ABSTRACT

WENG, QIFENG. Bayesian VAR Analysis in the Presence of Infrequent Shocks with Application to Analysis of Oil Price Shocks. (Under the direction of Atsushi Inoue.)

In this paper, we propose a so-called mean-plus-noise VAR (MPNVAR) model and applied it to the macroeconomics empirical study. Along with the model, we use Bayesian approach to estimate the parameters in the model.

Bayesian vector autoregressive (BVAR) models have been well evolved for decades independently from the development of mean-plus-noise (MPN) models. Study of MPN models is relatively a small branch in finance field for the use of long memory research. Econometricians have built beautiful theoretical frame work for both of them. Contributions of Bayesian VAR models in all kinks of empirical economic studies have been widely appreciated. For BVAR models, estimation and forecast results are better than than VAR model. In financial long memory and structure change research areas, researchers have modeled infrequent structural change to explain long memory in economic and financial research. Inspired by above promising features of both models, we introduce a new Bayesian VAR model whose disturbance terms take a form of MPN models called mean-plus-noise VAR (MPNVAR) model to empirical macroeconomic research in this paper. We use our MPNVAR model to analyze oil shocks for its infrequent shock feature estimated by Bayesian approach. We found parameters in our model can be consistently estimated. Moreover, impulse response functions analysis shows that the MPNVAR model helps interpret the effect of oil prices on the macro economy.

Bayesian VAR Analysis in the Presence of Infrequent Shocks
with Application to Analysis of Oil Price Shocks

by
Qifeng Weng

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Economics

Raleigh, North Carolina

2012

APPROVED BY:

Denis Pelletier	Huixia Wang

Atsushi Inoue
Chair of Advisory Committee

DEDICATION

To my grandparents, S. DING & Y. SHAO and my parents, L. DING & S. WENG.

BIOGRAPHY

This author was born in Wuxi, PRC. He finished his undergraduate in Shanghai University, majoring in Mathematics from 2006 to 2010 in Shanghai, PRC. In August, 2012, he started the economics master's program in NC State University. His current research interests include empirical macroeconomics and econometrics. Dr. Inoue, Atsushi is his advisor.

ACKNOWLEDGEMENTS

I would like to acknowledge my indebtedness to several people. First and foremost, I would like to express my gratefulness to my advisor, Dr. Atsushi Inoue. Actually, it is him who brought up this combined MPNVAR model. Without his help and guidance, this paper can not be finished. I also extend my gratitude to my advisory committee members Dr. Huixia Wang and Dr. Denis Pelletier for their insights and suggestions. I also thank all my professors in North Carolina State University.

I especially thank my grandparents Shuhua Ding and Yingjuan Shao for all what they did for me. At this time being, my grandfather Ding is struggling with his lung cancer toughly. This work is for him. I hope he will be fine soon. Then I can show it to him, making him proud of me. I love you so much, grandpa and grandma. Also I thank my parents, Liming Ding and Shijian Weng. Their support and guidance make this paper possible.

Special thanks my girlfriend Jia Geng for her understanding, love and company.

I also thank all other people make my life meaningful. Thank you, sincerely.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

This paper is motivated by the prior works of Bayesian vector autoregressive (BVAR) models and *mean-plus-noise* (MPN) model. We introduce a new BVAR model whose disturbance terms take a form of the MPN model. Namely, we use so-called MPNVAR, meaning M(ean)-P(lus)-N(oise) VAR. With this model's innovative features, we estimate a VAR model of oil shocks shocks based on Kilian's paper (2009). Vector autoregressive (VAR) models are useful modeling tools for economists to analyze both dynamic economic and financial problems. Along with its development, enormous estimation approaches have been widely explored over the past decades. Among them, Bayesian approaches have been proved to be a decent way to address estimation and specification problems of VARs. Milestones in Bayesian VAR literature include the following: Sims (1972 and 1980), in his seminal works, he first built VAR model in application to macroeconomic study. Since then, people start to use VAR model to model economic time series data and forecast. Litterman (1980) originally introduced Bayesian estimation approach to VAR model to address overfitting problem. His method avoided putting exact zero restrictions on certain coefficients by treating the uncertainty of parameters as a probability distribution. So this method can retain the uncertainty without compromising information from original data. Some extensions of Bayesian VARs include: Uhig (1997) extended it to stochastic volatility. Unlike ARCH models, his model captured sudden large movement by draws from a distribution with a randomly increased but unobserved variance; Regime-switching time series models (non-linear dynamic models) by Primiceri (2005) and Sims and Zha (2006) , etc. Primiceri concluded that there existed regime switching in both systematic and non-systematic monetary policy during the last forty years by his innovative modeling strategy for the law of motion of the variance covariance matrix. Sims and Zha gave out a multivariate regime-switching model for monetary policy. It is now widely appreciated that Bayesian VAR gives better estimation and forecasts than VARs estimated by the conventional approaches. Canova (1999) had specific discussions about this strength of BVAR models. Monetary policy researchers, central banker policy makers made bulk of papers going back and forth between simple univariate time series models to highly sophisticated nonlinear dynamic model for better forecast, are still in favor of Bayesian VAR.

Mean-plus-noise models have been introduced in the study of long memory and regime-switching in financial area by Chen and Tiao (1990) because of its infrequent sudden changed shocks can generate suspicious long memory behavior. This branch has been developed separately from the Bayesian VAR model. Granger and Hyung (2004) used this model to show the confusion between structural change and long memory. In an independent work, Diebold and Inoue (2000), found more general conclusions about this confusion by extending mean-plus-noise model in state space forms.

Inspired by above research and MPN model's infrequent shock character, we see its potentials in macroeconomics filed, especially for the oil shocks study started by Kilian (2009) or other fields like monetary policy research (Uhlig, 2005). Because both oil shocks and monetary policies don't change frequently over time. For instance, US interest rate, a benchmark of US monetary policy change, changes annually or even longer during stable economy, and quarterly or so during recession time. So monetary policy shocks occur infrequently. Major contributor to oil price changing as Kilian concluded in his paper is the precautionary demand shock. This shock presents when wars or turmoil occurs in mid-east or incident like 9/11 happens. So apparently, these shocks will not occur every day or monthly. Another feature of this infrequent shock is that once it changes, it is likely that the change will be bigger than conventional continuous disturbances and fades away in next few periods. Our MPNVAR model has the very power to handle both frequent and infrequent shocks simultaneously, which, we believe, will better describe economic phenomena like oil shocks and monetary policy. Our Monte Carlo experiment consistently verifies our belief.

The main contribution of this paper is that we proposed this new MPNVAR model and applied it to the macroeconomics empirical study. Along with the model, we use the Bayesian approach to estimate the parameters in the model. At the initial stage of this paper, we have considered the possibility to apply the MLE approach to estimate. Due to large numbers of latent variables $\{d_t\}$ (about 395 d_t's) in our empirical application, maximizing the likelihood function can lead to numerical calculation catastrophe. So we believe the Bayesian method is a better estimation strategy for our case. Here is the brief history of literature about Bayesian estimation approaches: Litterman (1986) introduced the famous Minnesota prior to discussion of prior choices, also known as Empirical Bayesian Estimation. The other prior choice solution known as Hierarchical Bayesian Estimation is brought up by Chib and Greenberg (1995). Numerical implementations of Bayesian method: e.g. the Gibbs sample introduced by Gelfand (1990), used as a substitution choice to get joint distribution; Metropolis-Hasting Algorithm, a MCMC method, first proposed by Metropolis et al. (1953), use as a tool for drawing random variables from unconventional distribution. For our estimation procedure, although the likelihood function of our MPNVAR model is different from conventional VARs', it has very similar properties as conventional ones. From what we discovered, most of parameters have conjugate distributions. Monte Carlo experiments have shown that they can be consistently estimated. Moreover, with the help of corresponding impulse response function, we show MPNVAR better explains the oil shocks.

This paper is organized as follows: We define MPNVAR model in chapter 2. Firstly, we describe the general form of the multivariate mean-plus-noise model as well as the conventional VAR(p) model, and then we extend the VAR(p) model to mean-plus-noise disturbance terms. In chapter 3, we mainly explore the estimation procedure for all parameters in MPNVAR model by conventional Bayesian approach. In chapter 4, we show the Bayesian estimation results for Kilian's trivariate model and for with the help of corresponding impulse response function model. We compare impulse response functions at the end. We conclude in chapter 5.

CHAPTER 2

Mean Plus Noise Models

2.1 Basic Form of Mean Plus Noise Models

First, we introduce a multivariate *Mean-Plus-Noise* (MPN) model, which only contains the disturbance terms. MPN model has two disturbance terms. The first one, U_t, is in a conventional form, which is normally distributed. The second disturbance term, V_t, is also normally distributed controlled by a binary scalar variable d_t, which is from a Bernoulli distribution. It represents as below:

$$\underset{n\times 1}{e_t} = \underset{n\times n}{A}\left[\underset{n\times 1}{U_t} + d_t \begin{pmatrix} \underset{n_1\times 1}{V_t} \\ \underset{n_2\times 1}{0} \end{pmatrix}\right] \tag{2.1}$$

where A is a $(n \times n)$ lower triangular matrix, $n = n_1 + n_2$, d_t is a scalar, $U_t \overset{\text{iid}}{\sim} N(0_{n\times 1}, I_n)$, $V_t \overset{\text{iid}}{\sim} N(0_{n_1\times 1}, \Sigma_V)$, Σ_V is a $(n_1 \times n_1)$ diagonal matrix, $d_t \overset{\text{iid}}{\sim} \text{Bernoulli}(p_d)$.

The other way to express e_t is:

$$e_t|d_t \overset{\text{iid}}{\sim} \begin{cases} N\left(0_{n\times 1}, \Sigma_1\right) & d_t = 0 \\ N\left(0_{n\times 1}, \Sigma_2\right) & d_t = 1 \end{cases} \tag{2.2}$$

where

$$\Sigma_1 = AA'$$
$$\Sigma_2 = A \begin{pmatrix} I_{n_1} + \Sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix} A' \tag{2.3}$$

We call the first disturbance term U_t as *Continuous Disturbance* and the second disturbance term V_t as *Discrete Disturbance*.

2.2 Extensions of MPN Models to VAR Models

Next, we extend pure MPN model to MPNVAR model by substituting conventional VAR models' disturbance term into MPN model. We will NOT assume this process is stationary. The reason why we exclude this important assumption is that nonstationarity will not affect the following Bayesian estimation procedure, but at the same time, if included, it will bring cumbersome Bayesian estimation technique, which we try to avoid here.

We define the following system of autoregressive model when $d_t = 1$, as:

$$\begin{aligned} y_{i,t} = \phi_{i,0} &+ \phi_{i,1}^{(1)} y_{1,t-1} + \phi_{i,2}^{(1)} y_{2,t-1} + \cdots \phi_{i,n}^{(1)} y_{n,t-1} \\ &+ \phi_{i,1}^{(2)} y_{1,t-2} + \phi_{i,2}^{(2)} y_{2,t-2} + \cdots \phi_{i,n}^{(2)} y_{n,t-2} \\ &+ \cdots + \phi_{i,1}^{(p)} y_{1,t-p} + \phi_{i,2}^{(p)} y_{2,t-p} + \cdots \phi_{i,n}^{(p)} y_{n,t-p} + u_{i,t} + d_t v_{i,t} \end{aligned} \tag{2.4}$$

and when $d_t = 0$, as:

$$\begin{aligned} y_{i,t} = \phi_{i,0} &+ \phi_{i,1}^{(1)} y_{1,t-1} + \phi_{i,2}^{(1)} y_{2,t-1} + \cdots \phi_{i,n}^{(1)} y_{n,t-1} \\ &+ \phi_{i,1}^{(2)} y_{1,t-2} + \phi_{i,2}^{(2)} y_{2,t-2} + \cdots \phi_{i,n}^{(2)} y_{n,t-2} \\ &+ \cdots + \phi_{i,1}^{(p)} y_{1,t-p} + \phi_{i,2}^{(p)} y_{2,t-p} + \cdots \phi_{i,n}^{(p)} y_{n,t-p} + u_{i,t} \end{aligned} \tag{2.5}$$

where c_i denotes the constant in ith equation and $\phi_{i,j}^{(l)}$ denotes the coefficient of jth variable at lth time lag in the ith equation. $i,\ j = 1, 2, \cdots, n$ and $l = 1, 2, \cdots, p$.

The **VAR(p)** representation form of above model system is:

$$y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + e_t \tag{2.6}$$

where y_t is a $(n \times 1)$ vector of n observations at time t. c denotes an $(n \times 1)$ vector of constants and Φ_i is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, \cdots, p$. And e_t, the special disturbance term is defined in $Eq.$ 2.1.

Further stack the equation system into a **VAR(1)** model, we reach:

$$\Phi(L)Y_t = c + e_t \tag{2.7}$$

where Y_t = $[y_{p+1}, y_{p+2}, \cdots, y_T]'$, $\Phi(L)$ = $[I_n$, -$\Phi_1 L$, -$\Phi_2 L^2$, $\cdots$, -$\Phi_p L^p]$, and L is the lag operator.

So far we have built the MPNVAR model. One thing we need to emphasize here is that this model can describe either stationary or nonstationary process. MPNVAR model estimation procedure by Bayesian approach will be presented in Chapter 3. Next we will show the tool that we can show MPNVAR model's strength: *Impulse Response Functions*.

2.3 Impulse Response Functions

One of the major goals in this paper is to investigate whether MPNVAR model gives better interpretation of oil shocks. We are using Impulse Response Function (IRF) to measure the reaction of all variables in response to external shocks. We calculate 20 steps ahead IRF of this model. And since MPNVAR model having continuous and discrete disturbance terms, which means we'll get two variance-covariance matrix of the error terms, we construct two separate IRF's for each of them.

From Eq. 2.6 we first build the *companion form* of reduced-from IRF, which is:

$$\Psi = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ & I_{n\cdot(p-1)} & & & 0_{n\cdot(p-1)\times n} \end{bmatrix} \tag{2.8}$$

Then $J\Psi^h J'$ where row vector $J = [I_n \quad 0_{n\times n\cdot(p-1)}]$, stands for the h step ahead impulse response of the all variables. (i, j) element of this $(n\times n)$ square sub-matrix of Ψ^h means ith element of y_t in response to the jth element of e_t.
After achieving the above reduced-form IRF, we can further calculate each structural impulse response functions for this model with respect to Σ_1 and Σ_2. Basically, according to Hamilton [9], first we get the Cholesky decomposition of $\Sigma_1 = A_1^{-1}A_1^{-1'}$ and $\Sigma_2 = A_2^{-1}A_2^{-1'}$. Then we calculate $J\Psi^h J'\cdot A_1^{-1}$ and $J\Psi^h J'\cdot A_2^{-1}$. The two $(n\times n)$ square sub-matrix we have now are the structural impulse response function in response to the continuous disturbance terms and the discrete disturbance terms. The final step is simply to plot impulse responses in response to external shock of each variable with respect to both continuous and discrete disturbances.

CHAPTER 3

Estimation Procedure

3.1 The Likelihood Function of the Model

In this chapter, we implement the Bayesian method to estimate all parameters in our MPNVAR model. To use Bayesian method, the first thing we need is to get the likelihood function of the model. Since the model has similarity to the conventional VAR model, we obtain MPNVAR model likelihood function as below:

$$\begin{aligned} L(y|\Phi,\Sigma_V,\Sigma_1,\Sigma_2,d) &= (2\pi)^{-3[\Sigma_{t=1}^{T}(1-d_t)]/2}|\Sigma_1|^{-\Sigma_{t=1}^{T}(1-d_t)/2}\exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma_1^{-1}(M_1\cdot e_t)'\cdot(M_1\cdot e_t))\right) \\ &\times (2\pi)^{-3[\Sigma_{t=1}^{T}d_t]/2}|\Sigma_2|^{-\Sigma_{t=1}^{T}d_t/2}\exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma_2^{-1}(M_2\cdot e_t)'\cdot(M_2\cdot e_t))\right) \\ &\times (1-p)^{\Sigma_{t=1}^{T}(1-d_t)}\times p^{\Sigma_{t=1}^{T}d_t} \end{aligned} \tag{3.1}$$

where $e_t=\Phi(L)Y_t-c; M_2=\operatorname{diag}(\{dt\})$, the diagonal elements of M_2 are from $\{d_t\}$, others are 0's; $M_1=I-\operatorname{diag}(\{d_t\}); d=\{d_1,d_2,\cdots,d_T\}$, Σ_1 and Σ_2 are defined in Eq. 2.3.

Or same likelihood in another expression:

$$
\begin{aligned}
L(y|\Phi,\Sigma_V,\Sigma_1,\Sigma_2,d) &= (2\pi)^{-3[\Sigma_{t=1}^{T}(1-d_t)]/2}|\Sigma_1|^{-\Sigma_{t=1}^{T}(1-d_t)/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}[(1-d_t)e_t'\Sigma_1^{-1}e_t]\right) \\
&\times\ (2\pi)^{-3[\Sigma_{t=1}^{T}d_t]/2}|\Sigma_2|^{-\Sigma_{t=1}^{T}d_t/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}[d_te_t'\Sigma_2^{-1}e_t]\right) \\
&\times\ (1-p)^{\Sigma_{t=1}^{T}(1-d_t)}\times p^{\Sigma_{t=1}^{T}d_t}
\end{aligned} \tag{3.2}
$$

where $e_t = Y_t - \Phi(L)Y_{t-1}$, Σ_1 and Σ_2 are defined in Eq. 2.3.

The reason why we give out two forms of likelihood function is that these two forms of likelihood function will be required during the derivation of different parameters' posterior distribution.

With the likelihood function of this model, we can implement Bayesian method to estimate each parameter sequentially in this model. we estimate the parameters in the following order: slope coefficients matrix $\Phi(L)$ or Φ, probability parameter p in the Bernoulli distribution, latent variable $\{d_t\}$, variance matrix Σ_V in discrete disturbance terms and Σ_1. Σ_2 is determined by Σ_V and Σ_1. This will not lose the generality since later we use the Gibbs Sampler method to simulate these estimates iteratively.

3.2 Estimation of Φ

We impose the prior of Φ which is multivariate normally distributed ($\Phi_0 \sim N(\mu_\Phi, \Sigma_\Phi)$), where Σ_Φ is considered as given. Since the likelihood function is similar to multivariate normal likelihood, the posterior of Φ likely have a conjugate distribution. The proof is showed in below.

First, we clean up the likelihood function, only keeping the terms having Φ, which would be:

$$
\exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}(M_1\cdot e_t)'\cdot(M_1\cdot e_t))\right)\cdot\exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma_2^{-1}(M_2\cdot e_t)'\cdot(M_2\cdot e_t))\right)
$$

For the convenience of deriving the posterior of Φ, mainly for dealing with the Kronecker operator, we first need to vectorize the whole equation system and use the technique below.

3.2.1 Vectorized VAR(q) model

In previous section we have built our VAR(1) model as:

$$
Y_t = \Phi(L)Y_{t-1} + e_t \qquad e_t = \begin{cases} u_t \sim N(0,\Sigma_1) & dt = 0 \\ v_t \sim N(0,\Sigma_2) & dt = 1 \end{cases} \tag{3.3}
$$

Since later we need to use vectorized VAR(1) model to achieve the posterior of Φ, referring to Canova's notes (2010) [2], we let $y = vec(Y_t)$, $X = [y_{t-1}, \cdots, y_{t-p}]$, $\Phi = vec(\Phi(L))$ and $e = vec(e_t)$. Then the vectorized VAR(1) model would be:

$$y = (I_3 \otimes X)\Phi + e \qquad e \sim N(0, \Sigma_1 \otimes M_1 + \Sigma_2 \otimes M_2) \tag{3.4}$$

where y, e are $(n \cdot T \times 1)$ vectors, I_3 is the identify matrix, and Φ is a $(n \cdot (n \cdot p + 1) \times 1)$ vector.
$M_1 = I_T - \text{diag}(\{dt\})$ and $M_2 = \text{diag}(\{dt\})$

Then the likelihood function would be:

$$\begin{aligned} L(y|\Phi, \Sigma_1, \Sigma_2, d) = & \frac{1}{(2\pi^{0.5 \cdot nT})} |\Sigma_1 \otimes M_1 + \Sigma_2 \otimes M_2|^{-0.5} \\ & \times \exp\{-0.5(y - (I_3 \otimes X)\Phi)'(\Sigma_1^{-1} \otimes M_1 + \Sigma_2^{-1} \otimes M_2)(y - (I_3 \otimes X)\Phi)\} \end{aligned} \tag{3.5}$$

where: $d = \{d_1, d_2, \cdots, d_T\}$

Now we can make some manipulations of the likelihood function focusing on terms having Φ:

$$\begin{aligned} & (y - (I_3 \otimes X)\Phi)'(\Sigma_1^{-1} \otimes M_1 + \Sigma_2^{-1} \otimes M_2)(y - (I_3 \otimes X)\Phi) \\ = & (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2) \cdot (y - (I_3 \otimes X)\Phi)' \cdot (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2) \cdot (y - (I_3 \otimes X)\Phi) \\ = & (\Delta y - \Xi\Phi)'(\Delta y - \Xi\Phi) \\ \text{where} \quad & \Delta = (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2) \qquad \Xi = (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2)(I_3 \otimes X) \\ \text{Also} \quad & (\Delta y - \Xi\Phi) = \Delta y - \Xi\hat{\Phi} + \Xi(\hat{\Phi} - \Phi) \\ \text{where} \quad & \hat{\Phi} = ((\Sigma_1^{-1} \otimes M_1 + \Sigma_2^{-1} \otimes M_2) \otimes X'X)^{-1}((\Sigma_1^{-1} \otimes M_1 + \Sigma_2^{-1} \otimes M_2) \otimes X)y \end{aligned}$$

3.2.2 Posterior of Φ Condition on Other Parameters

In this paper we are going to use $f_0(\cdot)$ to denote the prior pdf and $f_1(\cdot)$ as the posterior pdf. Now we can explicitly write out the prior of $\Phi_0 \sim N(\bar{\Phi}, \Sigma_\Phi)$,which is:

$$\begin{aligned} f_0(\Phi) \propto & |\Sigma_\Phi|^{-0.5} \exp[-0.5(\Phi - \bar{\Phi})'\Sigma_\Phi^{-1}(\Phi - \bar{\Phi})] \\ = & |\Sigma_\Phi|^{-0.5} \exp[-0.5(\Sigma_\Phi^{-0.5}(\Phi - \bar{\Phi}))'\Sigma_\Phi^{-0.5}(\Phi - \bar{\Phi})] \end{aligned} \tag{3.6}$$

Then the posterior of Φ would be proportional to the product of likelihood and prior:

$$\begin{aligned} f_1(\Phi|y;\Theta) &\propto g(\Phi)L(y|\Phi;\Theta) \\ &\propto \exp[-0.5(\Sigma_{\Phi}^{-0.5}(\Phi-\bar{\Phi}))'\Sigma_{\Phi}^{-0.5}(\Phi-\bar{\Phi})] \times \exp(\Delta y - \Xi\Phi)'(\Delta y - \Xi\Phi) \\ &= \exp\{-0.5[(\Phi-\Phi^*)'Z'Z(\Phi-\Phi^*) + (z-Z\Phi^*)'(z-Z\Phi^*)]\} \end{aligned} \tag{3.7}$$

where:

$$z = [\Sigma_{\Phi}^{-0.5} \cdot \Phi_0 \quad \Delta y]' \tag{3.8}$$

$$Z = [\Sigma_{\Phi}^{-0.5} \quad \Xi]' \tag{3.9}$$

$$\Phi^* = (Z'Z)^{-1}(Z'z) \tag{3.10}$$

where $\quad \Delta = (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2), \qquad \Xi = (\Sigma_1^{-0.5} \otimes M_1 + \Sigma_2^{-0.5} \otimes M_2)(I_3 \otimes X)$
Since Σ_1, Σ_2 and Σ_Φ are fixed, the second term above is a constant and

$$\begin{aligned} f_1(\Phi|y) &\propto \exp[-0.5(\Phi-\Phi^*)'Z'Z(\Phi-\Phi^*)] \\ &\propto \exp[-0.5(\Phi-\Phi^*)'(\Sigma_{\Phi}^*)^{-1}(\Phi-\Phi^*)] \end{aligned} \tag{3.11}$$

So Posterior of Φ is:

$$\begin{gathered} \Phi_1 \sim N(\Phi^*, \Sigma_{\Phi}^*) \\ \Phi^* = (Z'Z)^{-1}(Z'z) \qquad \Sigma_{\Phi}^* = (Z'Z)^{-1} \end{gathered} \tag{3.12}$$

3.3 Posterior of p Condition on Other Parameters

To estimate parameter p, we impose prior distribution of parameter p as beta distribution. So the pdf of Beta distribution is as below:

$$f_0(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{3.13}$$

Use the same way as we did in last section, we get the posterior of $f_1(p|\Phi,\Sigma_1,\Sigma_2,d)$ is proportional to the product of likelihood and its prior:

$$f_1(p|\Phi,d,\Sigma_V,\Sigma_1,\Sigma_2) \propto p^{\alpha-1+\sum_{t=1}^{T} d_t}(1-p)^{\beta-1+\sum_{t=1}^{T}(1-d_t)} \tag{3.14}$$

Obviously, the posterior of p is still conventional Beta distribution.

$$p_1 \sim \text{Beta}(\alpha + \sum_{t=1}^{T} d_t, \beta + \sum_{t=1}^{T}(1-d_t)) \tag{3.15}$$

3.4 Posterior of d Condition on Other Parameters

d denotes as $\{d_t; t = 1, 2, \cdots, T\}$. Unlike we did before, since d are latent variables in this model, we need to reorganize the model likelihood function (Eq. 3.2) only keep the terms having d. Then we have the posterior of d is:

$$f_1(d|\Phi, p, \Sigma_V, \Sigma_1, \Sigma_2) \propto [A(1-p)]^{T-\Sigma_{t=1}^{T} d_t} \cdot [B(p)]^{\Sigma_{t=1}^{T} dt} \tag{3.16}$$

where

$$\begin{aligned} A &= (2\pi)^{-3/2[\Sigma_{t=1}^{T}(1-d_t)]}|\Sigma|^{-1/2\Sigma_{t=1}^{T}(1-d_t)} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}[(1-d_t)e_t'\Sigma^{-1}e_t]\right) \\ B &= (2\pi)^{-3/2[\Sigma_{t=1}^{T} d_t]}|\Sigma_2|^{-1/2\Sigma_{t=1}^{T} d_t} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}[d_t e_t'\Sigma_2^{-1}e_t]\right) \end{aligned} \tag{3.17}$$

Or:

$$f_1(d|\Phi, p, \Sigma_V, \Sigma_1, \Sigma_2) \propto \begin{cases} \left(2\pi^{-\frac{3}{2}} \cdot |\Sigma_2|^{-\frac{1}{2}} \cdot p\right) \cdot \exp\left(-\frac{1}{2} \cdot e_t'\Sigma_2^{-1}e_t\right) & \text{if } d_t = 1 \\ \left(2\pi^{-\frac{3}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot (1-p)\right) \cdot \exp\left(-\frac{1}{2} \cdot e_t'\Sigma^{-1}e_t\right) & \text{if } d_t = 0 \end{cases} \tag{3.18}$$

Where: $e_t = Y_t - \Phi(L)Y_{t-1}$. Obviously, the posterior of d is still conventional Bernoulli distribution by normalizing above posterior pdf function.

3.5 Posterior of Σ_V Condition on Other Parameters

To estimate Σ_V, we first need to filter the likelihood function to keep those terms having Σ_V. Although we do not explicitly include Σ_V in the prior form of likelihood, Σ_V only relates to Σ_2. So we only keep terms of Σ_2 in Eq. 3.2 and substitute it with the form of Σ_V in Eq. 2.3:

$$
\begin{aligned}
& g(y|\Phi,\Sigma_V,\Sigma_1,\Sigma_2,d) \\
&= \left| A \begin{pmatrix} I_{n_1}+\sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix} A' \right|^{-\sum_{t=1}^{T} d_t/2} \exp\left(-\frac{1}{2}\operatorname{tr}(A'^{-1} \begin{pmatrix} I_{n_1}+\sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix}^{-1} A^{-1}(M_2\cdot e_t)'\cdot(M_2\cdot e_t)) \right) \\
&= \left| \begin{pmatrix} I_{n_1}+\sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix} \right|^{-\sum_{t=1}^{T} d_t/2} \exp\left(-\frac{1}{2}\operatorname{tr}(A'^{-1} \begin{pmatrix} I_{n_1}+\sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix}^{-1} A^{-1}(M_2\cdot e_t)'(M_2\cdot e_t)) \right) \quad (3.19)
\end{aligned}
$$

Then we impose a Inv-Wishart prior to Σ_V to draw its posterior. Following is the general pdf form of Wishart and inv-Wishart distribution for the use of drawing posterior of Σ_V. Assume the prior distribution of parameter Σ^{-1} follows Wishart distribution $f(\Lambda;n,\Sigma)$. Its pdf is as below:

$$
f(\Lambda;n,\Sigma) = \frac{|\Lambda|^{(n-d-1)/2}\exp(-(1/2)\operatorname{tr}(\Lambda\Sigma^{-1}))}{2^{dn/2}\pi^{d(d-1)/4}|\Sigma|^{n/2}\prod_{i=1}^{d}\Gamma((n+1-i)/2)} \quad (3.20)
$$

where $\Lambda = \sum_{a=1}^{N}(X_a-\bar{X})(X_a-\bar{X})'$ is a scale matrix in the pdf, d is the dimension of Σ, $X_a \overset{\text{iid}}{\sim} N(\mu,\Sigma)$ $|\Lambda| = det(\Lambda)$, $|\Sigma| = \det(\Sigma)$ and $f(\Lambda;n,\Sigma) = 0$ unless w is symmetric and positive definite

Substitute Σ_V^{-1} into Eq. 3.20, we have $\Sigma_V^{-1} \sim Wishart(n,\Lambda^{-1})$, which is equivalent to $\Sigma_V \sim Inv-Wishart$. Its pdf is:

$$
f_0(\Sigma_V^{-1};n,\Lambda^{-1}) = \frac{|\Sigma_V|^{-(n-d-1)/2}\exp(-(1/2)\operatorname{tr}(\Sigma_V^{-1}\Lambda))}{2^{dn/2}\pi^{d(d-1)/4}|\Lambda|^{-n/2}\prod_{i=1}^{d}\Gamma((n+1-i)/2)} \quad (3.21)
$$

where $|\Sigma_V| = det(\Sigma_V)$, d is the dimension of Σ, $|\Lambda| = \det(\Lambda)$, $n = \sum_{t=1}^{T} d_t$ and $f(\Sigma_V^{-1},n,\Lambda^{-1}) = 0$ unless Σ_V^{-1} is symmetric and positive definite.

The Posterior of $f_1(\Sigma_V|\Phi,p,d,\Sigma_1,\Sigma_2)$ is proportional to the product of likelihood and its prior:

$$
\begin{aligned}
& f_1(\Sigma_V|\Phi,p,d,\Sigma_1,\Sigma_2) \propto |\Sigma_V|^{-(\sum_{t=1}^{T} d_t-3-1)/2}\exp(-(1/2)\operatorname{tr}(\Sigma_V^{-1}\Lambda)) \\
& \times \left| \begin{pmatrix} I_{n_1}+\Sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix} \right|^{-\sum_{t=1}^{T} d_t/2} \exp\left(-\frac{1}{2}\operatorname{tr}(A'^{-1} \begin{pmatrix} I_{n_1}+\Sigma_V & 0_{n_1} \\ 0_{n_2} & I_{n_2} \end{pmatrix}^{-1} A^{-1}(M_2\cdot e_t)'(M_2\cdot e_t)) \right) \quad (3.22)
\end{aligned}
$$

where $\Lambda = \operatorname{tr}((A^{-1}M_2\cdot e_t)(A^{-1}M_2\cdot e_t)'+I_3)$ where $\Sigma_1 = AA'$.

To draw random variables from this messy posterior pdf, we use the Metropolis-Hastings algorithm.

3.6 Posterior of Σ_1 and Σ_2 Condition on Other Parameters

Because of the relationship between Σ_1 and Σ_2, once we estimated Σ_1, Σ_2 is also estimated. So to keep the proof as simple as possible, we only write out the posterior distribution of Σ_1. Use the same pdf of Wishart distribution expressed above (Eq. 3.21). The Posterior of $f_1(\Sigma_1|\Phi, p, d)$ is proportional to the product of likelihood and its prior:

$$\begin{aligned} f_1(\Sigma_1|\Phi, p, d, \Sigma_V) \propto & |\Sigma_1|^{-(n_0-3-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma_1^{-1}\Lambda_0)\right) \times \\ & |\Sigma_1|^{-\sum_{t=1}^{T}(1-d_t)/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}[(1-d_t)e_t'\Sigma_1^{-1}e_t]\right) \times \\ & |\Sigma_2|^{-\sum_{t=1}^{T}d_t/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}[d_t e_t'\Sigma_2^{-1}e_t]\right) \end{aligned} \tag{3.23}$$

where $e_t = Y_t - \Phi(L)Y_{t-1}$, $n_0 = \sum_{t=1}^{T}(1-d_t)$, $\Lambda_0 = e_t e_t'/(n_0 - np - 1)$. Keeping Σ_2 is because Σ_2 relates to Σ_1.

Unfortunately, this posterior also is not a conventional distribution. We need to implement MCMC method (Metropolis-Hastings algorithm) to draw samplers from this posterior within the Gibbs Sampler steps. The jumping distribution we use will be introduced in next chapter.

CHAPTER 4

Empirical Results

4.1 Empirical Model and Data

This paper is building on the work of Kilian (2009) [10]. We use processed data set from Kilian's paper and the same structural VAR(24) model to reach our results to compare it with Kilian's. To compare our Bayesian estimation results of MPNVAR model with his, we make a change in Kilian's estimation process that is instead of implementing bootstrap as Kilian did in his paper, we use Bayesian method to estimate his conventional VAR(24) model. This change allows us to compare two different results on the same paper.

We inherit Kilian's structural VAR(24) regression model with replaced new disturbance terms (MPN) to obtain our empirical MPNVAR model as (Eq. 4.1). This structural model is a little different from the reduced form VAR(p) model as we described in the last chapter for the structural coefficient matrix A_0. After some manipulation, it is easy to convert the structural model to reduced form model. Then we can follow the estimation procedure we developed previously to make inference of this model. In this particular case, the model system contains 3 variables. They are the *percent change in global crude oil production* denoted as $\Delta prod_t$, the logged *index of real economic activity* calculated by Kilian denoted as rea_t and the logged and deflated *real oil price* rpo_t. The VAR model extends to $p = 24$ time lags. The data is on monthly frequency from Jan. 1973 to Dec. 2007. The details of how Kilian collecting and manipulating his data set, you can find from his well written paper (2009). We impose the discrete shock to each of the shock factor independently. Hence the empirical MPNVAR models are:

$$\textbf{Case I}: \qquad A_0 y_t = c + \sum_{j=1}^{24} A_j y_{t-j} + U_t + \begin{pmatrix} d_t v_{t1} \\ 0 \\ 0 \end{pmatrix} \tag{4.1}$$

$$\textbf{Case II}: \qquad A_0 y_t = c + \sum_{j=1}^{24} A_j y_{t-j} + U_t + \begin{pmatrix} 0 \\ d_t v_{t2} \\ 0 \end{pmatrix} \tag{4.2}$$

$$\textbf{Case III}: \qquad A_0 y_t = c + \sum_{j=1}^{24} A_j y_{t-j} + U_t + \begin{pmatrix} 0 \\ 0 \\ d_t v_{t3} \end{pmatrix} \tag{4.3}$$

where $U_t \overset{\text{iid}}{\sim} N(0_{3\times 1}, I_3)$, $v_{ti} \overset{\text{iid}}{\sim} N(0, \sigma_{vi}^2)$, $d_t \overset{\text{iid}}{\sim} \text{Bernoulli}(p_d)$.

U_t denotes the vector of serially and mutually uncorrelated *continuous structural innovations* and v_t's denotes the corresponding discrete shocks in each independent model, i.e. v_{t1} denotes *discrete oil supply shock*, v_{t3} denotes *discrete aggregate demand shock* and v_{t3} denotes *discrete oil specific-demand shock*. They are all indexed by a binary variable d_t which indicates if an infrequent event triggers discrete oil shocks. Like what Kilian did in his paper, we postulate that A_0^{-1} has a recursive structure such that the reduced form errors e_t can be decomposed according to $e_t = A_0^{-1} U_t + d_t A_0^{-1} V_t$ by labeling $A_0^{-1} = A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. So we have:

Case I:

$$e_t \equiv \begin{pmatrix} e_t^{\Delta prod_t} \\ e_t^{rea_t} \\ e_t^{rpo_t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \left[\begin{pmatrix} u_t^{cons\ oil\ supply\ shock} \\ u_t^{cons\ aggregate\ demand\ shock} \\ u_t^{cons\ oil\ specific-demand\ shock} \end{pmatrix} + d_t \begin{pmatrix} v_{t1} \\ 0 \\ 0 \end{pmatrix} \right] \tag{4.4}$$

where $\Sigma_1 = A_0^{-1} A_0^{-1\prime}$, $\Sigma_2 = A_0^{-1} \begin{pmatrix} 1+\sigma_v^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} A_0^{-1\prime}$.

Case II:

$$e_t \equiv \begin{pmatrix} e_t^{\Delta prod_t} \\ e_t^{rea_t} \\ e_t^{rpo_t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \left[\begin{pmatrix} u_t^{cons\ oil\ supply\ shock} \\ u_t^{cons\ aggregate\ demand\ shock} \\ u_t^{cons\ oil\ specific-demand\ shock} \end{pmatrix} + d_t \begin{pmatrix} 0 \\ v_{t2} \\ 0 \end{pmatrix} \right] \tag{4.5}$$

where $\Sigma_1 = A_0^{-1}A_0^{-1\prime}$, $\Sigma_2 = A_0^{-1}\begin{pmatrix}1 & 0 & 0\\ 0 & 1+\sigma_v^2 & 0\\ 0 & 0 & 1\end{pmatrix}A_0^{-1\prime}$.

Case III:

$$e_t \equiv \begin{pmatrix} e_t^{\Delta prod_t} \\ e_t^{rea_t} \\ e_t^{rpo_t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \left[\begin{pmatrix} u_t^{cons\ oil\ supply\ shock} \\ u_t^{cons\ aggregate\ demand\ shock} \\ u_t^{cons\ oil\ specific-demand\ shock} \end{pmatrix} + d_t \begin{pmatrix} 0 \\ 0 \\ v_{t3} \end{pmatrix} \right] \tag{4.6}$$

where $\Sigma_1 = A_0^{-1}A_0^{-1\prime}$, $\Sigma_2 = A_0^{-1}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1+\sigma_v^2\end{pmatrix}A_0^{-1\prime}$.

By multiplying A_0^{-1} on both sides of the structural VAR(24) models, we can achieve the reduced form VAR(24) models to implement the estimation method we developed in the previous chapter:

$$y_t = c + \sum_{j=1}^{24} \Phi_j y_{t-j} + e_{ti} \qquad i = 1,2,3 \tag{4.7}$$

where $\Phi_j = A_0^{-1}A_j$, $t = 1,2,\cdots,395$ and this reduced form model (Eq. 4.7) is the same as Eq. 2.6. The corresponding parameters in Eq. 2.7 are: $n = 3$, $p = 24$, $T = 419 - p = 395$, the Φ_j is a (3×3) matrix, y_t is a (3×1) matrix and e_{ti} hereby is a (3×1) matrix. Now we finally build our reduced form MPNVAR(24) model.

4.2 Bayesian Estimation Results for Kilian's Trivatiate VAR Model

In Kilian's original paper, he used bootstrap method to estimate his trivariate VAR(24) model. In this case, bootstrap is a good choice for its estimation efficiency and accuracy. Again since we use Bayesian approach to estimate parameters, it is more reasonable to compare our results with same Bayesian estimation results from his trivariate model. Hence, we use Bayesian approach to replicate his estimation results. Figure 4.1 is the estimated impulse response of Global Crude Oil Production, Real Economic Activity and Real Oil Price to 3 conventional structural disturbances imposed by Kilian: Continuous Oil Supply Shock, Continuous Aggregate Demand Shock, Continuous Oil Specific Demand Shock by implementing Bayesian method. Although we use different approaches to estimate the same model, we found little difference from his result after 10,000 replications in Figure 4.1. All 9 plots are shaped in exact the same way. Bandwidth difference is caused by instead of using one- and two-standard error bands by Kilian, we using 68% and 95% band to plot IRF. This experiment verifies Bayesian method is applicable for his oil shocks investigation case.

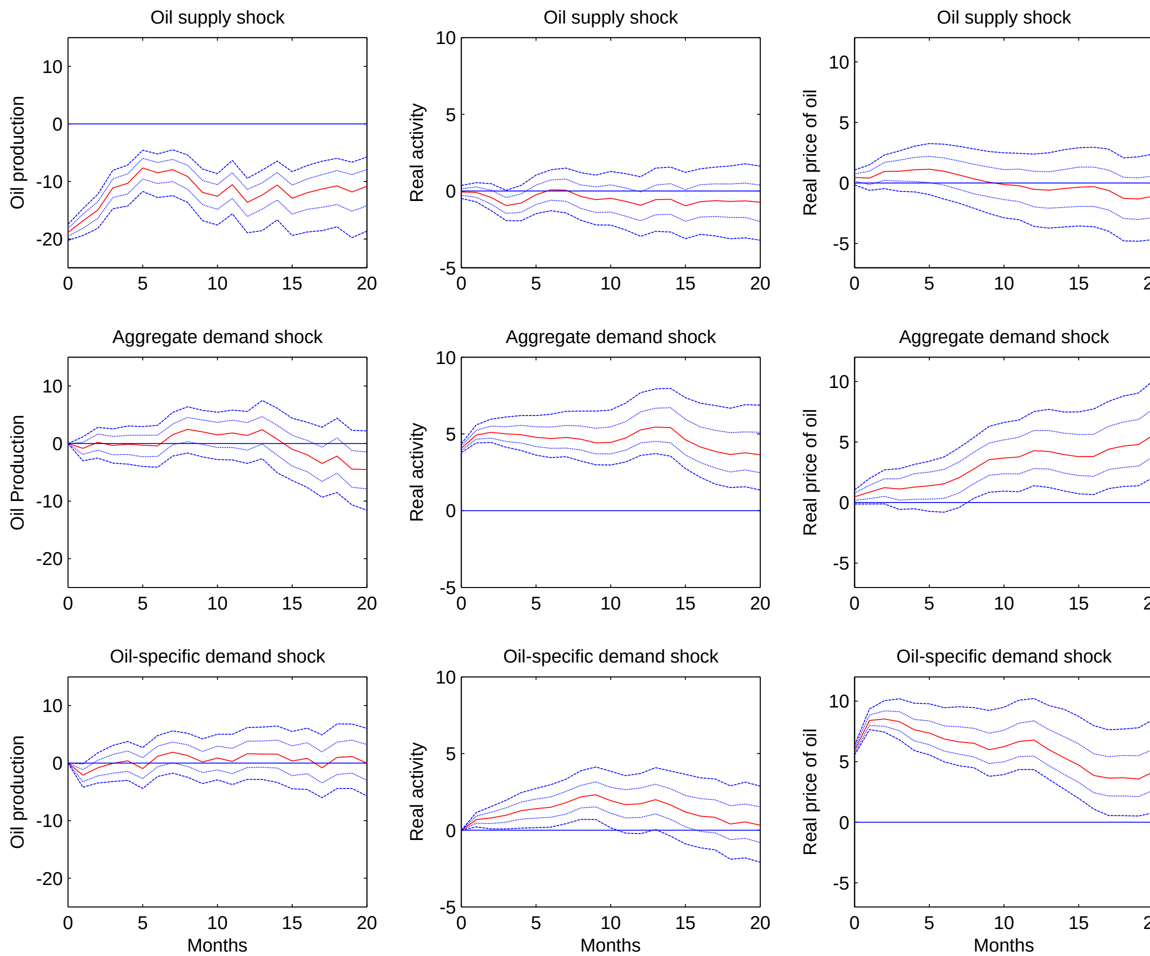

Figure 4.1: Kilian's Trivatiate VAR Model IRF with 95% and 68% Credible Intervals (Median Line in Middle) by Bayesian Estimation

4.3 Details of Parameters Estimation Procedure for MPNVAR Model

After this preparation, we move forward to estimate our innovative MPNVAR model. All three cases are estimated in the same way. We follow the procedure in chapter 3. Here we give out all details about setups, priors we use, etc.

Before the MCMC step, we setup all initial parameters value by their corresponding OLS estimated values. For the Φ prior, we use Φ_{OLS} as its mean, and build a (219×1) variance matrix with random numbers in uniform distribution from 1 to 10 automatically generated by MATLAB built-in function. With the prior we can easily generate random values from Φ's posterior distribution $N(\Phi^*, \Sigma_\Phi^*)$, then update residuals by new Φ for remainder steps.

p's posterior distribution is straightforward. We simply use MATLAB built-in Beta function to generate random variables by setting $\alpha = 6$ and $\beta = 8$. We tried many other value, this combination makes convergence speed faster than smaller values and seems more reasonable than bigger values. And these initial values setting also helps avoiding getting extreme σ_v value.

With updated information of Φ, new residuals and p, we generate d series one by one from a binomial function in MATLAB using normalized parameter. By saying normalizing, we calculate corresponding values of A and B in Eq. 3.17, and then set fraction $B/(A+B)$ as the probability parameter in binomial distribution.

For σ_v we use Metropolis-Hastings algorithm to draw random variables from the messy posterior pdf. One thing needs to be mentioned is that since this particular model has 24 lags and 395 time periods, they cause over-float issues in MATLAB when we calculate some powers in the expression. So we first take the log to avoid over-float issue, then take the exp at the accept-reject step in Metropolis-Hastings. And the jumping distribution we choose here is simply uniform distribution.

Last parameters we estimate are Σ_1 and Σ_2. We use pretty much the same approach as we did in the σ_v step for their similarity except for the choice of jumping distributions we use. During the exploration of the jumping distribution, we find that the biggest contributor to the $\log(\cdot)$ value of the posterior is from the $\frac{-(n_0-3-1)}{2}\log(\det(\Sigma_1))$ term, which is 1000 times larger than the log value of the rest terms. So to eliminate this effect we impose jumping the distribution's pdf to be:

$$\frac{J_t(\theta^*|\theta^{(t-1)})}{J_t(\theta^{(t-1)}|\theta^*)} = \frac{\exp(\det(\Sigma_1^{(t-1)}))^{n_0-np-1}}{\exp(\det(\Sigma_1^*))^{n_0-np-1}}$$

where Σ_1^* denotes the candidate of Σ_1 and $\Sigma_1^{(t-1)}$ denotes the most recent accepted Σ_1 draw.

Along with each iteration of parameter estimation, we record the structural impulse response functions of the model with continuously updated parameters we simulated before.

4.4 Estimates Report

Table 4.1 shows the estimates after 25,000 draws (first 10% draws discarded) from the Gibbs Sampler method for each case. Note that since the slope coefficient matrix Φ contains 219 elements, giving out all estimates of them is not helpful. The estimate we show below is a (3×3) matrix averaged over all time lags. Its expression is:

$$\hat{\bar{\Phi}} = \frac{1}{24}\sum_{j=1}^{24}\Phi_j \qquad j = 1, 2, \cdots, 24$$

Table 4.1: MPNVAR Model Estimates (Posterior Mean)

Parameters	Results	95% Credible Interval
Case I		
$\hat{\sigma}_v$	8.2567	(4.0229, 17.7276)
$\hat{\hat{\Phi}}$	$\begin{pmatrix} -0.6594 & 0.0292 & 0.0695 \\ -0.0484 & 0.9719 & 0.0749 \\ -0.0360 & 0.0031 & 0.9857 \end{pmatrix}$	
$\hat{\Sigma}_1$	$\begin{pmatrix} 353.1727 & 5.3157 & -10.5550 \\ 5.3157 & 17.9109 & 1.0338 \\ -10.5550 & 1.0338 & 35.1841 \end{pmatrix}$	
$\hat{\Sigma}_2$	$\begin{pmatrix} 715.3917 & 9.1030 & -18.8757 \\ 9.1030 & 18.0416 & 0.9212 \\ -18.8757 & 0.9212 & 223.8713 \end{pmatrix}$	
$\hat{p}$	0.0699_{mean} 0.0506_{mode}	(0.0254, 0.1366)
Case II		
$\hat{\sigma}_v$	5.9403	(3.7262, 10.5315)
$\hat{\hat{\Phi}}$	$\begin{pmatrix} -0.6929 & 0.0282 & 0.0743 \\ -0.0487 & 0.9718 & 0.0742 \\ -0.0367 & 0.0032 & 0.9864 \end{pmatrix}$	
$\hat{\Sigma}_1$	$\begin{pmatrix} 345.5075 & 11.2860 & -12.5989 \\ 11.2860 & 16.5503 & 0.5892 \\ -12.5989 & 0.5892 & 37.4496 \end{pmatrix}$	
$\hat{\Sigma}_2$	$\begin{pmatrix} 345.5075 & 11.2860 & -12.5989 \\ 11.2860 & 80.6248 & 0.5892 \\ -12.5989 & 0.5892 & 37.4496 \end{pmatrix}$	
$\hat{p}$	0.1071_{mean} 0.1013_{mode}	(0.0401, 0.1915)
Case III		
$\hat{\sigma}_v$	8.042	(3.945, 18.2527)
$\hat{\hat{\Phi}}$	$\begin{pmatrix} -0.6929 & 0.0282 & 0.0743 \\ -0.0487 & 0.9718 & 0.0742 \\ -0.0367 & 0.0032 & 0.9864 \end{pmatrix}$	
$\hat{\Sigma}_1$	$\begin{pmatrix} 360.1163 & 5.7069 & -11.3732 \\ 5.7069 & 17.8107 & 0.9248 \\ -11.3732 & 1.0332 & 34.8246 \end{pmatrix}$	
$\hat{\Sigma}_2$	$\begin{pmatrix} 360.1163 & 5.7069 & -11.3732 \\ 5.7069 & 17.8107 & 0.9248 \\ -11.3732 & 1.0332 & 266.4268 \end{pmatrix}$	
$\hat{p}$	0.0653_{mean} 0.0481_{mode}	(0.0236, 0.1270)

4.5 Convergence Test Results for Mixture Model

After achieving tens of thousands of draws to estimate the parameters in this model, we need to make sure all estimates have come to convergence when we stop the simulation procedure. Since we iteratively get the draws by implementing Gibbs Sampler method, these draws can be considered as time series data. To assess whether or not these estimates have converged to the desired posterior distribution in a certain finite sample, we use the *convergence diagnosis (CD) test* developed by Geweke (1992) [7]. Following is Geweke's CD test algorithm. We have θ_1 denotes the first subsample, which having total T_1 draws and θ_2 from the second subsample which having T_2 draws. From Geweke's paper, T_1 is the first 10% of M draws. M here is draws after discarding first 10% burn-in period of the original data draws. T_2 is the bottom 50% of M draws. Geweke found that if T_1/M and T_2/M are fixed, and $(T_1+T_2)/M < 1$, then as $M \to \infty$ then:

$$\frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{Avar(\bar{\theta}_1)/T_1 + Avar(\bar{\theta}_2)/T_2}} \Rightarrow N(0,1) \tag{4.8}$$

where $Avar(\bar{\theta}_1)$ and $Avar(\bar{\theta}_2)$ are the long-run asymptotic variance of θ_1 and θ_2. And their estimates are defined as below:

$$\begin{cases} \hat{A}var(\bar{\theta}_1) = \hat{\Gamma}_0^{(1)} + 2\sum_{j=1}^{K_1}(1 - \frac{j}{K_1+1})\hat{\Gamma}_j^{(1)} \\ \hat{A}var(\bar{\theta}_2) = \hat{\Gamma}_0^{(2)} + 2\sum_{j=1}^{K_2}(1 - \frac{j}{K_2+1})\hat{\Gamma}_j^{(2)} \end{cases} \tag{4.9}$$

In the Eq. 4.9, $\hat{\Gamma}_j^{(i)}$ denotes the consistent estimate for the jth sample auto-covariance matrix in the ith subsample set. $(1 - \frac{j}{K_i+1})$ is called *Bartlett Kernel*. and the K_i is the *Bandwidth* for the CD test. Literature shows that the CD statistics are sensitive to the choice of both Bartlett Kernel and Bandwidth. And the choice of Bandwidth is the key for the test. In this thesis we use the method of choosing bandwidth has been brought by Newey and West (1994) [14]. There are also other methods available to choose the bandwidth. Here is how we apply Newey and West's automatic bandwidth selection method:

$$\begin{aligned} \hat{\Gamma}_j^{(i)} &= (1/T_{(i)}) \sum_{t=j+1}^{T_{(i)}} (\theta_t^{(i)} - \bar{\theta}^{(i)})(\theta_{t-j}^{(i)} - \bar{\theta}^{(i)}) \\ n &= [4(T_{(i)}/100)^{2/9}] \\ \hat{s}_a &= 2\sum_{j=1}^{n} j\hat{\Gamma}_j^{(i)} \qquad \hat{s}_b = \Gamma_0^{(i)} + 2\sum_{j=1}^{n} \hat{\Gamma}_j^{(i)} \\ \hat{\gamma} &= 1.1447(\{\hat{s}_a/\hat{s}_b\}^2)^{1/3} \\ K_{(i)} &= [\hat{\gamma} \cdot T_{(i)}^{1/3}] \end{aligned} \tag{4.10}$$

where $[\cdot]$ is the integer part.

We have the H_0 hypothesis: parameter simulated by Gibbs Sampler method are converged in certain finite samples. And we calculate the convergence test statistic for θ by:

$$CD(\theta) = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{Avar(\bar{\theta}_1)/T_1 + Avar(\bar{\theta}_2)/T_2}} \tag{4.11}$$

If $|CD(\theta)|$ is greater than 1.96, we reject the H_0 hypothesis with Type I error $\alpha = 5\%$. If not then we fail to reject the H_0 hypothesis with Type I error $\alpha = 5\%$.

With above test method, we ran the convergence test for each of the parameters in our MPN model with significant level 5% (22,500 draws left after burn-in period). Table 4.2 shows the results:

Table 4.2: Convergence Test Results

Parameters	Numbers of Elements	Rejection Ratio
	Case I	
p	1	100%
σ_v	1	100%
Σ_1 & Σ_2	18	66.67%
$\Phi(L)$	219	39.73%
	Case II	
p	1	0%
σ_v	1	0%
Σ_1 & Σ_2	18	5.56%
$\Phi(L)$	219	7.31%
	Case III	
p	1	0%
σ_v	1	0%
Σ_1 & Σ_2	18	11.11%
$\Phi(L)$	219	2.48%

Since HAC test will over-reject more than 5% and there are only small amount of elements in the first two parameters, our test result meets our expectation quite well. So we can say our estimates have converged after 25,000 replications of MCMC method.

4.6 Monte Carlo Experiment Report

To test if our estimation procedure is plausible, we do some Monte Carlo experiment. We use estimated parameters to generate simulated data. Then we repeat the estimation procedure in Chapter 3 to get estimates of simulated data. Compare them with results showed in section 4.4 to see if they are reasonably close. The Monte Carlo experiment results (Table 4.3) reasonably indicate that the estimation procedure is plausible:

4.7 Inference from MPNVAR model

From the estimation results, case I is not reasonable. First reason is that, the supply disturbance is quite stable according to the historic data. The estimation results are more frequent than expected. The other reason why we drop it is that the estimates never converge in our convergence test. Case II and Case III's results meet our initial anticipation. We make inference from Case II and Case III only.

For Case II: After achieving consistently estimated parameters, we analyze them in this section. We found that $\hat{p}_{\text{mean}} = 0.1071$ and $\hat{p}_{\text{mode}} = 0.1013$. This means, by monthly frequency data, discrete oil specific shocks occasionally occur every 9.3 month (by mean). And the other important parameter is $\sigma_v = 5.9$. 5.9 means the discrete disturbance is about 6 times larger than the continuous shocks. As you can see in Figure 4.2 (oil specific-demand shocks plot), if we count peaks beyond +0.29 and -0.29

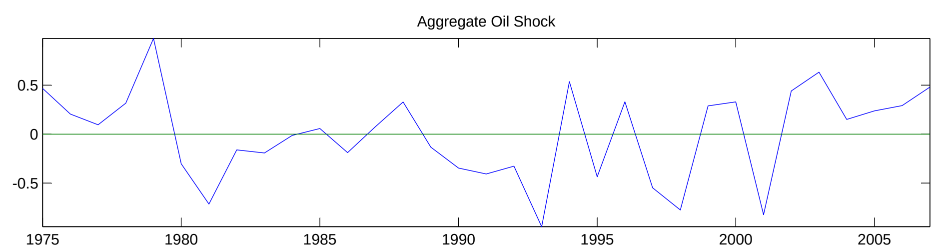

Figure 4.2: Kilian's Historical Decomposition of Real Price of Oil Plot

bound as the discrete disturbance, we get 40 peaks, which is about every 10.2 month one big shock. This approximately matches our 9.3 month frequency in mode sense. +0.29 and -0.29 bound is calculated from $\pm((+1-(-1))/7) = \pm 0.29$.

We use Σ_1 and Σ_2 to build corresponding impulse response function for each case. Figure 4.4 is the IRF plot of continuous disturbance terms in our Mean plus Noise model by ignoring the discrete error term over all time period, i.e. same disturbance in Kilian's model and Figure 4.5 is the IRF plot of discrete disturbance terms over all time period, i.e. at each time period, the oil shocks having both frequent and infrequent shock factors.

Σ_1's IRF is basically the same thing as Bayesian results of Kilian's original model. And no wonder

Table 4.3: Monte Carlo Experiment Estimates (Posterior Mean)

Parameters	Estimated Results	95% Credible Interval
Case I		
$\widehat{(\sigma_v)}_{MC}$	4.3045	(2.3346, 6.3487)
$\widehat{(\Phi)}_{MC}$	$\begin{pmatrix} -1.4361 & -0.0587 & 0.1006 \\ 0.2341 & 0.9930 & 0.0299 \\ -0.1505 & -0.0079 & 0.9895 \end{pmatrix}$	
$\widehat{(\Sigma_1)}_{MC}$	$\begin{pmatrix} 337.3448 & 24.2653 & -15.5950 \\ 24.2653 & 24.2759 & -3.4370 \\ -15.5950 & -3.4370 & 48.5735 \end{pmatrix}$	
$\widehat{(\Sigma_2)}_{MC}$	$\begin{pmatrix} 145.69 & 104.2 & -68.0 \\ 104.2 & 30.3 & -7.3 \\ -68.0 & -7.3 & 51.7 \end{pmatrix}$	
$\hat{p}$	0.0618_{mean} 0.0506_{mode}	(0.0224, 0.1145)
Case II		
$\widehat{(\sigma_v)}_{MC}$	8.4251	(4.6504, 14.6052)
$\widehat{(\Phi)}_{MC}$	$\begin{pmatrix} -1.2300 & -0.0020 & -0.1515 \\ 0.0391 & 0.9204 & 0.0430 \\ -0.0858 & -0.0016 & 0.9863 \end{pmatrix}$	
$\widehat{(\Sigma_1)}_{MC}$	$\begin{pmatrix} 389.1351 & 16.8761 & -33.0577 \\ 16.8761 & 20.6361 & 8.0430 \\ -33.0577 & 8.0430 & 50.1577 \end{pmatrix}$	
$\widehat{(\Sigma_2)}_{MC}$	$\begin{pmatrix} 389.1351 & 16.8761 & -33.0577 \\ 16.8761 & 191.6584 & 8.0430 \\ -33.0577 & 8.0430 & 50.1577 \end{pmatrix}$	
$\hat{p}$	0.0608_{mean} 0.0481_{mode}	(0.0294, 0.1049)
Case III		
$\widehat{(\sigma_v)}_{MC}$	6.265	(3.157, 10.650)
$\widehat{(\Phi)}_{MC}$	$\begin{pmatrix} -1.2035 & 0.0450 & -0.0476 \\ -0.1826 & 0.9586 & 0.0777 \\ -0.0342 & 0.0093 & 0.9767 \end{pmatrix}$	
$\widehat{(\Sigma_1)}_{MC}$	$\begin{pmatrix} 357.7975 & 27.2786 & -10.8794 \\ 27.2786 & 23.7256 & -2.1590 \\ -10.8794 & -2.1590 & 37.8577 \end{pmatrix}$	
$\widehat{(\Sigma_2)}_{MC}$	$\begin{pmatrix} 357.7975 & 27.2786 & -10.8794 \\ 27.2786 & 23.7256 & -2.1590 \\ -10.8794 & -2.1590 & 209.2842 \end{pmatrix}$	
$\hat{p}$	0.0819_{mean} 0.0557_{mode}	(0.0272, 0.1726)

that plot is also consistent as our assumption that discrete innovation is only on oil specific-demand shocks. Because only bottom 3 subplots have dramatic difference from the original model's IRF. The band width is about 4 times lager than Σ_1's, noticing that we change the y-axis scale in the Σ_2's IRF plot. However, the shocks forecast behaviors are very similar in these two IRF plots. Each of them has the correspondingly same curve.

For Case III: After achieving consistently estimated parameters, we analyze them in this section. We found that $\hat{p}_{\text{mean}} = 0.066$ and $\hat{p}_{\text{mode}} = 0.048$. This means, by monthly frequency data, discrete oil specific shocks occasionally occur every 1.7 years (by mode). And the other important parameter is $\sigma_v = 8.0$. 8.0 means the discrete disturbance is about 8 times larger than the continuous shocks. As you can see in Figure 4.3 (oil specific-demand shocks plot), if we count peaks beyond +0.11 and -0.11

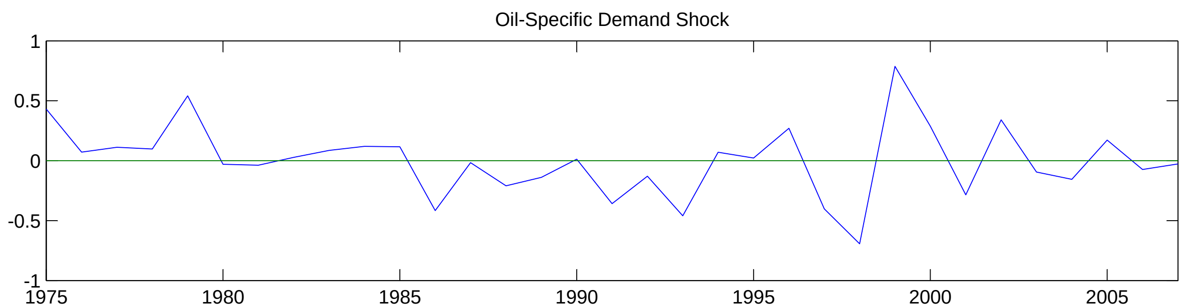

Figure 4.3: Kilian's Historical Decomposition of Real Price of Oil Plot

bound as the discrete disturbance, we get 21 peaks, which is about every 1.6 year one big shock. This approximately matches our 1.7 year frequency in mode sense. +0.11 and -0.11 bound is calculated from $\pm((+1-(-1))/9) = \pm 0.11$.

We use Σ_1 and Σ_2 to build corresponding impulse response function for each case. Figure 4.6 is the IRF plot of continuous disturbance terms in our Mean plus Noise model by ignoring the discrete error term over all time period, i.e. same disturbance in Kilian's model and Figure 4.7 is the IRF plot of discrete disturbance terms over all time period, i.e. at each time period, the oil shocks having both frequent and infrequent shock factors.

Σ_1's IRF is basically the same thing as Bayesian results of Kilian's original model. And no wonder that plot is also consistent as our assumption that discrete innovation is only on oil specific-demand shocks. Because only bottom 3 subplots have dramatic difference from the original model's IRF. The band width is about 4 times lager than Σ_1's, noticing that we change the y-axis scale in the Σ_2's IRF plot. However, the shocks forecast behaviors are very similar in these two IRF plots. Each of them has the correspondingly same curve.

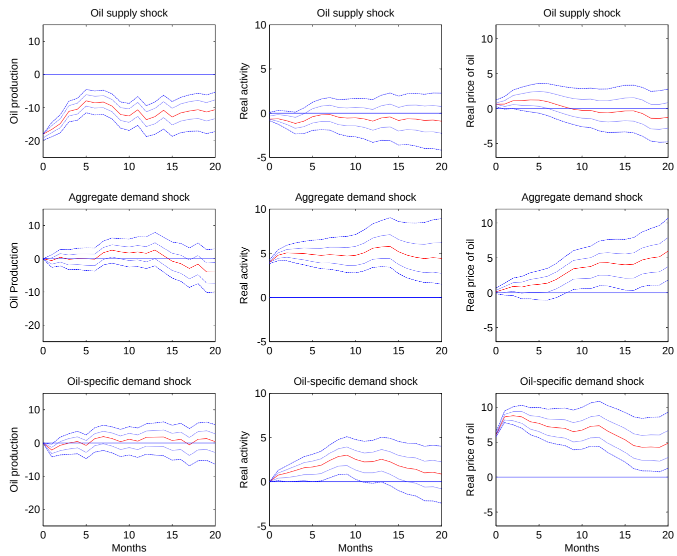

Figure 4.4: Case II: MPN Model IRF of Σ_1 with 95% and 68% Credible Intervals (Median in Middle)

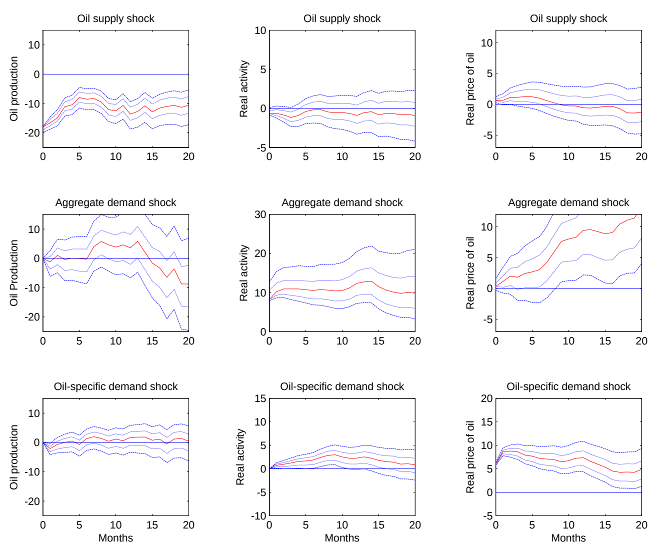

Figure 4.5: Case II: MPN Model IRF of Σ_2 with 95% and 68% Credible Intervals (Median in Middle)

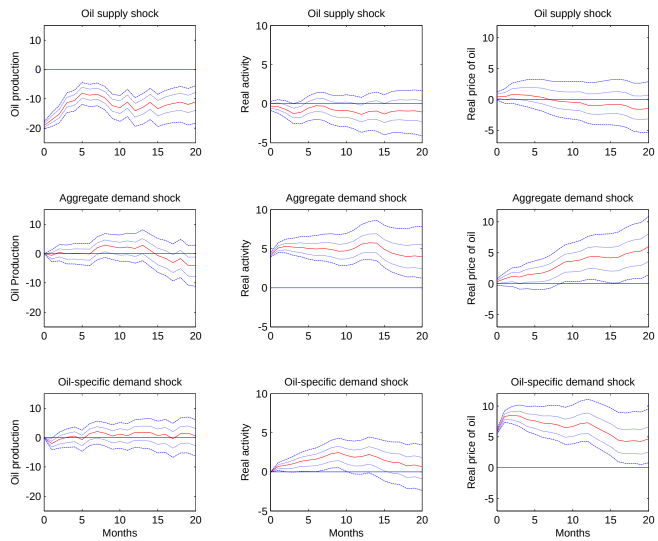

Figure 4.6: Case III: MPN Model IRF of Σ_1 with 95% and 68% Credible Intervals (Median in Middle)

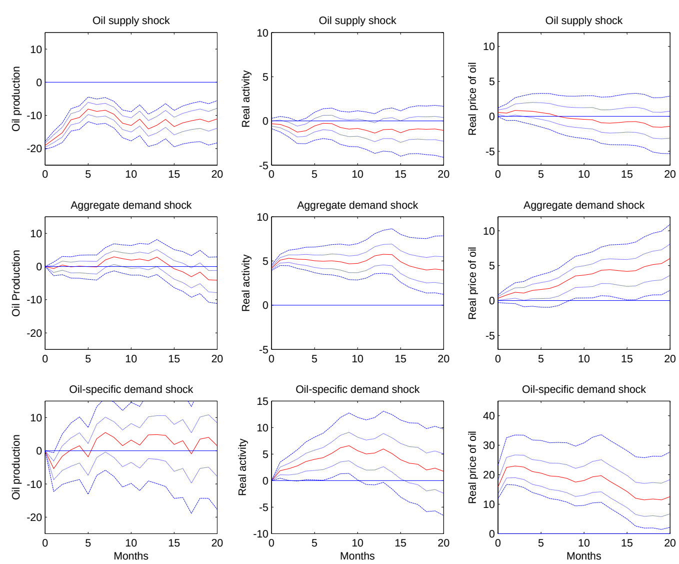

Figure 4.7: Case III: MPN Model IRF of Σ_2 with 95% and 68% Credible Intervals (Median in Middle)

CHAPTER 5

Concluding Remarks

We propose MPNVAR model in structura and reduced VAR forms and apply them to the study of oil shocks. We find that oil specific-demand shock is the biggest contributor to the oil price fluctuation, it roughly occurs every 1.7 years. And the magnitude of this discrete shock is about 8 times of conventional continuous shock. This conclusion matches the historical data in Kilian (2009).

The IRF results show a similar pattern to those of the conventional VAR model. In the future, we will bring MPN model to other dynamic analysis models, trying to find a way get different shaped IRF from the new model. And we are expecting to apply our MPNVAR model to other areas like monetary policy analysis or so on.

REFERENCES

[1] Fabio Canova. *Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting*, volume Volume 1:, page 482. Wiley-Blackwell, 1999.

[2] Fabio Canova. Bayesian var models. Unpublished Note, 2010.

[3] Chung Chen and George C. Tiao. Random level-shift time series models, arima approximations, and level-shift detection. *Journal of Business & Economic Statistics*, 8(1):pp. 83–97, 1990.

[4] Siddhartha Chib and Edward Greenberg. Hierarchical analysis of sur models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, 68(2):339 – 360, 1995.

[5] Francis X. Diebold and Atsushi Inoue. Long memory and regime switching. *Journal of Econometrics*, 105(1):131 – 159, 2001.

[6] Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith. Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412):pp. 972–985, 1990.

[7] John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, 1991.

[8] Clive W.J. Granger and Namwon Hyung. Occasional structural breaks and long memory with an application to the sp 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399 – 421, 2004.

[9] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[10] Lutz Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69, 2009.

[11] Robert Litterman. Forecasting with bayesian vector autoregressions – five years of experience: Robert b. litterman, journal of business and economic statistics 4 (1986) 25-38. *International Journal of Forecasting*, 2(4):497–498, 1986.

[12] Robert B. Litterman. Techniques of forecasting using vector autoregressions. Technical report, 1980.

[13] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[14] Whitney K Newey and Kenneth D West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4):631–53, October 1994.

[15] Giorgio E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):pp. 821–852, 2005.

[16] Christopher A. Sims. Money, income, and causality. *The American Economic Review*, 62(4):pp. 540–552, 1972.

[17] Christopher A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):pp. 1–48, 1980.

[18] Christopher A. Sims and Tao Zha. Were there regime switches in u.s. monetary policy? *The American Economic Review*, 96(1):pp. 54–81, 2006.

[19] Harald Uhlig. Bayesian vector autoregressions with stochastic volatility. *Econometrica*, 65(1):pp. 59–73, 1997.

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