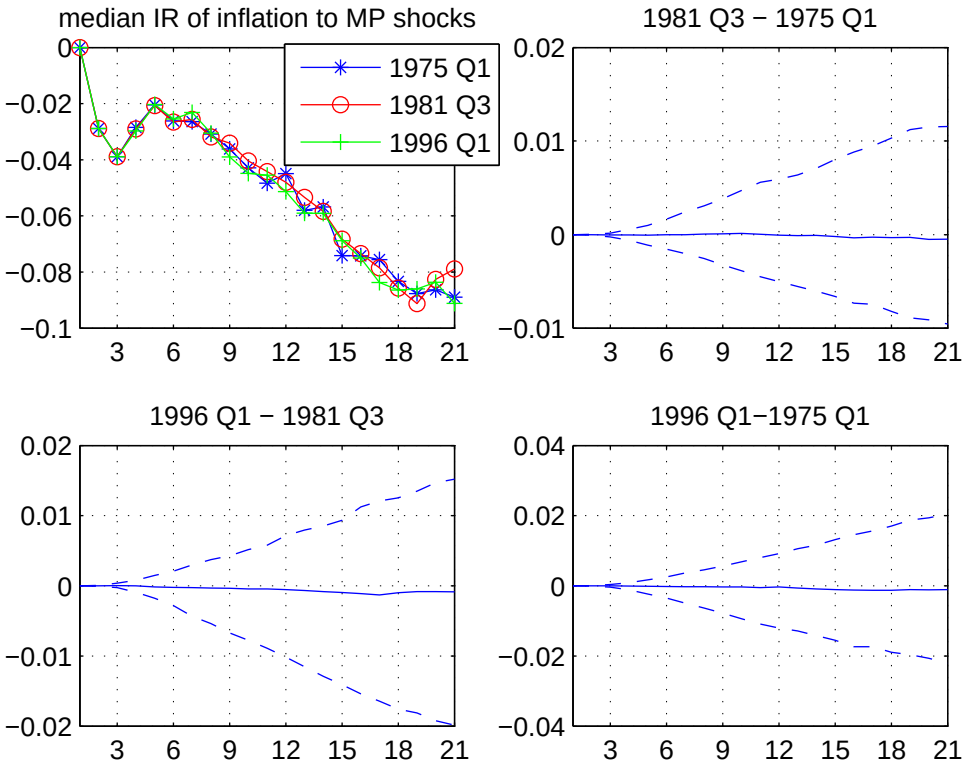


Figure 2.9: TVP-VAR-CV-lag3-factor3-IR-inflation

Notes: This is TVP-VAR with 3 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

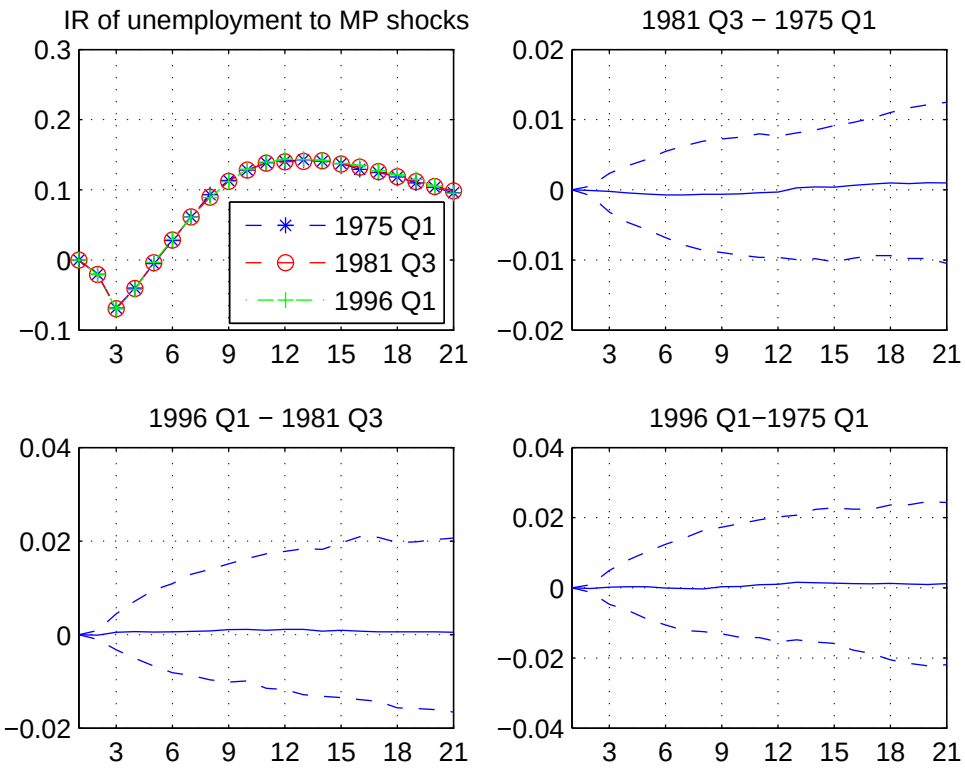


Figure 2.10: TVP-VAR-CV-lag3-factor3-IR-unemployment rate

Notes: This is TVP-VAR with 3 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of unempoyment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

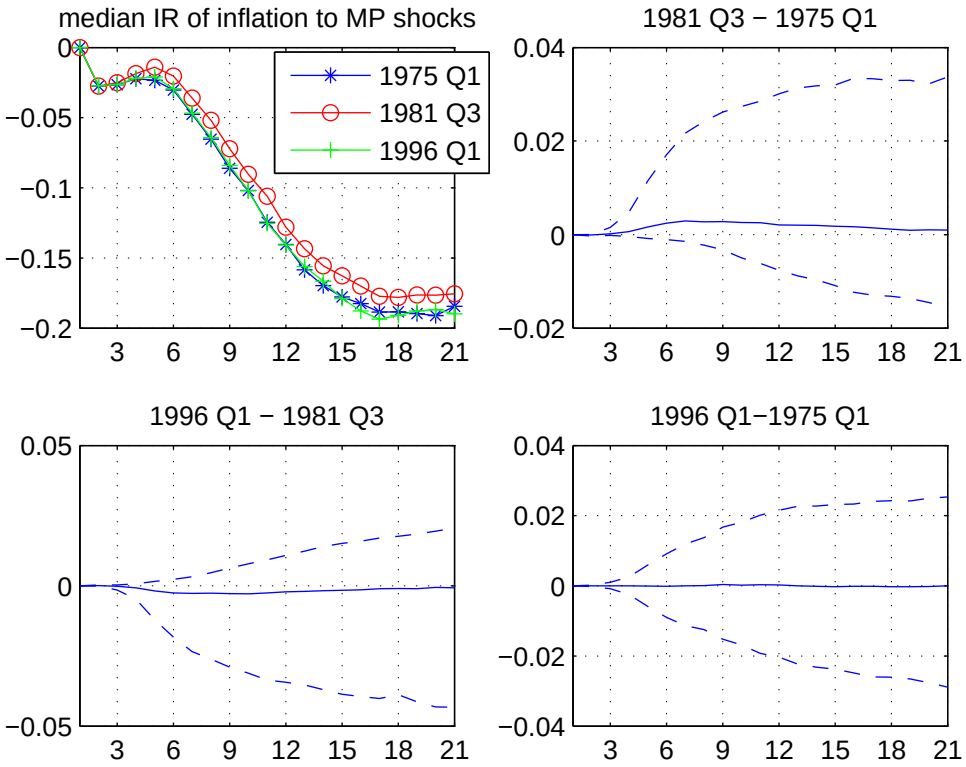


Figure 2.11: TVP-VAR-CV-lag4-factor3-IR-inflation

Notes: This is TVP-VAR with 4 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of inflation to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

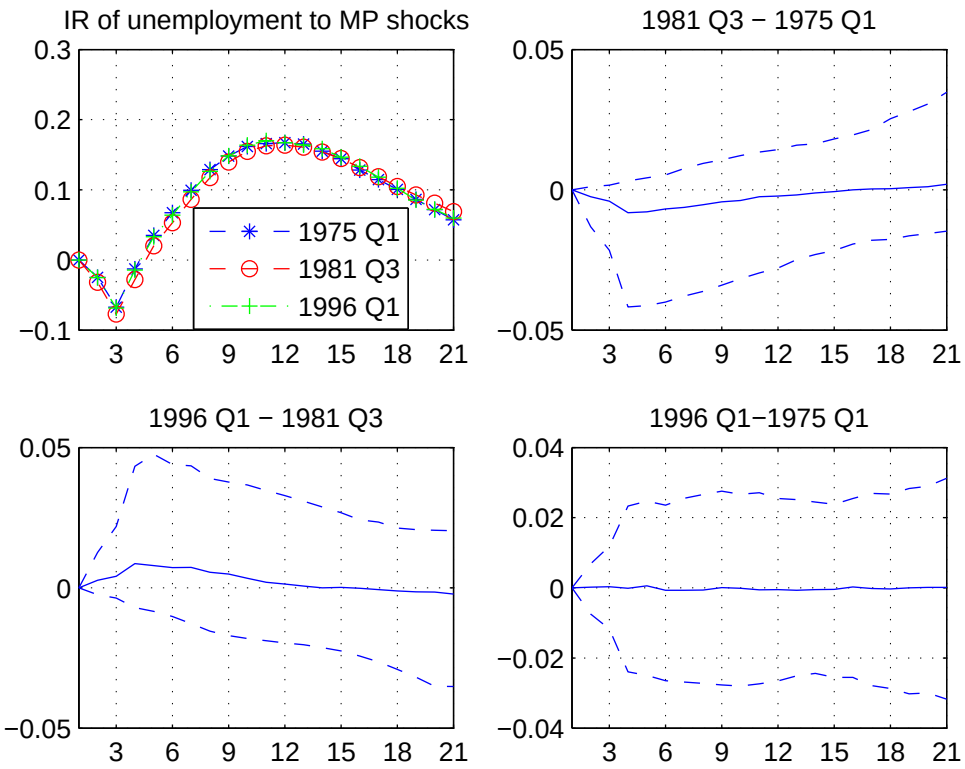


Figure 2.12: TVP-VAR-CV-lag4-factor3-IR-unemployment rate

Notes: This is TVP-VAR with 4 lag, 3 factors driving time variation in coefficients on regressors and common volatility. The upper left panel plots median impulse responses of unemployment rate to monetary policy shocks in 1975 Q1, 1981 Q3 and 1996 Q1. The remaining three panels plot difference of responses between every two periods mutually, with solid line representing median and dashed lines for 16^{th} percentile and 84^{th} percentile respectively.

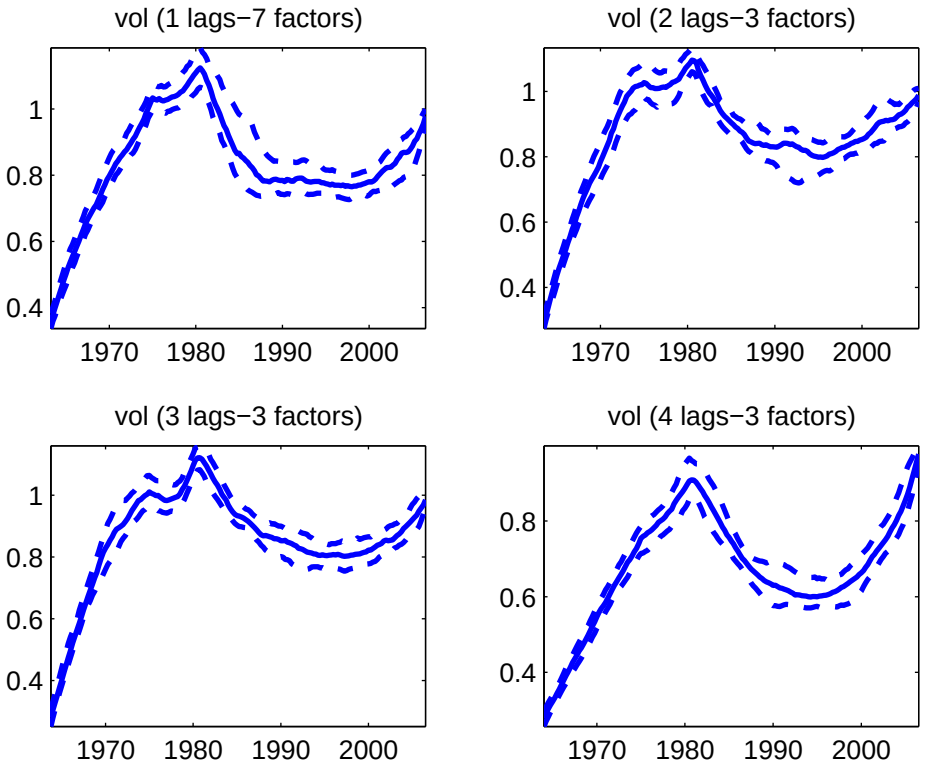


Figure 2.13: Common volatility in different model settings

early 2000s, then gradully increased at the end of the sample.

Connecting the above analysis together, we find that i) in the model of one lag even with enough 7 factors capturing almost all the variation in coefficients, we still find price puzzle. As discussed in empirical test via principal component analysis, one lag can not sufficiently release and probably distort the dynamic inter relationships among variables. This is a problem of lags, not a problem of factor driving. Therefore, using the first lag to drive the other lags is not a good choice. ii) When having enough lags and even more lags to make the model more over parameterization, the factor driving model works very well on every setting, giving almost the same results on impulse response function and volatility. All these fully demonstrate that the factor-driven model is flexible enough to capture main driving forces and give consistent interpretation in the context of the model with over-fitting problem, especially when the extent of over parameterization is very serious (the model with 4 lags).

2.5 Concluding remarks

In this chapter, we present a model that gives a general treatment via factor driving on the three parts of conventional TVP-VAR model, namely, the part of time varying coefficients on regressors, the part of lags and the part of stochastic volatility. These three portions are sources of parameter proliferation. The general model uses several factors to drive the dynamic process of all the time varying coefficients; uses the first several lags to drive other remaining lags; and uses a latent factor to drive the volatilities of all variables, i.e. the common volatility. These factor drivings can be jointly or separately used depending on specific object one meets.

We also provide a Bayesian estimation procedure on this general model step by step, which can be nested, divided and modified according to the factor driving portion one needs.

We finally conduct an empirical analysis on the typical small scale monetary TVP-VAR. Strong evidences of over parameterization are found over one lag to four lags. This is the starting point of our general parsimonious treatment of the standard TVP-VAR model. Empirical analysis shows that the factor-driven model is strong enough and flexible enough to capture the main driving forces even in a serious over-fitting setting and still give consistent results.

A point should be pointed out is that the factor-idea-based model must stand on the precondition that the dynamic process among data set is correctly expressed. If not, the model is estimable, but not instructive.

The future research could be applying the model to large data set or conducting forecasting exercise.

Appendix

A. Bayesian estimation of restricted linear regression model

Here, we consider a multivariate linear regression case.

Consider that the linear regression model has the form:

$$y_t = X_t\beta + u_t$$

where y_t is an $n \times 1$ vector of regressands, X_t is $n \times k$ matrix of regressors, β is $k \times 1$ corresponding parameters and $u_t \sim N(0, \Sigma_t)$. Stacking the above equation over time periods and holding all the data together, one can obtain

$$y = X\beta + u$$

where $y = \left[y_1^{'}, \ldots . y_T^{'}\right]^{'}$, $X = \left[X_1^{'}, \ldots, X_T^{'}\right]^{'}$ and $u = \left[u_1^{'}, \ldots, u_T^{'}\right]^{'}$ with $u \sim N(0, \Omega)$ and $\Omega = Blkdiag\left(\{\Sigma_t\}_{t=1}^{T}\right)$.

We first consider Bayesian estimation of unrestricted linear regression model. Set prior for β:

$$\beta \sim N\left(\underline{\beta}, \underline{V}_\beta\right) \tag{2.36}$$

then the posterior is

$$\beta \sim N\left(\bar{\beta}, \bar{V}_\beta\right)$$

with

$$\bar{\beta} = \bar{V}_\beta\left(X^{'}\Omega^{-1}y + \underline{V}_\beta^{-1}\underline{\beta}\right) \tag{2.37}$$

$$\bar{V}_\beta = \left(X^{'}\Omega^{-1}X + \underline{V}_\beta^{-1}\right)^{-1} \tag{2.38}$$

Now consider linear restriction on coefficients

$$R\beta = 0 \tag{2.39}$$

where R is a $q \times k$ matrix with $rank(R) = q$ and $q \leq k$. The posterior of β can be derived by pooling the linear regression model, the prior information and the linear restriction condition above together:

$$y = X\beta + u \tag{2.40}$$

$$\underline{\beta} = \beta + v \tag{2.41}$$

$$0 = R\beta + \eta \tag{2.42}$$

where $v \sim N(0, \underline{V}_\beta)$ and $\eta \sim N\left(0, \frac{1}{\lambda} I_q\right)$. Equation (2.41) is prior inforamtion in equation (2.36) and equation (2.42) is linear restriction of (2.39) when $\lambda \to \infty$. The posterior of restricted β for a given λ can be obtained by conducting generalized least square estimation of the pooled regression model (2.40) - (2.42):

$$\beta_{res}(\lambda) \sim N\left(\tilde{\beta}(\lambda), \tilde{V}(\lambda)\right)$$

with

$$\tilde{\beta}(\lambda) = \left(X'\Omega^{-1}X + \underline{V}_\beta^{-1} + \lambda R'R\right)^{-1}\left(X'\Omega^{-1}y + \underline{V}_\beta^{-1}\underline{\beta}\right) \tag{2.43}$$

$$\tilde{V}(\lambda) = \left(X'\Omega^{-1}X + \underline{V}_\beta^{-1} + \lambda R'R\right)^{-1} \tag{2.44}$$

Following De Wind and Gambetti (2014), (2.44) is further expanded, by matrix inversion lemma, to[13]

$$\tilde{V}(\lambda) = \left(X'\Omega^{-1}X + \underline{V}_\beta^{-1}\right)^{-1} - \left(X'\Omega^{-1}X + \underline{V}_\beta^{-1}\right)^{-1} R'\left(\lambda^{-1} I_q + R\left(X'\Omega^{-1}X + \underline{V}_\beta^{-1}\right)^{-1} R'\right)^{-1} R\left(X'\Omega^{-1}X + \underline{V}_\beta^{-1}\right)^{-1}$$

[13]Matrix inversion lemma: $(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}$

Since equation (2.37) and equation (2.38), posterior of $\beta_{res}(\lambda)$ for a given λ can be expressed in terms of unrestricted posterior of β:

$$\tilde{\beta}(\lambda) = \left(I_k - \bar{V}_\beta R' \left(\lambda^{-1} I_q + R \bar{V}_\beta R' \right)^{-1} R \right) \bar{\beta}$$

$$\tilde{V}(\lambda) = \left(I_k - \bar{V}_\beta R' \left(\lambda^{-1} I_q + R \bar{V}_\beta R' \right) R \right) \bar{V}_\beta$$

With above prepared expression, letting $\lambda \to \infty$, the posterior of β_{res} follows

$$\beta_{res} \sim N\left(\tilde{\beta}_{res}, \tilde{V}_{res} \right)$$

with

$$\tilde{\beta}_{res} = \left(I_k - \bar{V}_\beta R' \left(R \bar{V}_\beta R' \right)^{-1} R \right) \bar{\beta}$$

$$\tilde{V}_{res} = \left(I_k - \bar{V}_\beta R' \left(R \bar{V}_\beta R' \right)^{-1} R \right) \bar{V}_\beta$$

and $\bar{\beta}$ is from (2.37), $\bar{V}_\beta$ from (2.38).

Chapter 3

Structual Analysis in a Large Bayesian VAR

3.1 Introduction

In previous two chapters, We have studied time-varying parameters vector autoregressive regression models with stochastic volatility or common volatility. As we know that vector autoregessive models are unrestricted models and relatively easily tractable with good properties such as impulse response function and variance decompositon for the future dynamics or historical decomposition for the past, therefore it is widely used by practioners, reseachers and policy-makers in economic, policy ananlysis and forecasting. In addition, since endogenous variables in a vector inter act each other that can capture very complicated relationships in an economy which economists are interested in, and most importantly some of the relationship among those variabes are still vague, disputable or even unknown, such as typiclly the effects of monetary policy on the performance of economy, the relationship between financial market and real economy, and recently the uncertainty shocks that is paid more and more attention after global financial crisis of 2008 and in the subsequent still slow recovery, VARs are used again and again to find inner possible relationships that have not been explored by theoretical models and thus VARs are also good instructions for the direction of theoretical analysis.

Nevertheless, everything has its two sides. VAR is not a parsimonious model due to its unrestricted structures that are very suitable to expose blackboxes in economy but it is hard to estimate in parctice when the number of observations in a vector increase. Its advantege cause its disadvantage. Typically, in empirical analysis, a VAR has three or five, at most ten variables, but seen otherwise without any restrictions, no matter in static VAR models or TVP-VAR models as we have discussed in the previous chapers. Chapter 1 focuses on the significance of the time varying parameters over each identity and each time via stochastic variable selection and provide a method to estimate them efficiently in practice; Chapter 2 lies in reducing dimension of the TVP-VAR with SV via factor idea. Namely, we use factors to represent and reduce number of lags, coefficients on regressors and volatilities in order to make a parsionious estimation with potentional over-fitting problem. That is different with the ever increasing literature that mainly on reducing the dimension of obsevations when one meet rich data enviroment such as facor models begining with Geweke

(1977) and factor augmented VAR recommended by Bernake, Boivin and Eliasz (2005) and Stock and Watson (2005).

In this chapter, we still look at vector autoregressive models in constant parameters, rather than time variant, discussed in Chapter 1 and Chapter 2. The VAR model with constant parameters probably is the most widely used one in paractise. Economist and practioners have again and agian proved that VARs with constant parameters still do a good job in economic analysis, policy making and forecasting compared to other type of VARs like TVP-VAR or regime switching VAR. Banbura, Giannone and Reichlin (2010) argue that vector regression with Bayesian shrinkage is an appropriate tool for large dynamic models. Building on the results of De Mol and co-authors (2008), they show that when the degree of shrinkage is set in relation to the cross-sectional dimension, the forecasting performance of small monetary VARs can be improved by adding additional maroeconomic variables and sectoral information. In addition, VARs with shrinkage produce credible impulse responses and are suitable for structural analysis.

Small VARs have the limitation that the information incorporated in the model is very small, while in real economic world, we have hundreds of maroeconomic data published by goverrtment agency and research institution. For example, central banks usually obseve and investigate a large amount of data in decision making of their policy; Economic agency in an economy can not determine their final behavior within small number of observations. A small information-involved VAR sometimes cause misleading and misunderstanding, a typical case is of 'prize puzzle' due to the missing of forward looking variables in VAR analysis. However, if a VAR with more variables will inevitably lead parameter proliferation. As suggested by Banbura et al. (2010), the way to deal with the problem is to impose proper priors of shrinkage on the parameters. On the priors with property of shrinkage, the parameters that are important for the variables will be strenghen while less important will shrink to zero or a limit value.

Follwing Banbura et al. (2010) with proper priors, we conduct an empirical analysis on three important shocks: monetary policy shock, uncertainty shock and financial shock. Monetary policy shock and its transmission mechanism, as we known, are very important to central banks and

economic agencies. It reflects central banks' stance to the current economic performance and what kind of policy they conduct, and meanwhile how agencies react to such stance and policy implementaion. The second - uncertainty shocks and the third - financial shocks recently are more and more jointly considered, in paticular, in the current situation of slow recovery. Researchers find that uncertainty and financial market are strongly tangled each other. They are both channels and shocks. They are both potentially important drivers in business cycle and display strong nonlinearity in different economic stages. We use a large data set which contains 28 variables that cover a broad range of an economy such as goods market, labor market, financial market and so on. By the large VAR, we jointly identify and investigate the three shocks and their transmisson. We find some interesting results via Bayesian estimation with shrinkage priors.

This chapter is organized as follows. In the second section, we discuss some issues on the large Bayesian VAR and the inter relationship among financial market, uncertainty and monetary policy. In section 3, we give the theoretical background of the prior on shrinkage, model specification and posterior in terms of large cross-sectional data set. Section 4 conducts empirical analysis and gives some important findings we obtain and implications we infer from the large Bayesian VAR. The last section concludes.

3.2 The literature

In this section, some issues will be discussed in current literature on the model and on the shocks and their transmission, respectively.

We first look at the model.

In econometric literature on VAR models, a very important issue is how to deal with the potential over parameterization problem inherited in VARs and the accessibility of high dimensional data set. This two typically join togther. Researchers would like to include more observations in the dynamic model in order to expose and explore complicated inner relationships that samll scale model is unable to do, at the same time if more variables are incorporated, the number of

parameters will nonlinearly, dramatically proliferate, making the estimation and inference infeasible even large data is availabe. Generally, there are three branches in dealing with this dilemma. One is on reducing data dimension by factors. Since the Geweke (1977), factor modes have been the most common way of achieving this goal. Applications such as Forni and Reichlin (1998), Stock and Watson (1999, 2002b), Bernake and Boivin (2003), have popularized factor methds among macroeconomists. Bernake, Boivin and Eliasz (2005) and Stock and Watson (2005) have combined factor methods with VAR methods. Del Negro and Otrok (2008) and Korobilis (2013a) provide further time varying parameter extension to these models. The second is on lowering the dimension of parameters of the model still using the factor idea. Indeed, factors that drive large observations can substantially limit the over-fitting problem, which is specially prominent in constant-parameter VARs, however, when it comes to time varying ones, the number of parameters in the case of constant-paramer model will be multiplied by the periods of sample size. The number of parameters again becomes very large even in a small scale TVP-VAR as discussed in Chapter 2. To deal with the problem, Kim and Yamamoto (2012) use coefficients on recent lags as factors to drive distant lags; De Wind and Gambetti (2014) decompose the covariance matrix of innovations to time varying parameters and extract several factors to drive whole dynamic process of all the time varying coefficients. Note that the covariance matrix associated with the parameters become reduced rank; Carriero, Clark and Marcellino (2012) present a model applying latent factor, i.e. common volatility to represent fully stochatic volatility on each variable justfied by the observation that most maroeconomic variables share very similar pattern of estimated volatility. They are all trying to limit parameter dimension in the context of time varying coefficient framwork. The last is on shrinking the parameters via imposing priors. Probably the most classic one is Minnesota prior (see Doan, Litterman and Sims, 1984 and Litterman, 1986). The basic idea of the prior is that it makes model implement like a random walk process. There are also other priors on stochastic search variable selection (SSVS) (see George, Sun and Ni, 2008). This paper focuses on the third branch, namely, setting suitable priors to shrink the parameters of the large VAR via Bayesian method. For the first and the third branch, Koop and Korobilis (2010) have given a good

survey on them.

The second issue is about the empirical analysis. Here, we discuss financial market and uncertainty jointly. The literature has identified at least three channels through which uncertainty shocks impose impact on economic activity. First, unceratin can affect the behavior of firms (Bernake, 1983; Bloom, 2009). A key concept in this framwork is irreversibility in investment. If investment decisions are irreversible, firms must take investment decisions that trade of the extra returns from early commitment agianst the benefit of having more informatioin by waiting. Bernake's real options framwork captures the notion that when uncertainty is high, the option value of waiting increases as it may be beneficial for firms to wait and acquire more information before deciding to invest in a real asset. The second, higher uncertainty may induce households to save more as higher uncertainty about future income will delay consumer spending, in particular on durable goods (Romer, 1990). The last channel is via financial maket for which a more recent strand of research places financial rather than real frictions at the center of the transmission mechanism (Arellano et al., 2012; Christiano et al., 2014; Gilchrist et al., 2014). If financial contracts are subject to agency problem or moral hazard probem, a rise in economic uncertainty increase the premium on external finance, leading to an increase in the cost of capital faced by firms or borrowers and thus a fall in investment. These three channels mixed together, especially the last channel makes that uncertainty and financial market should be considered together.

In empirical works, Beetsma and Giuliodori (2012) use linear VARs via rolling windows and show that the impact of uncertainty shocks on output in the US has decreased over the last five decades. Hartmann et at. (2012) use a regime switch (with fixed probability) model to estimate the links between financial stress and macroeconomic variables. They find that financial shocks have more serious impact on real variables in high financial stress regime than in normal times. Bijsterbosch and Guerin (2014) also use a regime swithing model to study regime-dependent relationship between uncertainty and economic activity. They find that only the third regime - the highest uncertainty regime are associated with a weaker growth performacne and sharp decline in stock price.

Caggiano et al. (2014) use instead a smooth - transition VAR (with continous probability) where parameters are allowed to depend on the state of the business cycle. They find that, using U.S. quarterly post-war data, uncertainty shocks have a stronger impact on unemployment during recessions. The above researches only independently study the uncertainty or financial conditions, not jointly together. It is crucial if the 'financial view' of the third channel discussed above plays a very important role. To our best knowledge, the Alessandri and Mumtaz (2014) probably the first to study the joint relationship of uncertainty and financial condition to check the importance of the 'financial view' in literature. Their model is very innovative in twofolds. First, the effects of uncertainty is dependent on the different financial conditions which is determined by a threhold variable that represents financial conditions on different stage rather than business cycle stages defined by NBER. Second, the uncertainty is generated endogenously from the common stachastic volatility as we discussed in Chaper 2 and the common volatility can be regarded as uncertainty proxy that is placed in the regressor, then the volatility generated from the model itself have the effects on the endogenous variables. This model is so called volatility in mean. The idea of using a single volatility process in a multivariate model has been introduced by Carriero et al. (2012) while volatility in mean effects are studied in the context of otherwise linear VAR models by Mumtaz and Thedoridis (2012), Mumtaz and Surio (2013) and Mumtaz and Zanetti (2013).

The above linear or nonliner VAR models as tool to study uncertainty and financial market jointly or indepently all belong to small scale VAR models. Caggiano et al. (2012) have four variables, namely the uncertainty proxy of VIX index and inflation, umeployment and federal funds rates. Alessandri and Mumtazi (2014) also use four variables including a financial condition proxy. Bloom (2009), perhaps the most, contained eight observations. Except for the Banbura et al. (2010), large Bayesian VAR models are seldom used in structual analysis though recommended by them in terms of large information of great view. We only find three works related with it. Gupta et al. (2012) study the effects of monetary policy on housing sector dynamics in a large sacle Bayesian VAR with 143 monthly macrovariables. Auer (2014) consider another direction of monetary policy shock on foreign investment income from a large VAR with 32 variables. Sanjani

(2014) use quarterly data of 34 variables to study financial frictions. In this chapter, we use a large data set of 28 observables to jointly identify and analyse the effect of each structual shock on different dimension of the U.S. economy. In the next section, the large Bayesian VAR model is presented and the priors on shrinkage are discussed.

3.3 The model

3.3.1 The likelihood of VAR

Let's consider a VAR model with p lags. It can be typically written as

$$y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \tag{3.1}$$

where y_t is an $n \times 1$ vector containing n endogenous observations; c is also $n \times 1$ constant term; coefficients on lags are from B_1 to B_{t-p} with corresponding dimension $n \times n$ and u_t is residual follows $u_t \sim iddN(0, \Sigma)$. The above equation is the expression for each period, it can be rewritten in the form of nesting all the data, that is in a compact form

$$Y = XB + U \tag{3.2}$$

where equation (3.2) is obtained with matrices $x_t = \left[y_{t-1}^{'}, \ldots, y_{t-p}^{'}, 1\right]^{'}$, then the $X = [x_1, \ldots, x_T]^{'}$ and accordingly $B = [B_1, \ldots, B_{t-p}, c]^{'}$ as well as $U = [u_1, \ldots, u_T]^{'}$. By column stacking operator of both sides of equation (3.2), we further have the form

$$y = (I_n \otimes X)\beta + u \tag{3.3}$$

where $y = vec(Y)$, $\beta = vec(B)$, $u = vec(U)$ and $u \sim N(0, \Sigma \otimes I_T)$ with vec column stacking operator and $\otimes$ Kronecker product. The regressors in the brackets are obtained by $vec(X \cdot B \cdot I)$.[1]

[1] $vec(ABC) = \left(C^{'} \otimes A\right) vec(B)$

Note that $y = [y_1; \ldots; y_n]$ where y_i is a $T \times 1$ vector collecting all observations belonging to i, therefore y is a $Tn \times 1$ column. u has the similar structure as y. The likelihood of (3.3) based on the multivariate normal distribution of u is written as

$$f(y|\beta,\Sigma) \propto |\Sigma|^{\frac{T}{2}} \exp\left(\frac{1}{2}u'(\Sigma \otimes I_T)^{-1}u\right)$$

by the transformation formula $tr(ABC) = vec\left(A'\right)'(I \otimes B)\,vec(C)$, the above is equal to

$$f(y|\beta,\Sigma) \propto |\Sigma|^{\frac{T}{2}} \exp\left(\frac{1}{2}\mathrm{tr}\left((Y-XB)'(Y-XB)\Sigma^{-1}\right)\right) \tag{3.4}$$

The above likelihood after leaving out the constant term that does not affect finding distribution kernels, can be decomposed into two components: $f(\beta|\Sigma,y)$ and $f(\Sigma|y)$. The former follows multi-variate normal distribution and the latter follows inverse wishart distribution. By the equation that $(Y-XB)'(Y-XB) = \left(Y-X\hat{B}\right)'\left(Y-X\hat{B}\right) + \left(\hat{B}-B\right)'X'X\left(\hat{B}-B\right)$, where $\hat{B} = \left(X'X\right)^{-1}X'Y$ is the least square estimate of B, equation (3.4) gives

$$f(y|\beta,\Sigma) \propto |\Sigma|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\hat{B}-B\right)X'X\left(\hat{B}-B\right)\Sigma^{-1}\right)\right) \tag{3.5}$$

$$|\Sigma^{-\frac{T-K}{2}}| \exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(Y-X\hat{B}\right)'\left(Y-X\hat{B}\right)\Sigma^{-1}\right)\right) \tag{3.6}$$

the first equation (3.5) is the kernel of the matrix normal distribution and the second (3.6) is the kernel of the inverse wishart distribution. It is more convenient to rewrite the matrix normal distribution in terms of the multivariate normal distribution, so that

$$\beta|\Sigma,y \sim N\left(\hat{\beta},\Sigma \otimes \left(X'X\right)^{-1}\right) \tag{3.7}$$

$$\Sigma|y \sim IW\left(\hat{S},T-k-n-1\right) \tag{3.8}$$

where $\hat{\beta} = vec\left(\hat{B}\right)$, $\hat{S} = \left(Y - X\hat{B}\right)' \left(Y - X\hat{B}\right)$ and $k = np + 1$

From the above derivation, we know that the likelihood function of a VAR is a product of two conditional distributions.

3.3.2 The priors

First we introduce the Minnesota prior. The basic idea of Minnesota prior is that all the equations are 'centered' around the random walk with drift, which means that $Y_t = c + Y_{t-1} + u_t$. This means diagonal of B_1 shrink to one and the remaining elements of coefficient matrix from B_1 to B_p towards to zero. Any element of coefficient matrices independently follows a normal distribution satisfying the moment conditions

$$E\left[(B_k)_{ij}\right] = \begin{cases} \delta_i, & j = i, k = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

$$V\left[(B_k)_{ij}\right] = \begin{cases} \frac{\lambda^2}{k^2}, & j = i \\ \nu \frac{\lambda^2}{k^2} \frac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise} \end{cases} \tag{3.10}$$

which reflects that diagonal elements of coefficient matrix on the first lag centered on δ_i. If δ_i shrinks to 1, it is random walk; If δ_i converges to0, it is white noise. The diagonal covariance matrix V on each B_k implies two beliefs that more recent lags are more important than more distant lags which can be seen that as $k \to p$, the magnitude of the variance, i.e. the range of flucuation become narrower and narrower and accordingly coefficients on far lags shrink more quickly to zero and that own lags of a variable have more powerful interpretation than foreign lags which can be found on the value of $\nu \in (0,1)$. In the original version of Minnesota prior, it has a hyper parameter to control tightness on each line of (3.10). In Banbura et al. (2010), the hyper parameter on the first line has been normalized, therefore ν takes the range. We follow the way of Banbura. Finally, let's look at the λ which controls the overall tightness of the prior distribution.

The λ shows up on both lines of (3.10), which means that as $\lambda \to 0$, the posterior is equivalent to the prior and the data is perfectly dominated by prior information; when $\lambda \to \infty$, the very flat prior, then the posterior equals likelihood and prior plays no role. De Mol et al. (2008) show that the parameters should be shrunk more so as to alleviate over-fitting when the number of variables increases. σ_i^2/σ_j^2 shows the scale and variability of the data. These values can be set by residual variance on AR(p) regression on each variable in a large data set.

The original Minnesota prior assumes that covariance matrix Σ of u_t is fixed and diagonal. If one is interested in structural analysis, Σ should be imposed as a full matrix to allow for possible correlation among residual of different variables. To overcome the problem, following Kadiyala and Kalsson (1977), Banbura et al. (2010) impose a normal inverse wishart prior on (β, Σ) under the condition that $\nu = 1$. This prior can retain the principles of the Minnesota prior. The normal inverse wishart prior has the form

$$vec(B)|\Sigma \sim N(vec(B_0), \Sigma \otimes \Omega_0) \tag{3.11}$$

$$\Sigma \sim IW(S_0, \alpha_0) \tag{3.12}$$

Since (3.7) and (3.8) are also the same distributions, the product of (3.11) and (3.12) is the likelihood of prior information. By this link, imposing prior on likelihood is equivalent to adding dummy observations to the original data set which demonstrates the advantage of natural conjugate prior of normal inverse wishart prior. Following (3.7) and (3.8), the prior parameters B_0, Ω_0, S_0 and α_0 can be expressed by dummy observations: $B_0 = \left(X_d' X_d\right)^{-1} X_d' Y_d$, $\Omega_0 = \left(X_d' X_d\right)^{-1}$, $S_0 = (Y_d - X_d B_0)' (Y_d - X_d B_0)$ and $\alpha_0 = T_d - k - n - 1$ where $k = np + 1$.

It can be shown that the dummy variables X_d and Y_d involving Minnesota moments of (3.9) and

(3.10) are

$$Y_d = \begin{pmatrix} \text{diag}\left(\delta_1\sigma_1,\ldots,\delta_n\sigma\right)/\lambda \\ 0_{n(p-1)\times n} \\ \cdots \\ \text{diag}\left(\sigma,\ldots,\sigma_n\right) \\ \cdots \\ 0_{1\times n} \end{pmatrix}$$

$$X_d = \begin{pmatrix} J_p \otimes \text{diag}\left(\sigma_1,\ldots,\sigma_n\right)/\lambda & 0_{np\times 1} \\ \cdots & \\ 0_{n\times np} & 0_{n\times 1} \\ \cdots & \\ 0_{1\times np} & \varepsilon \end{pmatrix}$$

where $J_p = \text{diag}\left(1,2,\ldots,p\right)$.

We also include an additonal prior, which implements a so-called 'inexact differencing' of the data. More preciselty, rewrite the VAR equation (3.1) in an error correction form:

$$\triangle y_t = c + \Pi y_{t-1} + B_1^\star \triangle y_{t-1} + \cdots + B_{p-1}^\star \triangle y_{t-p+1} + u_t$$

where $B_s^\star = -B_{s+1} - \cdots - B_p$, $s = 1,\ldots,p-1$ and $\Pi = B_1 + \cdots + B_p - I_n$.

A VAR in first difference implies the restriction $\Pi = 0$ or $A_1 + \cdots + A_p = I_n$. We follow Doan et al. (1984) and set a prior that shrinks Π to zeros. Specially, we set a prior that is centered at 1 for sum of coefficients on the own lags for each variable, and at 0 for the sum of coefficients on other variables' lags. This prior introduces correlations among the coefficients on each variable in each equation. The tightness of this prior on the sum of 'coefficients' is controlled by the hyparameter τ. As τ goes infinity, the prior becomes diffuse, whereas as it goes to zero, we approach the case of exact differencing, which implies the presence of a unit root in each equation. In the literature, it is usually implemented by adding the following dummy observations:

$$Y_d^\star = \text{diag}\left(\delta_1\mu_1,\ldots,\delta_n\mu_n\right)/\tau \tag{3.13}$$

$$X_d^\star = \begin{pmatrix} \left(1_{1\times p}\right)\otimes \text{diag}\left(\delta_1\mu_1,\ldots,\delta_n\mu_n\right)/\tau & 0_{n\times 1} \end{pmatrix} \tag{3.14}$$

The incorporation of this sum of coefficients prior serves to increase accuracy of forecast and avoid confidence band explosiveness in long horizons for impulse response functions.

3.3.3 The posterior

As discussed above, implementation of priors is equivalent to adding dummy observations transformed from priors, now the equation (3.2) is augmented to

$$Y^\star = X^\star B + U \tag{3.15}$$

where $T^\star = T + T_d + T_d^\star$, $Y^\star = \left(Y^{'}, Y_d^{'}, Y_d^{\star'}\right)^{'}$ and $X^\star = \left(X^{'}, X_d^{'}, X_d^{\star'}\right)$ with $T_d = n(p+1)+1$, $T_d^\star = n$. To ensure the existence of the prior expectation of Σ, it is necessary to add an improper prior $\Sigma \sim |\Sigma|^{-(n+3)/2}$. After that the posterior has the form

$$vec\left(B\right)|\Sigma, Y \sim N\left(vec\left(\tilde{B}, \Sigma\otimes\left(X^{\star'}X^\star\right)^{-1}\right)\right)$$

$$\Sigma|Y \sim IW\left(\tilde{\Sigma}, T_d + T_d^\star + 2 + T - k\right)$$

where $\tilde{B} = \left(X^{\star'}X^\star\right)^{-1}X^{\star'}Y^\star$ and $\tilde{\Sigma} = \left(Y^\star - X^\star\tilde{B}\right)^{'}\left(Y^\star - X^\star\tilde{B}\right)$ which are typical setting for normal inverse wishart distribution as seen in (3.7) and (3.8).

3.4 Empirical analysis

In this section, we apply the large Bayesian VAR model with priors of shrinkage to a large data set which contains 28 variables covering a broad range of the U.S. economy. We focus three shocks which are monetary policy shocks, uncertainty shocks and financial shocks.

3.4.1 The data

Since large Bayesian VAR with proper shrinkage priors is able to deal with large data set, the whole information in the data can be directly used to explore complicated dynamic process among different variables avoiding possible missing variable bias. Following Sanjani (2014), we select 28 variables that generally can give a comprehensive description of the economy of U.S.

This data covers labor market, housing market, labor market, government bonds market, corporate bonds market and so on. To identify uncertainty shock and financial shock, we include VIX index proxy for uncertainty measure and Chicago Fed national financial condition index for financial condition measure.

As for uncertaity, three types of measure can be found: financial market indicators, survey based measures including forecast dispersion measures and media measures based on the number of citations of a specific term. Other measures are more microeconomic in nature and based on various indicators of dispersion at individual company or industry level. In this paper, we choose the VIX index due to two reasons. First it is widely used as standard uncertainty measure (Bloom, 2009) and second it covers long period since 1962 the third quarter.

For the financial conditons, the Chicago Fed national financial condition index is chosed. It is a real-time indicator of financial distress constructed and maintained by the Chicago Fed and described extensively in Brave and Butters (2012). The index extracted using dynamic factor analysis from a set of 120 series that describe a broad range of monetary, debt and equity markets as well as the leverage of the financial industry. Another advantage of the data is that it covers the longest periods since 1973 the first quarter.

We use quarterly data from 1973 the first quarter to 2009 the last quarter that covers the longest period for all the 28 variables.

3.4.2 Empirical findings

We use 28 quarterly data form 1973 Q1 to 2009 Q4 with 2 lags in large VAR in Bayesian estimation. We also tried more lags, however, the results are robust.

The identification of the monetary policy shocks, uncertainty shocks and financial shocks are implemented in a recursive manner. For the identification of monetary policy shocks, the observations are divided into three blocks. The first one is slow moving block that contain variables not sensitive in response to monetary policy shocks and react with one period lag. The second block only contains the instrument of conventional monetary policy, namely the effctive fedral fund rate. The last block contains variables that are very sensitive to the changes in monetary policy – the surprise. All the variables in the third block are financial variables that cover government bonds, corporate bonds, stock , exchange rate markets.

As for the identification of the uncertainty shock, I place VIX index for uncertainty measure in the first place in the first block, that is VIX is the first variable among all the observations. Caggiano et al. (2012) in a small scale nonlinear VAR for the U.S economy and Kamber et al. (2013) in a factor augmented VAR model for small open economy of New Zealand, both place uncertainy measure of VIX in the first. In order to make the uncertainty measure as compatible as possible with the identification restriction - uncertainty is not affected immediately by other shocks, following Kamber et al. (2013), the quarterly data is constructed by only choosing the first month of the quarter. Regarding financial condition variable, it is ordered the last in the third block. That is, the national financial condition index is the last one among all the observations in the sense that this indicator responses to all the variables contemporarily. In practice, identification of these shocks are conducted by cholesky decomposition of the covariance matirx of the residuals in the large VAR.

Let's first look at the effects of monetary policy shocks on the economy. Figue 3.1 plots the response functions in median for all the 28 variables with dashed lines representing confidence interval between 16^{th} percentile and 84^{th} percentile. From the responses, it is reasonable to believe that the non-systematic monetary policy is well identified. One standard deviation monetary policy shock cause almost all the responses in line with economic theory. Non-systematic monetary shocks increase uncertainty that is recently empasized by Baker et al. (2013) in the current slow recovery stage after the global financial crisis of 2008. Both real GDP and industrial production of

different output measure decline in response to tightening policy. Capacity utilization gets down in response to high interest rate. Residential investment responses much stronger than non-resdential investment due to that they are more sensitive to financing cost. Consumer confidence is also depressed by the tightening policy. With the increase in federal fund rate - the benchmark rate, rates of all the government bonds and corporate bonds rise. Stock price declines as expected. U.S dollar appreciates due to high return in international capital market. M1 and M2 decline reflecting liquidity. Finally, tighten monetary policy elevates financial stress.

Then we look at the uncertainy shocks in Fig 3.2. Uncertainty shocks have countercyclical effect on the economy. Capacity utilization decline. Real output, consumption, investment all get down consistent with different economic theory on uncertainty. What is interesting is that comparing with monetary shock, nonresidential investment decline facing futural uncertaity while residential investment seems nonsensitive to the uncertainty relative to borrowing cost caused by tightening monetary policy. Uncertainty also depresses consumer confidence. In respose to uncertainty shock, federal fund rate lowers in order to stimulate the economic activity. Decline in federal funds rate decreases government bonds rates while increase the Baa return. This means that with high uncertainty, the effects of requiring high risk premium dominate the decline in short rates for corporate bonds with low credit rating. M1 and M2 is improved by Fed policy. Stock price decline and stay on a low level for the whole five years. Uncertainty cause national financial conditon worse than before which is evidenced by the last graph in Fig 3.2.

As discussed in section 1 and section 2, financial market and uncertaity links together and affect each other. On the main diagonal of Fig 3.3, financial shocks increase the uncertainty. Financial shocks affect real activity for a long period. Compared response of industrial production in Fig 3.3 with that in Fig 3.2, the deline in IP caused by financial shock last for five years, while the effect caused by uncertainy recover back after three or four quarters. We can find the same case in real GDP, employment, hours worked, capacity utilization and nonresidential investment as opposed to uncertainty shocks. Again residential investment becomes sensitive with financial tightness – the same response with positive monetary shocks. In addtion, financial shocks seem to have no

significant effects on financial variables except for financial condition index itself; In contract, uncertainty shocks have strong effects on financial variables over government bonds, corporate bonds, money, stock and foreign exchange rate markets.

In sum, through the large Bayesian VAR with data set covering broad range of the U.S. economy, we have three findings. First, increase in uncertainty cause tight financial conditions and tight financial conditons lead high uncertainty (see the main diagonal of Fig 3.2 and Fig 3.3); Second, both financial and uncertaity shocks cause counter-cyclical effects on real activity but financial shocks have the effects kept for a long time; The last is that financial variables are much more sensitive to uncertainty shocks compared to financial shocks.

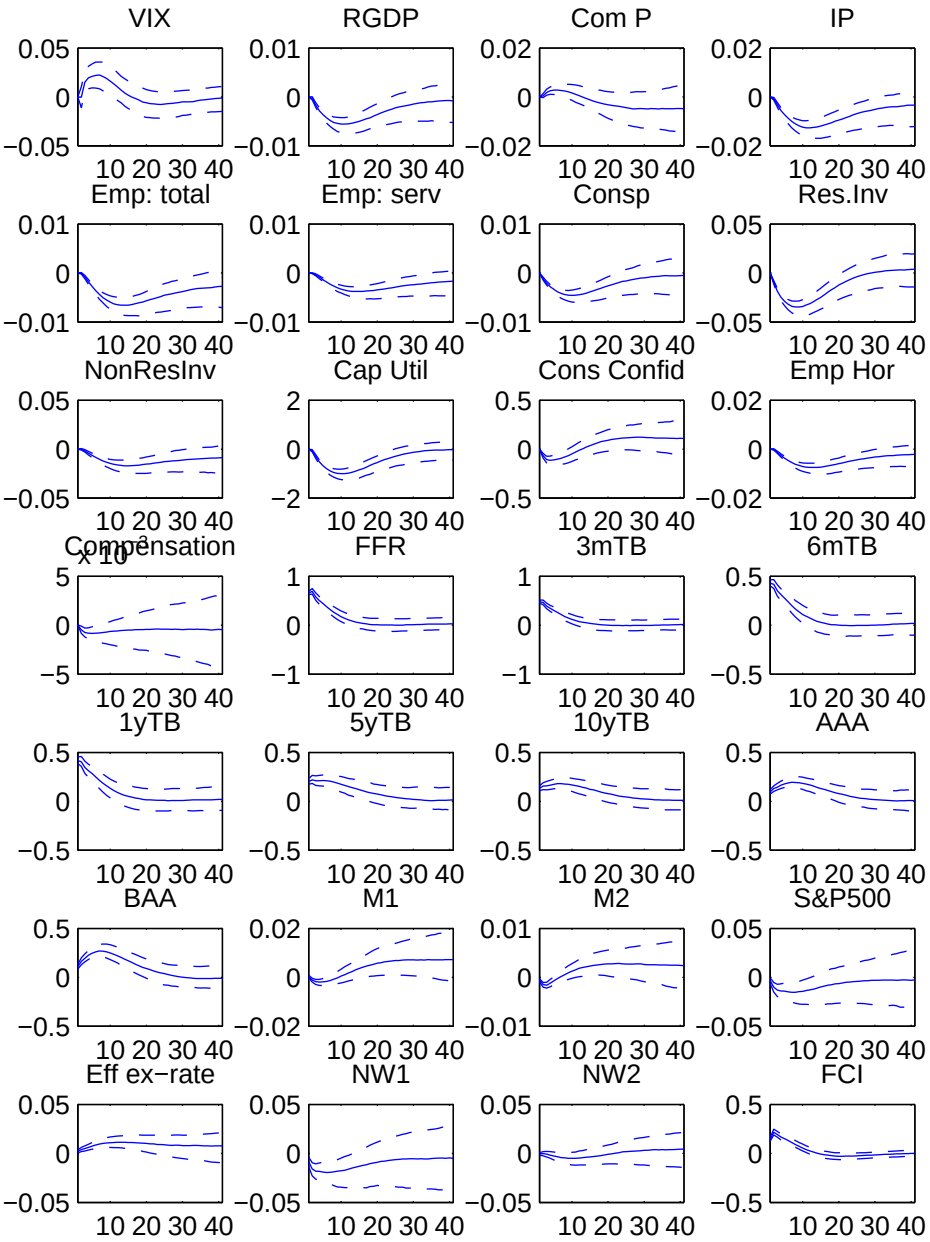
3.5 Concluding remarks

In this chapter, we switch from time varying parameter VAR to constant parameter VAR in a rich data environment. With the proper priors of shrinkage property imposed on the coefficients and estimated by Bayesian method, the model conducts a good performance in structual analysis.

We jointly identified three shocks which are monetary policy shocks, financial shocks and uncertainty shocks in a structural analysis. For the effects of monetary shocks, the impulse response functions are in line with theoretical predictions. For the financial shocks and uncertainty shocks, we analyse them together. Financial condition and uncertainty affect each other. Tight financial condition elevates uncertainty and in turn, high uncertainty exacerbates financial condition. Both positive financial and uncertainty shocks have negative effects on real activities, however, financial shocks have more persistent effects on these real variables than uncertainty shocks. We also find that financial variables care more for uncertainty shocks compared to financial shocks.

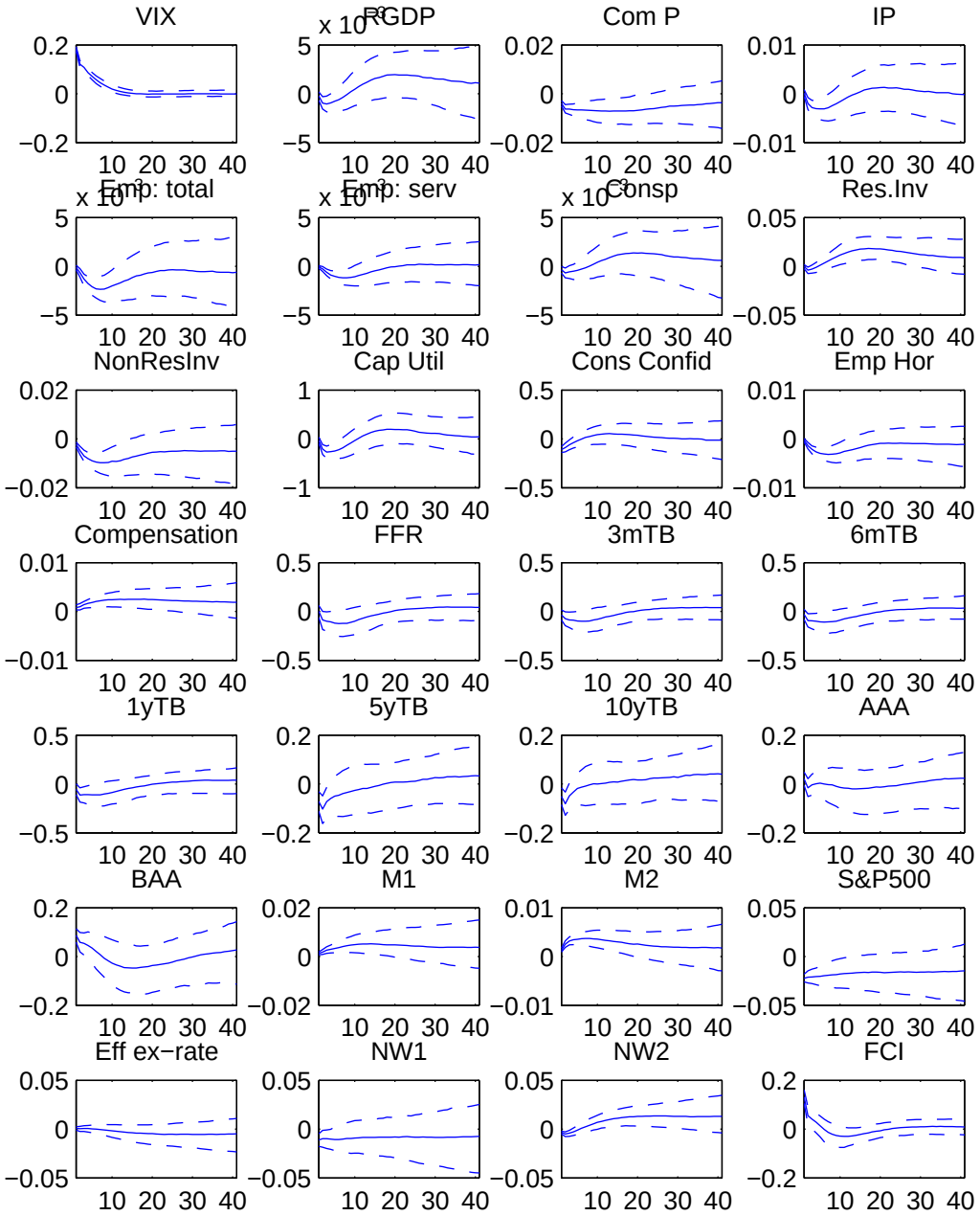
The empirical analysis implies that in a theoretical framwork such as DSGE model, the financial market and uncertainty should be combined together beacuse they are mutually as shocks and channels, affecting each other and multiple dimensions of the economy evidenced by the large Bayesian VAR.

Figure 3.1: IRs to monetary policy shocks



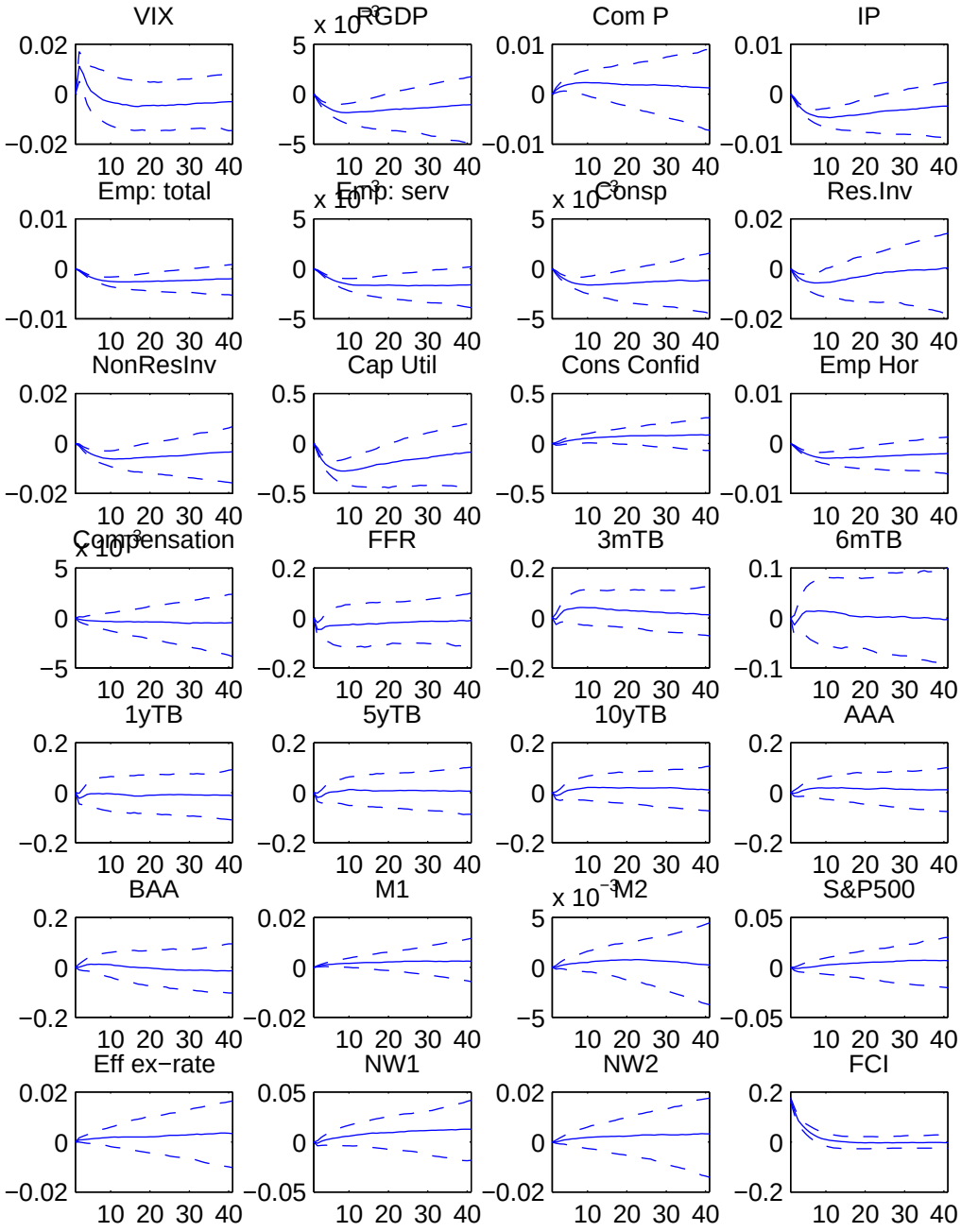
Notes: Impulse responses of 28 variables to monetary policy shocks with solid line representing posterior median and dashed lines covering confidence interval between 16^{th} percentile and 84^{th} percentile.

Figure 3.2: IRs to uncertainty shocks



Notes: Impulse responses of 28 variables to uncertainty shocks with solid line representing posterior median and dashed lines covering confidence interval between 16^{th} percentile and 84^{th} percentile.

Figure 3.3: IRs to financial shocks



Notes: Impulse responses of 28 variables to financial shocks with solid line representing posterior median and dashed lines covering confidence interval between16th percentile and 84th percentile.

Appendix

A. Data specification

Notes: The data set consists of 28 U.S. quarterly variables from 1973 Q1 to 2009 Q4. Except for consumer confidence index, effective exchange rate index and VIX volatility index, other data can be found from Federal Reserve Economic Data - FRED - St. Louis Fed. The consumer confidence index is from OECD. The effective exchage rate is from BIS. For the VIX index, the stock market volatility is measured by realized volatility before 1986 and by Black - Scholes implied volatility after 1986 (Choi, 2013).[2] The quarterly data of VIX is constructed by choosing the first month of each quarter following Kamber et al. (2013) to allow for lag effect in response to other shocks. For the other quaterly data, they are all constructed by monthly average for a specific quarter. The '#' column lists the order of the variables in the large VAR. In the 'Tcode' column, 1 means level, 2 means log level and 'sa' represents seasonally adjusted as well as 'nsa' not seasonally adjusted. In the last 'Id' column, 'f' implies fast moving and 's' slow moving.

[2]The monthly VIX index is kindly provided by Sangyup Choi.

Table 3.1: Data specification

#	Mnemonic	Description	Tcode	Id
1	VIX	CBOE Volatility Index	2 nsa	s
2	RGDP	Real gross domestic product	2 sa	s
3	PPI	Producer Price Index: finished goods	2 sa	s
4	IP: total	Industrial Production Index	2 sa	s
5	Emp: total	All employees: total private	2 sa	s
6	Emp: services	All employees: service providing	2 sa	s
7	consp	Real personal consumption expenditures	2 sa	s
8	Res.Inv	Residential private domestic investment	2 sa	s
9	NonRes.Inv	Nonresidential private domestic investment	2 sa	s
10	Cap Util	Capacity utilization: total industry	1 sa	s
11	Cons Confid	Consumer confidence index (OECD)	1 nsa	s
12	Emp. Hours	Hours of all persons: nonfarm	2 sa	s
13	Real Comp/Hour	Real compensation per hour: nonfarm	2 sa	s
14	FFR	Effective Federal Funds Rate	1 nsa	-
15	3mTB	Treasury Bill rate 3 months: second market	1 nsa	f
16	6mTB	Treasury Bill rate 6 months: second market	1 nsa	f
17	1yTB	Treasury Bill rate 1 year: constant maturity rate	1 nsa	f
18	5yTB	Treasury Bill rate 5 year: constant maturity rate	1 nsa	f
19	10yTB	Treasury Bill rate 10 year: constant maturity rate	1 nsa	f
20	AAA yield	AAA corporate bond yield	1 nsa	f
21	BAA yield	BAA corporate bond yield	1 nsa	f
22	M1	M1 money stock	2 sa	f
23	M2	M2 money stock	2 sa	f
24	S&P 500	S&P's common stock price index	2 nsa	f
25	Ex rate	Effective exchange rate index (BIS)	2 nsa	f
26	NW 1	Market value of equities outstanding	2 nsa	f
27	NW 2	Owners' equity in household real estate	2 nsa	f
28	FCI	Chicago Fed national financial condition index	1 nsa	f

References

Alessandri, P. & Mumtaz, H. (2014). Financial regimes and uncertainty shocks. *Working paper, No.729, Queen Mary University of London, School of Economics and Finance*.

Aoki, N. (1990). *State space modeling of time series*. Cambridge Univ Press.

Arellano, C., Bai, Y., & Kehoe, P. (2010). Financial markets and fluctuations in uncertainty. *Federal Reserve Bank of Minneapolis Working Paper*.

Auer, S. (2014). Monetary policy shocks and foreign investment income: evidence from a large bayesian var. *SNB working papers, Swiss National Bank*.

Baker, S. R., Bloom, N., & Davis, S. J. (2013). Measuring economic policy uncertainty. *Chicago Booth research paper*.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1), 71–92.

Baumeister, C. & Peersman, G. (2013). Time-varying effects of oil supply shocks on the u.s. economy. *American Economic Journal: Macroeconomics*, 5(4), 1–28.

Bauwens, L., Koop, G., Korobilis, D., & Rombouts, J. V. (2014). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics, early online publication*.

Beetsma, R. & Giuliodori, M. (2012). The changing macroeconomic response to stock market volatility shocks. *Journal of Macroeconomics*, 34(2), 281–293.

Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *The Quarterly Journal of Economics*, 98(1), 85–106.

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

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

Bernanke, B. S. & Mihov, I. (1998). Measuring monetary policy. *Quarterly Journal of Economics*, 113(3), 869–902.

Bijsterbosch, M. & Guérin, P. (2013). Characterizing very high uncertainty episodes. *Economics Letters*, 121(2), 239–243.

Blanchard, O. & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4), 1329–1368.

Blanchard, O. & Simon, J. (2001). The long and large decline in u.s. output volatility. *Brookings papers on economic activity*, 2001(1), 135–174.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623–685.

Boivin, J. (2001). The fed's conduct of monetary policy: Has it changed and does it matter? *Columbia Business School, mimeo*.

Boivin, J. & Giannoni, M. P. (2006). Has monetary policy become more effective? *The Review of Economics and Statistics*, 88(3), 445–462.

Brave, S. & Butters, A. (2012). Diagnosing the financial system: financial conditions and financial stress. *International Journal of Central Banking*, 8(2), 191–239.

Caggiano, G., Castelnuovo, E., & Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in u.s. recessions. *Journal of Monetary Economics*, 67, 78–92.

Canova, F. (2007). *Methods for applied macroeconomic research*, volume 13. Princeton University Press.

Canova, F. & Gambetti, L. (2009). Structural changes in the u.s. economy: Is there a role for monetary policy? *Journal of Economic dynamics and control*, 33(2), 477–490.

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418), 493–500.

Carriero, A., Clark, T. E., & Marcellino, M. G. (2012). Common drifting volatility in large bayesian vars. *CEPR Discussion Paper, No. 8894*.

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

Carter, C. K. & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3), 541–553.

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

Choi, S. (2013). Are the effects of bloom's uncertainty shocks robust? *Economics Letters*, 119(2), 216–220.

Christiano, L. J., Motto, R., & Rostagno, M. (2014). Risk shocks. *The American Economic Review*, 104(1), 27–65.

Ciccarelli, M. & Rebucci, A. (2003). Measuring contagion with a bayesian, time-varying coefficient model. *IMF working paper*.

Cogley, T., Primiceri, G. E., & Sargent, T. J. (2010). Inflation-gap persistence in the u.s. *American Economic Journal: Macroeconomics*, 2(1), 43–69.

Cogley, T. & Sargent, T. J. (2001). Evolving post-world war ii u.s. inflation dynamics. *NBER Macroeconomics Annual*, 16(1), 331–373.

Cogley, T. & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii u.s. *Review of Economic dynamics*, 8(2), 262–302.

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

De Wind, J. & Gambetti, L. (2014). Reduced-rank time-varing vector autoregressions. *CPB Discussion paper, N0. 270, CPB Netherlands Bureau for Economic Policy Analysis*.

Del Negro, M. & Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *FRB of New York Staff Report, No.326*.

Del Negro, M. & Primiceri, G. E. (2013). Time-varying structural vector autoregressions and monetary policy: a corrigendum. *FRB of New York Staff Report, No.619*.

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

Fernandez-Villaverde, J., Rubio-Ramirez, J. F., Sargent, T. J., & Watson, M. W. W. (2007). Abcs (and ds) of understanding vars. *American Economic Review*, 97(3), 1021–1026.

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

Forni, M. & Reichlin, L. (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies*, 65(3), 453–473.

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics*, 142(1), 553–580.

Geweke, J. (1977). *The dynamic factor analysis of economic time series*. Amsterdam: North-Holland.

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

Gilchrist, S., Sim, J. W., & Zakrajšek, E. (2014). Uncertainty, financial frictions, and investment dynamics. *NBER Working paper, No.20038*.

Gupta, R., Jurgilas, M., Kabundi, A., & Miller, S. M. (2012). Monetary policy and housing sector dynamics in a large-scale bayesian vector autoregressive model. *International Journal of Strategic Property Management*, 16(1), 1–20.

Hanson, M. S. (2006). Varying monetary policy regimes: A vector autoregressive investigation. *Journal of Economics and Business*, 58(5), 407–427.

Hartmann, P., Hubrich, K., Kremer, M., & Tetlow, R. J. (2014). Melting down: Systemic financial instability and the macroeconomy. *Available at SSRN 2462567*.

Jacquier, E., Polson, N. G., & Ross (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 12(4), 371–389.

Kadiyala, K. & Karlsson, S. (1997). Numerical methods for estimation and inference in bayesian var models. *Journal of Applied Econometrics*, 12(2), 99–132.

Kamber, G., Karagedikli, O., Ryan, M., & Vehbi, T. (2013). International spill-overs of uncertainty shocks: Evidence from a favar. *Reserve Bank of New Zealand, mimeo*.

Kim, C.-J. & Nelson, C. R. (1999). *State space models with regime switching: classical and gibbs-sampling approach with applications*. MIT Press.

Kim, D. & Yamamoto, Y. (2013). Time instability of the u.s. monetary system: Multiple break tests and reduced rank tvp-var. *Discussion Paper*.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3), 361–393.

Koop, G. & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and trends(R) in Econometrics*, 3(4), 267–358.

Koop, G. & Korobilis, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, 177(2), 185–198.

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4), 997–1017.

Koop, G. M. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, 28(2), 177–203.

Korobilis, D. (2013a). Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics*, 75(2), 157–179.

Korobilis, D. (2013b). Var forecasting using bayesian variable selection. *Journal of Applied Econometrics*, 28(2), 204–230.

Kuo, L. & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 60(1), 65–81.

Leeper, E. M. & Zha, T. (2003). Modest policy interventions. *Journal of Monetary Economics*, 50(8), 1673–1700.

Litterman, R. B. (1986). A statistical approach to economic forecasting. *Journal of Business & Economic Statistics*, 4(1), 1–4.

Lubik, T. A. & Schorfheide, F. (2004). Testing for indeterminacy: an application to u.s. monetary policy. *American Economic Review*, 94(1), 190–217.

Morris, S. D. (2012). Var (1) representation of dsge models. *University of California, San Diego working paper*.

Mumtaz, H. & Surico, P. (2013). Policy uncertainty and aggregate fluctuations. *CEPR Discussion Paper, No.8894*.

Mumtaz, H. & Thedoridis, K. (2012). The international transmission of volatility shocks. *Bank of England Working paper, No. 463*.

Mumtaz, H. & Thedoridis, K. (2014). The changing transmission of uncertainty shocks in the u.s.: an empirical analysis. *Working paper, No. 735, Queen Mary University of London, School of Economics and Finance*.

Mumtaz, H. & Zanetti, F. (2013). The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking*, 45(4), 535–558.

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2), 425–449.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3), 821–852.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, 54(7), 2048–2064.

Romer, C. D. (1990). The great crash and the onset of the great depression. *The Quarterly Journal of Economics*, 105(3), 597–624.

Sargan, J. D. & Bhargava, A. (1983). Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. *Econometrica: Journal of the Econometric Society*, 51(3), 799–820.

Shephard, N. G. & Harvey, A. C. (1990). On the probability of estimating a deterministic component in the local level model. *Journal of time series analysis*, 11(4), 339–347.

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

Sims, C. A. (1999). Drift and breaks in monetary policy. *Manuscript, Princeton University*.

Sims, C. A. (2001). Stability and instability in u.s. monetary policy behavior. *Princeton University, photocopy*.

Sims, C. A. & Zha, T. (2006). Were there regime switches in us monetary policy? *The American Economic Review*, 96(1), 54–81.

Stock, J. H. & Watson, M. W. (1998). Asymptotically median unbiased estimation of coefficient variance in a time varying parameter model. *Journal of the American Statistical Association*, 93(441), 349–358.

Stock, J. H. & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2), 293–335.

Stock, J. H. & Watson, M. W. (2002a). Has the business cycle changed and why? *NBER Macroeconomics Annual*, 17, 159–230.

Stock, J. H. & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162.

Stock, J. H. & Watson, M. W. (2005). Implications of dynamic factor models for var analysis. *NBER Working paper, No.11467*.

Taheri Sanjani, M. (2014). Financial frictions in data: Evidence and impact. *IMF working paper*.

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