Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications

Jouchi Nakajima

This paper aims to provide a comprehensive overview of the estimation methodology for the time-varying parameter structural vector autoregression (TVP-VAR) with stochastic volatility, in both methodology and empirical applications. The TVP-VAR model, combined with stochastic volatility, enables us to capture possible changes in underlying structure of the economy in a flexible and robust manner. In this respect, as shown in simulation exercises in the paper, the incorporation of stochastic volatility into the TVP estimation significantly improves estimation performance. The Markov chain Monte Carlo method is employed for the estimation of the TVP-VAR models with stochastic volatility. As an example of empirical application, the TVP-VAR model with stochastic volatility is estimated using the Japanese data with significant structural changes in the dynamic relationship between the macroeconomic variables.

Keywords: Bayesian inference; Markov chain Monte Carlo; Monetary policy; State space model; Structural vector autoregression; Stochastic volatility; Time-varying parameter

JEL Classification: C11, C15, E52

Economist, Institute for Monetary and Economic Studies, Bank of Japan. Currently in the Personnel and Corporate Affairs Department (studying at Duke University) (E-mail: jouchi.nakajima@stat.duke.edu)

The author would like to thank Shigeru Iwata, Han Li, Toshiaki Watanabe, Tomoyoshi Yabu, and the staff of the Institute for Monetary and Economic Studies (IMES), Bank of Japan (BOJ), for their useful comments. Views expressed in this paper are those of the author and do not necessarily reflect the official views of the BOJ.

I. Introduction

A vector autoregression (VAR) is a basic econometric tool in econometric analysis with a wide range of applications. Among them, a time-varying parameter VAR (TVP-VAR) model with stochastic volatility, proposed by Primiceri (2005), is broadly used, especially in analyzing macroeconomic issues. The TVP-VAR model enables us to capture the potential time-varying nature of the underlying structure in the economy in a flexible and robust manner. All parameters in the VAR specification are assumed to follow the first-order random walk process, thus allowing both a temporary and permanent shift in the parameters.

Stochastic volatility plays an important role in the TVP-VAR model, although the idea of stochastic volatility is originally proposed by Black (1976), followed by numerous developments in financial econometrics (see, e.g., Ghysels, Harvey, and Renault [2002] and Shephard [2005]). In recent years, stochastic volatility is also more frequently incorporated into the empirical analysis in macroeconomics (e.g., Uhlig [1997], Cogley and Sargent [2005], and Primiceri [2005]). In many cases, a data-generating process of economic variables seems to have drifting coefficients and shocks of stochastic volatility. If that is the case, then application of a model with time-varying coefficients but constant volatility raises the question of whether the estimated time-varying coefficients are likely to be biased because a possible variation of the volatility in disturbances is ignored. To avoid this misspecification, stochastic volatility is assumed in the TVP-VAR model. Although stochastic volatility makes the estimation difficult because the likelihood function becomes intractable, the model can be estimated using Markov chain Monte Carlo (MCMC) methods in the context of a Bayesian inference.

To illustrate the estimation procedure of the TVP-VAR model, this paper begins by reviewing an estimation algorithm for a TVP regression model with stochastic volatility, which is a univariate case of the TVP-VAR model. Then the paper extends the estimation algorithm to the multivariate case. The paper also provides simulation exercises of the TVP regression model to examine its estimation performance against the possibility of structural changes using simulated data. Such simulation exercises show the important role of stochastic volatility in improving the estimation performance.[1]

Regarding the empirical application of the TVP-VAR model, this paper provides empirical illustrations using Japanese macroeconomic data. The estimation results for standard three-variable models reveal the time-varying structure of the Japanese economy and the Bank of Japan's (BOJ's) monetary policy from 1977 to 2007. During the three decades of the sample period, the Japanese economy shows significantly different macroeconomic performance, thus implying the possibility of important structural changes in the economy over time. The time-varying impulse responses show remarkable changes in the relations between the macroeconomic variables.

1. In this regard, the estimation performance of the TVP-VAR model differs significantly, depending on whether the stochastic volatility is incorporated or not. Thus, we use the expression "TVP-VAR model with stochastic volatility" if the inclusion of the stochastic volatility needs to be emphasized. Otherwise, we use just "TVP-VAR model" for simplicity.

The paper is organized as follows. In Section II, the estimation methodology of the TVP regression model is developed. Section III illustrates the simulation study of the TVP regression model focusing on stochastic volatility. In Section IV, the model specification, the estimation scheme, and the literature survey of the TVP-VAR model are provided. Section V presents the empirical results of the TVP-VAR model for Japanese macroeconomic variables. Finally, Section VI concludes the paper.

II. TVP Regression Model with Stochastic Volatility

This section explains the basic estimation methodology of the TVP-VAR model by reviewing an estimation algorithm for a univariate TVP regression model with stochastic volatility.

A. Model

Consider the TVP regression model:

(Regression)

$$y_t = x_t'\beta + z_t'\alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad t = 1, \ldots, n, \tag{1}$$

(Time-varying coefficients)

$$\alpha_{t+1} = \alpha_t + u_t, \quad u_t \sim N(0, \Sigma), \quad t = 0, \ldots, n-1, \tag{2}$$

(Stochastic volatility)

$$\sigma_t^2 = \gamma \exp(h_t), \quad h_{t+1} = \phi h_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad t = 0, \ldots, n-1, \tag{3}$$

where y_t is a scalar of response; x_t and z_t are $(k \times 1)$ and $(p \times 1)$ vectors of covariates, respectively; β is a $(k \times 1)$ vector of constant coefficients; α_t is a $(p \times 1)$ vector of time-varying coefficients; and h_t is stochastic volatility. We assume that $\alpha_0 = 0$, $u_0 \sim N(0, \Sigma_0)$, $\gamma > 0$, and $h_0 = 0$.

Equation (1) has two parts of covariates; one corresponds to the constant coefficients (β) and the other to the time-varying coefficients (α_t). The effects of x_t on y_t are assumed to be time-invariant, while the regression relations of z_t to y_t are assumed to change over time.

The time-varying coefficients α_t are formulated to follow the first-order random walk process in equation (2). It allows both temporary and permanent shifts in the coefficients. The drifting coefficient is meant to capture a possible nonlinearity, such as a gradual change or a structural break. In practice, this assumption implies a possibility that the time-varying coefficients capture not only the true movement but also some spurious movements, because the α_t can *freely move* under the random-walk assumption. In other words, there is a risk that the time-varying coefficients overfit the data if the relations of z_t and y_t are obscure. To avoid such a situation, it might be better to assume a stationarity for the time-varying coefficients. For example, each coefficient can be modeled to follow an AR(1) process where the absolute value of the persistence parameter is less than one. However, in this formulation, a structural change or a permanent shift of the coefficient would be difficult to estimate even if it existed. After all, it is important to choose the model specification of the time-varying coefficients that is considered to be suitable to data of interest, economic theories, and the purpose of analysis (see, e.g., West and Harrison [1997]).

The disturbance of the regression, denoted by ε_t, follows the normal distribution with the time-varying variance σ_t^2. The log-volatility, $h_t = \log \sigma_t^2/\gamma$, is modeled to follow the AR(1) process in equation (3). Similar to the discussion on the assumption of the time-varying coefficients above, the process of log-volatility can be modeled following both stationary and non-stationary processes. For the following analysis in this section, we assume that $|\phi| < 1$ and the initial condition is set based on the stationary distribution as $\eta_0 \sim N(0, \sigma_\eta^2/(1-\phi^2))$. In the case of $\phi = 1$, the log-volatility follows the random walk process. The estimation algorithm for the random-walk case requires only a slight modification for the algorithm developed below.[2]

We can consider reduced models in the class of the TVP regression model. If the regression has only constant coefficients (i.e., $z_t'\alpha_t \equiv 0$), the model reduces to a standard (constant-parameter) linear regression model. If we assume that $\sigma_t^2 = \sigma^2$, for $t = 1, \ldots, n$, the model forms the TVP regression model with the constant variance.

B. Estimation Methodology

1. State space model

Regarding α_t and h_t as state variables, TVP regression forms the state space model. The state space model has been well studied in many fields (see, e.g., Harvey [1993] and Durbin and Koopman [2001] for econometric issues). To estimate the state space model, several methods have been developed. For the TVP regression models, if the variance of disturbance is assumed to be time-invariant (i.e., time-varying coefficient and constant volatility), the parameters are easily estimated using the standard Kalman filter for a linear Gaussian state space model (e.g., West and Harrison [1997]). However, if it has stochastic volatility, the maximum likelihood estimation requires a heavy computational burden to repeat the filtering many times to evaluate the likelihood function for each set of parameters until we reach the maximum, because the model forms a nonlinear state space model. Therefore, we alternatively take a Bayesian approach using the MCMC method for a precise and efficient estimation of the TVP regression model. This also has a great advantage when the model is extended to the TVP-VAR model, as shown later.

2. Bayesian inference and MCMC sampling method

The MCMC method has become popular in econometrics. In recent years, a considerable number of works on empirical macroeconomics have employed the MCMC method. The MCMC method is considered in the context of Bayesian inference, and its goal is to assess the joint posterior distribution of parameters of interest under a certain prior probability density that the researchers set in advance. Given data, we repeatedly sample a Markov chain whose invariant (stationary) distribution is the posterior

2. The estimation algorithm in the case of $\phi = 1$ is provided in the appendix of Nakajima and Teranishi (2009). See also Sekine (2006) and Sekine and Teranishi (2008) for investigation of the macroeconomic issues using the TVP regression model with random-walk stochastic volatility.

distribution. There are many ways to construct the Markov chain with this property (e.g., Chib and Greenberg [1996] and Chib [2001]).[3]

In the Bayesian inference, we specify the prior density, denoted by $\pi(\theta)$, for a vector of the unknown parameters θ. Let $f(y\,|\,\theta)$ denote the likelihood function for data $y = \{y_1, \ldots, y_n\}$. Inference is then based on the posterior distribution, denoted by $\pi(\theta\,|\,y)$, which is obtained by the Bayes' theorem,

$$\pi(\theta\,|\,y) = \frac{f(y\,|\,\theta)\pi(\theta)}{\int f(y\,|\,\theta)\pi(\theta)\,d\theta}.$$

In principle, the prior information concerning θ is updated by observing the data y. The quantity $m(y) = \int f(y\,|\,\theta)\pi(\theta)\,d\theta$ is called the normalizing constant or marginal distribution. In the case where the likelihood function or the normalizing constant is intractable, the posterior distribution does not have a closed form. To overcome this difficulty, many computational methods are developed for sampling from the posterior distribution. Among them, the MCMC sampling methods are popular and powerful algorithms that enable us to sample from the posterior distribution without computing the normalizing constant. The MCMC algorithm proceeds by sampling recursively the conditional posterior distribution where the most recent values of the conditioning parameters are used in the simulation.

The Gibbs sampler is one of the well-known MCMC methods. Consider a vector of unknown parameters $\theta = (\theta_1, \ldots, \theta_p)$. The procedure is constructed as follows:

(1) Choose an arbitrary starting point $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$, and set $i = 0$.

(2) Given $\theta^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$,

- (a) generate $\theta_1^{(i+1)}$ from the conditional posterior distribution $\pi(\theta_1^{(i+1)}\,|\,\theta_2^{(i)}, \ldots, \theta_p^{(i)})$,
- (b) generate $\theta_2^{(i+1)}$ from $\pi(\theta_2^{(i+1)}\,|\,\theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_p^{(i)})$,
- (c) generate $\theta_3^{(i+1)}$ from $\pi(\theta_3^{(i+1)}\,|\,\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, \ldots, \theta_p^{(i)})$,
- (d) generate $\theta_4^{(i+1)}, \ldots, \theta_p^{(i+1)}$, in the same way.

(3) Set $i = i + 1$, and go to (2).

These draws can be used as the basis for making inferences by appealing to suitable ergodic theorems for Markov chains.

For the estimation of the TVP regression model, there are several reasons to use the Bayesian inference and MCMC sampling method. First, the likelihood function is intractable because the model includes the nonlinear state equations of stochastic volatility, which precludes the maximum likelihood estimation method. Also, we cannot assess the normalizing constant and therefore the posterior distribution analytically. Second, using the MCMC method, since not only the parameters $\theta \equiv (\beta, \Sigma, \phi, \sigma_\eta, \gamma)$ but also the state variables $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ and $h = \{h_1, \ldots, h_n\}$ are sampled simultaneously, we can make the inference for the state variables with the uncertainty of the parameters θ. Third, we can estimate the function of the parameters such as an impulse

3. Koop (2003) and Lancaster (2003) are helpful for understanding Bayesian econometrics as a primer. Geweke (2005) and Gamerman and Lopes (2006) cover more comprehensive theories and practices of the MCMC method.

response function with the uncertainty of the parameters θ taken into consideration by using the sample drawn through the MCMC procedure.

C. MCMC Algorithm for the TVP Regression Model

For the TVP regression model, specifying the prior density as $\pi(\theta)$, we obtain the posterior distribution, $\pi(\theta, \alpha, h \mid y)$.[4] There are several ways to implement the MCMC algorithm to explore this posterior distribution, though we develop the implementation using the following algorithm:

(1) Initialize θ, α, and h.
(2) Sample $\beta \mid \gamma, \alpha, h, y$.
(3) Sample $\alpha \mid \beta, \Sigma, \gamma, h, y$.
(4) Sample $\Sigma \mid \alpha$.
(5) Sample $h \mid \beta, \gamma, \phi, \sigma_\eta, \alpha, y$.
(6) Sample $\phi \mid \sigma_\eta, h$.
(7) Sample $\sigma_\eta \mid \phi, h$.
(8) Sample $\gamma \mid \beta, \alpha, h, y$.
(9) Go to (2).

The details of the procedure are illustrated as follows.

1. Sample β

We specify the prior for β as $\beta \sim N(\beta_0, B_0)$. We explore the conditional posterior density of β given by

$$\begin{aligned}
&\pi(\beta \mid \gamma, \alpha, h, y) \\
&\propto \exp\left\{-\frac{1}{2}(\beta - \beta_0)' B_0^{-1}(\beta - \beta_0)\right\} \times \left\{-\frac{\sum_{t=1}^{n}(y_t - x_t'\beta - z_t'\alpha_t)^2}{2\gamma e^{h_t}}\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\beta - \hat{\beta})' \hat{B}^{-1}(\beta - \hat{\beta})\right\},
\end{aligned}$$

where

$$\hat{B} = \left(B_0^{-1} + \sum_{t=1}^{n} \frac{x_t x_t'}{\gamma e^{h_t}}\right)^{-1}, \quad \hat{\beta} = \hat{B}\left(B_0^{-1}\beta_0 + \sum_{t=1}^{n} \frac{x_t \hat{y}_t}{\gamma e^{h_t}}\right),$$

and $\hat{y}_t = y_t - z_t'\alpha_t$, for $t = 1, \ldots, n$. The conditional posterior density is proportional to the kernel of the normal distribution whose mean and variance are $\hat{\beta}$ and $\hat{B}$, respectively. Then, we draw a sample as $\beta \mid \gamma, \alpha, h, y \sim N(\hat{\beta}, \hat{B})$.

2. Sample α

We consider how to sample α from its conditional posterior distribution. Regarding α as the state variable, the model given by equations (1) and (2) forms the linear Gaussian state space model. Given the parameters $(\beta, \Sigma, \gamma, h)$, a primitive way to sample α is to assess the conditional posterior density of α_t given $(\beta, \Sigma, \gamma, h, y, \alpha^t)$, where α^t is

4. Section A of the Appendix provides the functional form of the joint posterior distribution.

the α excluding α_t, i.e., $\alpha^t = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_n)$. This manner of sampling is often called a single-move sampler. The single-move sampler is quite simple, but inefficient in the sense that the autocorrelation of the MCMC sample often goes extremely high. For instance, after the α_t is sampled given α^t (including α_{t+1}), the α_{t+1} is sampled given α^{t+1} (including the α_t, which has been just drawn). The recursive chain depending on both sides of the sampled state variable yields an undesirable high autocorrelation. If the MCMC sample has a high autocorrelation, the convergence of the Markov chain is slow and an inference requires considerably many samples. To reduce the sample autocorrelation for α, we introduce the simulation smoother developed by de Jong and Shephard (1995) and Durbin and Koopman (2002). This enables us to sample α simultaneously from the conditional posterior distribution $\pi(\alpha \mid \beta, \Sigma, \gamma, h, y)$, which can reduce the autocorrelation of the MCMC sample.

Following de Jong and Shephard (1995), we show the algorithm of the simulation smoother on the state space model

$$
\begin{aligned}
y_t &= X_t\beta + Z_t\alpha_t + G_t u_t, \quad t = 1, \ldots, n, \\
\alpha_{t+1} &= T_t\alpha_t + H_t u_t, \quad t = 0, \ldots, n-1,
\end{aligned} \tag{4}
$$

where $\alpha_0 = 0$, $u_t \sim N(0, I)$, and $G_t H_t' = O$. The simulation smoother draws $\eta = (\eta_0, \ldots, \eta_t) \sim \pi(\eta \mid \omega, y)$, where $\eta_t = H_t u_t$, for $t = 0, \ldots, n$, and ω denotes all the parameters in the model. We initialize $a_1 = 0$, $P_1 = H_0 H_0'$, and recursively run the Kalman filter:

$$
\begin{aligned}
&e_t = y_t - X_t\beta - Z_t a_t, \quad D_t = Z_t P_t Z_t' + G_t G_t', \quad K_t = T_t P_t Z_t' D_t^{-1}, \\
&L_t = T_t - K_t Z_t, \quad a_{t+1} = T_t a_t + K_t e_t, \quad P_{t+1} = T_t P_t L_t' + H_t H_t',
\end{aligned}
$$

for $t = 1, \ldots, n$. Then, letting $r_n = U_n = 0$, and $\Lambda_t = H_t H_t'$, we run the simulation smoother:

$$
\begin{aligned}
&C_t = \Lambda_t - \Lambda_t U_t \Lambda_t, \quad \eta_t = \Lambda_t r_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, C_t), \quad V_t = \Lambda_t U_t L_t, \\
&r_{t-1} = Z_t' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} \varepsilon_t, \quad U_{t-1} = Z_t' D_t^{-1} Z_t + L_t' U_t L_t + V_t' C_t^{-1} V_t,
\end{aligned}
$$

for $t = n, n-1, \ldots, 1$. For the initial state, we draw $\eta_0 = \Lambda_0 r_0 + \varepsilon_0$, $\varepsilon_0 \sim N(0, C_0)$ with $C_0 = \Lambda_0 - \Lambda_0 U_0 \Lambda_0$. Once η is drawn, we can compute α_t using the state equation (4), replacing $H_t u_t$ by η_t.

In the case of the TVP regression model to sample α, the correspondence of the variables is as follows:

$$
\begin{aligned}
&X_t\beta = x_t'\beta, \quad Z_t = z_t', \quad G_t = (\sqrt{\gamma} e^{h_t/2}, 0_p'), \\
&T_t = I_p, \quad H_t = (0_p, \Sigma^{1/2}), \quad H_0 = (0_p, \Sigma_0^{1/2}),
\end{aligned}
$$

where 0_p is a $p \times 1$ zero vector, and I_p is a $p \times p$ identity matrix.

3. Sample Σ

We derive the conditional posterior density of Σ. If we specify the prior as $\Sigma \sim IW(\nu_0, \Omega_0^{-1})$, where IW denotes the inverse-Wishart distribution, we obtain the conditional posterior distribution for Σ as

$$\begin{aligned}\pi(\Sigma \,|\, \alpha) \propto |\Sigma|^{-\frac{\nu_0+p+1}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{tr}(\Omega_0\Sigma^{-1})\right\} \\ \times \prod_{t=1}^{n-1} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\alpha_{t+1}-\alpha_t)'\Sigma^{-1}(\alpha_{t+1}-\alpha_t)\right\} \\ \propto |\Sigma|^{-\frac{\hat{\nu}+p+1}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{tr}(\hat{\Omega}\Sigma^{-1})\right\}, \end{aligned} \tag{5}$$

where

$$\hat{\nu} = \nu_0 + n - 1, \quad \hat{\Omega} = \Omega_0 + \sum_{t=1}^{n-1}(\alpha_{t+1}-\alpha_t)(\alpha_{t+1}-\alpha_t)'.$$

Note that the posterior distribution for Σ depends on only α and (5) forms the kernel of the inverse-Wishart distribution. Then, we draw the sample as $\Sigma \,|\, \alpha \sim IW(\hat{\nu}, \hat{\Omega}^{-1})$.

4. Sample *h*

Regarding stochastic volatility h, the equations (1) and (3) form a nonlinear and non-Gaussian state space model. We need more technical methods for sampling h. A simple way of sampling h is to assess the conditional posterior distribution of h_t given $(h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$ and other parameters. This method is called a single-move sampler, similar to sampling α, and yields an undesirable high autocorrelation in MCMC sample.

There are mainly two efficient methods for sampling stochastic volatility developed in the literature. One way to sample stochastic volatility is the approach of Kim, Shephard, and Chib (1998), called the mixture sampler. The mixture sampler has been widely used in financial and macroeconomics literature (Cogley and Sargent [2005] and Primiceri [2005]). The other way is the multi-move sampler of Shephard and Pitt (1997), modified by Watanabe and Omori (2004). The idea of the former method is to approximate the nonlinear and non-Gaussian state space model by the normal mixture distribution, converting the original model to the linear Gaussian state space form. Though we draw samples from the posterior distribution based on the approximated model, its approximation error is small enough to implement the original model, and can be corrected by reweighting steps, as discussed by Kim, Shephard, and Chib (1998), and Omori *et al.* (2007). On the other hand, the latter algorithm approaches to the model by drawing samples from the exact posterior distribution of the original model. Both methods are more efficient to draw samples of stochastic volatility than a single-move sampler, while we use the latter one in this paper. The details of the multi-move sampler are illustrated in Section B of the Appendix.

5. Sample ϕ

We write the prior of ϕ as $\pi(\phi)$, and assume that $(\phi + 1)/2 \sim \text{Beta}(\alpha_{\phi_0}, \beta_{\phi_0})$. This beta distribution is chosen to satisfy the restriction $|\phi| < 1$. The conditional posterior distribution of ϕ is given by

$$
\begin{aligned}
&\pi(\phi \,|\, \sigma_\eta, h) \\
&\propto \pi(\phi) \times \sqrt{1 - \phi^2} \exp\left\{-\frac{(1 - \phi^2)h_1^2}{2\sigma_\eta^2}\right\} \times \exp\left\{-\frac{\sum_{t=1}^{n-1}(h_{t+1} - \phi h_t)^2}{2\sigma_\eta^2}\right\} \\
&\propto \pi(\phi)\sqrt{1 - \phi^2} \times \exp\left\{-\frac{\sum_{t=2}^{n-1} h_t^2}{2\sigma_\eta^2}\left(\phi - \frac{\sum_{t=1}^{n-1} h_t h_{t+1}}{\sum_{t=2}^{n-1} h_t^2}\right)^2\right\}.
\end{aligned}
$$

The conditional posterior density does not form any basic distribution from which we can easily sample. If the term $\pi(\phi)\sqrt{1 - \phi^2}$ is omitted, the rest of the term corresponds to a kernel of the normal distribution. In this case, we use the Metropolis-Hasting (MH) algorithm (e.g., Chib and Greenberg [1995]).

The idea of the MH algorithm is as follows. First, we draw samples (which we call *candidates*) from a certain distribution (*proposal distribution*) that is close to the conditional posterior distribution we want to sample from. We had better choose the proposal distribution whose random sample can be easily generated. Next, we accept the candidate as a new sample with a certain probability. When the candidate is rejected, we use the old (current) sample we have just drawn in the previous iteration as the new sample. Under certain conditions, the iterations of these steps produce the sample from the target conditional posterior distribution (see, e.g., Chib and Greenberg [1995]). There are many ways to choose the proposal density, which often depends on the target conditional posterior distribution.

Specifically, let $q(\theta^* \,|\, \theta^{(i)})$ denote the probability density function of the proposal given the current point $\theta^{(i)}$, and $\alpha(\theta_0, \theta^*)$ denote the acceptance rate from the current point θ_0 to the proposal θ^*. The MH algorithm is written as the following algorithm:

(1) Choose an arbitrary starting point $\theta^{(0)}$, and set $i = 0$.
(2) Generate a candidate θ^* from the proposal $q(\theta^* \,|\, \theta^{(i)})$.
(3) Accept θ^* with the probability $\alpha(\theta^{(i)}, \theta^*)$, and set $\theta^{(i+1)} = \theta^*$. Otherwise, set $\theta^{(i+1)} = \theta^{(i)}$.
(4) Set $i = i + 1$, and go to (2).

The acceptance rate is given by

$$
\alpha(\theta_0, \theta^*) = \min\left\{1, \frac{\pi(\theta^* \,|\, y)q(\theta_0 \,|\, \theta^*)}{\pi(\theta_0 \,|\, y)q(\theta^* \,|\, \theta_0)}\right\},
$$

where $\pi(\theta \,|\, y)$ denotes the target posterior distribution.

To sample ϕ in our model, we first draw a candidate as $\phi^* \sim TN_{[-1,1]}(\mu_\phi, \sigma_\phi^2)$, where *TN* refers to the truncated normal distribution on the domain $-1 < \phi < 1$, and

$$\mu_\phi = \frac{\sum_{t=1}^{n-1} h_t h_{t+1}}{\sum_{t=2}^{n-1} h_t^2}, \quad \sigma_\phi^2 = \frac{\sigma_\eta^2}{\sum_{t=2}^{n-1} h_t^2}.$$

This proposal density is the one excluding the term $\pi(\phi)\sqrt{1-\phi^2}$ from the conditional posterior distribution, considered to be close to our target conditional posterior distribution and truncated for the same domain of the target. Next, we calculate the probability for acceptance. Let $q(\phi)$ denote the probability density function of the proposal and ϕ_0 denote the old sample (current point) drawn in the previous iteration. The acceptance rate for the candidate ϕ^* from the current point ϕ_0, denoted by $\alpha(\phi_0, \phi^*)$, is given by

$$\alpha(\phi_0, \phi^*) = \min\left\{1, \frac{\pi(\phi^* \mid \sigma_\eta, h) q(\phi_0)}{\pi(\phi_0 \mid \sigma_\eta, h) q(\phi^*)}\right\} = \min\left\{1, \frac{\pi(\phi^*)\sqrt{1-\phi^{*2}}}{\pi(\phi_0)\sqrt{1-\phi_0^2}}\right\}.$$

The acceptance rate is the ratio of the terms omitted from the conditional posterior distribution. The acceptance step can be implemented by drawing a uniform random number $u \sim U(0,1)$ to accept the candidate ϕ^* when $u < \alpha(\phi_0, \phi^*)$.

6. Sample σ_η

We assume the prior of σ_η as $\sigma_\eta^2 \sim IG(v_0/2, V_0/2)$, where *IG* refers to the inverse gamma distribution. The conditional posterior distribution for σ_η is obtained as

$$\begin{aligned}\pi(\sigma_\eta \mid \phi, h) &\propto \sigma_\eta^{-(\frac{v_0}{2}+1)} \exp\left(-\frac{V_0}{2\sigma_\eta}\right) \\ &\quad \times \frac{1}{\sigma_\eta} \exp\left\{-\frac{(1-\phi^2)h_1^2}{2\sigma_\eta^2}\right\} \times \prod_{t=1}^{n-1} \frac{1}{\sigma_\eta} \exp\left\{-\frac{(h_{t+1}-\phi h_t)^2}{2\sigma_\eta^2}\right\} \\ &\propto \sigma_\eta^{-(\frac{v_0+n}{2}+1)} \exp\left\{-\frac{V_0 + (1-\phi^2)h_1^2 + \sum_{t=1}^{n-1}(h_{t+1}-\phi h_t)^2}{2\sigma_\eta}\right\}.\end{aligned}$$

The conditional posterior distribution forms the kernel of the inverse gamma distribution. Thus, we draw samples as $\sigma_\eta^2 \mid \phi, h \sim IG(\hat{v}/2, \hat{V}/2)$, where

$$\hat{v} = v_0 + n, \quad \hat{V} = V_0 + (1-\phi^2)h_1^2 + \sum_{t=1}^{n-1}(h_{t+1} - \phi h_t)^2.$$

7. Sample γ

Sampling γ can be implemented in the same way as sampling σ_η. We set the prior as $\gamma \sim IG(\gamma_0/2, W_0/2)$. Then, the conditional posterior distribution for γ is given by $\gamma \mid h \sim IG(\hat{\gamma}/2, \hat{W}/2)$, where

$$\hat{\gamma} = \gamma_0 + n, \quad \hat{W} = W_0 + \sum_{t=1}^{n}(y_t - x_t'\beta - z_t'\alpha_t)^2 / e^{h_t}.$$

III. Simulation Study

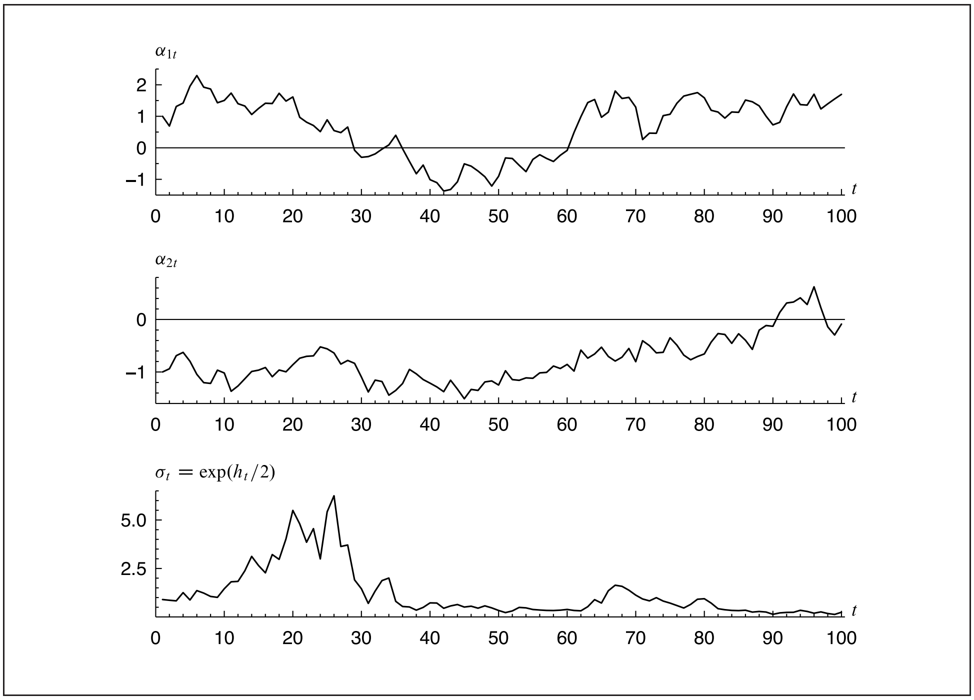
This section carries out simulation exercises of the TVP regression model to examine its estimation performance against the possibility of structural changes using simulated data, with emphasis on the role of stochastic volatility.

A. Setup

The performance of the proposed estimation method for the TVP regression model is illustrated using simulated data. In this simulation study, we investigate how the parameters are estimated, and how the assumption of stochastic volatility affects the estimates of other parameters.

Based on the TVP regression model of equations (1)–(3) with $n = 100$, $k = 2$, and $p = 2$, we generate $\{x_t\}_{t=1}^n$ and $\{z_t\}_{t=1}^n$ as $x_{it} \sim U(-0.5, 0.5)$, $z_{jt} \sim U(-0.5, 0.5)$ for $i, j = 1, 2$, where $x_t = (x_{1t}, x_{2t})'$, $z_t = (z_{1t}, z_{2t})'$, and $U(a, b)$ denotes the uniform distribution on the domain (a, b). Setting the true parameters as $\beta = (4, -3)'$, $\alpha_1 = (1, -1)'$, $\Sigma = \mathrm{diag}(0.1, 0.03)$, $\phi = 0.95$, $\sigma_\eta = 0.7$, and $\gamma = 0.1$, where $\mathrm{diag}(\cdot)$ refers to a diagonal matrix with the diagonal elements in the arguments, we generate α, h, and y recursively on the TVP regression model. The simulated state variables α and h are plotted in Figure 1. The volatility temporarily increases around $t = 20$.

Figure 1 Simulated State Variables α and h ($n = 100$)



B. Parameter Estimates

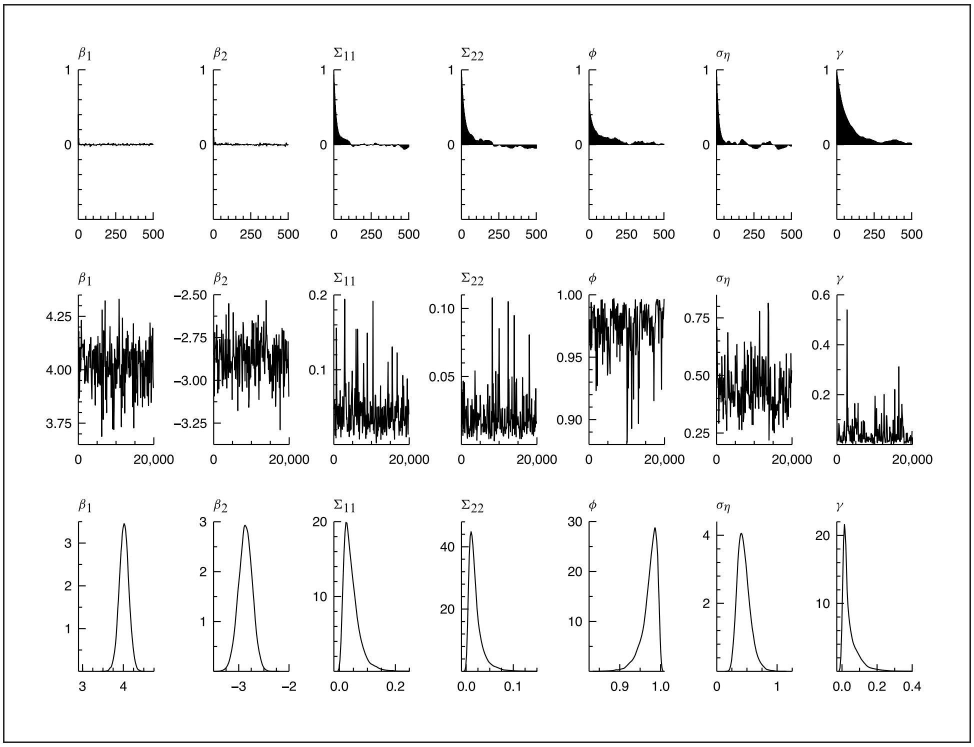
We estimate the TVP regression model using the simulated data by drawing $M = 20{,}000$ samples, after the initial 2,000 samples are discarded by assuming the following prior distributions:[5]

$$\beta \sim N(0, 10 \times I), \quad \Sigma \sim IW(4, 40 \times I), \quad \alpha_1 \sim N(0, 10 \times I),$$

$$\frac{\phi + 1}{2} \sim \text{Beta}(20, 1.5), \quad \sigma_\eta^2 \sim IG(2, 0.02), \quad \gamma \sim IG(2, 0.02).$$

Figure 2 shows the sample autocorrelation function, the sample paths, and the posterior densities for the selected parameters. After discarding the samples in the burn-in period (initial 2,000 samples), the sample paths look stable and the sample autocorrelations drop stably, indicating that our sampling method efficiently produces the samples with low autocorrelation.

Figure 2 Estimation Results of the TVP Regression Model (With Stochastic Volatility) for the Simulated Data



Note: Sample autocorrelations (top), sample paths (middle), and posterior densities (bottom).

5. The computational results are generated using Ox version 4.02 (Doornik [2006]). All the codes for the algorithms illustrated in this paper are available at http://sites.google.com/site/jnakajimaweb/program.

Table 1 Estimation Results of the TVP Regression Model for the Simulated Data with (1) Stochastic Volatility and (2) Constant Volatility

[1] TVP Regression Model with *Stochastic* Volatility

Parameter	True	Mean	Stdev.	95 percent interval	CD	Inefficiency
β_1	4.0	4.0155	0.1166	[3.7837, 4.2441]	0.833	2.46
β_2	−3.0	−2.8668	0.1371	[−3.1409, −2.6019]	0.909	4.37
Σ_{11}	0.1	0.0440	0.0303	[0.0096, 0.1221]	0.144	38.02
Σ_{22}	0.03	0.0201	0.0168	[0.0043, 0.0656]	0.217	57.05
ϕ	0.95	0.9735	0.0197	[0.9224, 0.9967]	0.895	52.39
σ_η	0.5	0.4508	0.1084	[0.2808, 0.7057]	0.506	33.55
γ	0.1	0.0445	0.0511	[0.0052, 0.1865]	0.908	116.44

[2] TVP Regression Model with *Constant* Volatility

Parameter	True	Mean	Stdev.	95 percent interval	CD	Inefficiency
β_1	4.0	4.2373	0.3118	[3.6256, 4.8447]	0.472	1.03
β_2	−3.0	−2.7760	0.3369	[−3.4188, −2.1054]	0.398	1.52
Σ_{11}	0.1	0.0173	0.0206	[0.0029, 0.0689]	0.533	68.50
Σ_{22}	0.03	0.0123	0.0133	[0.0025, 0.0444]	0.136	70.39
σ	—	0.9451	0.0688	[0.8215, 1.0922]	0.456	1.87

Note: The true model is stochastic volatility.

Table 1 gives the estimates for posterior means, standard deviations, the 95 percent credible intervals,[6] the convergence diagnostics (CD) of Geweke (1992), and inefficiency factors, which are computed using the MCMC sample.[7] In the estimated result, the null hypothesis of the convergence to the posterior distribution is not rejected for the parameters at the 5 percent significance level based on the CD statistics, and the inefficiency factors are quite low except for γ, which indicates an efficient sampling for the parameters and state variables. Even for γ, the inefficiency factor is about 100, which implies that we obtain about $M/100 = 200$ uncorrelated samples. It is considered to be sufficient for the posterior inference. In addition, the estimated posterior mean is

6. In Bayesian inference, we use "credible intervals" to describe the uncertainty of the parameters, instead of "confidence intervals" in the frequentist approach. In the MCMC analysis, we usually report the 2.5 percent and 97.5 percent quantiles of posterior draws, as taken here.

7. To check the convergence of the Markov chain, Geweke (1992) suggests the comparison between the first n_0 draws and the last n_1 draws, dropping out the middle draws. The CD statistics are computed by

$$\mathrm{CD} = (\bar{x}_0 - \bar{x}_1)/\sqrt{\hat{\sigma}_0^2/n_0 + \hat{\sigma}_1^2/n_1},$$

where $\bar{x}_j = (1/n_j)\sum_{i=m_j}^{m_j+n_j-1} x^{(i)}$, $x^{(i)}$ is the i-th draw, and $\sqrt{\hat{\sigma}_j^2/n_j}$ is the standard error of $\bar{x}_j$ respectively for $j = 0, 1$. If the sequence of the MCMC sampling is stationary, it converges in distribution to a standard normal. We set $m_0 = 1$, $n_0 = 1{,}000$, $m_1 = 5{,}001$, and $n_1 = 5{,}000$. The $\hat{\sigma}_j^2$ is computed using a Parzen window with bandwidth, $B_m = 500$. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{B_m} \rho_s$, where ρ_s is the sample autocorrelation at lag s, which is computed to measure how well the MCMC chain mixes (see, e.g., Chib [2001]). It is the ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. The inverse of the inefficiency factor is also known as relative numerical efficiency (Geweke [1992]). When the inefficiency factor is equal to m, we need to draw the MCMC sample m times as many as the uncorrelated sample.

close to the true value of the parameter, and the 95 percent credible intervals include it for each parameter listed in Table 1 [1].

C. The Role of Stochastic Volatility

To assess the function of stochastic volatility in the TVP regression model, we estimate the TVP regression model with constant volatility for the same simulated data. Because the true specification is stochastic volatility, we investigate how the estimation result changes with the misspecification. As mentioned in Section II.A, constant volatility is specified by $\sigma_t^2 = \sigma^2$, for $t = 1, \ldots, n$. If we assume the prior as $\sigma^2 \sim IG(s_0/2, S_0/2)$, then the conditional posterior distribution of σ is given by $\sigma^2 \mid \beta, \alpha, y \sim IG(\hat{s}/2, \hat{S}/2)$, where $\hat{s} = s_0 + n$, and $\hat{S} = S_0 + \sum_{t=1}^{n}(y_t - x_t'\beta - z_t'\alpha_t)^2$. For the MCMC algorithm for the TVP regression model, Steps 4–7 are replaced by the step of sampling σ for constant volatility.

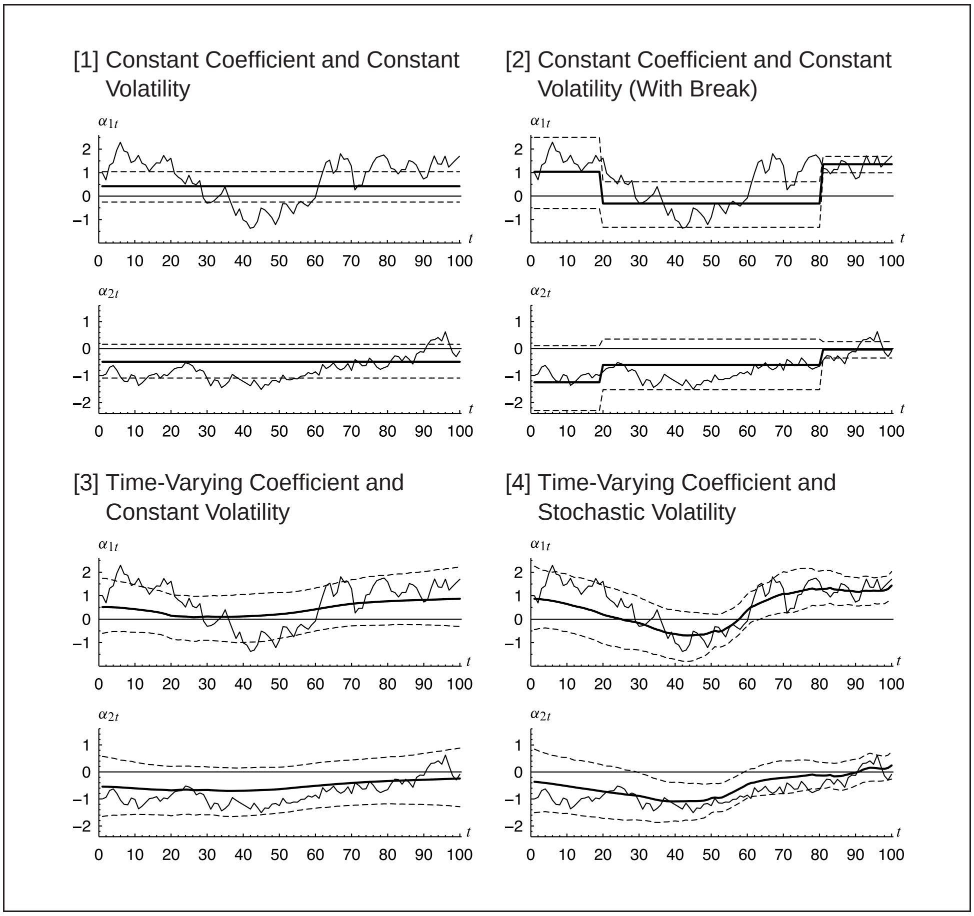
In the simulation study, the prior $\sigma^2 \sim IG(2, 0.02)$ is additionally assumed, and the estimation procedure is the same as the TVP regression model with stochastic volatility discussed above. Table 1 [2] reports the estimation results of the TVP regression model with constant volatility for the simulated data. The standard deviations of (β_1, β_2) are evidently wider than the stochastic volatility model, and the posterior means are slightly apart from the true value. The posterior means of $(\Sigma_{11}, \Sigma_{22})$ are estimated lower than the stochastic volatility model.

We check how the time-varying coefficients are estimated. In addition to the above two models, the constant coefficient and constant volatility model is estimated. The posterior estimates of α are plotted in Figure 3. Figure 3 [1] clearly shows that the constant coefficient model is unable to capture the time variation of the coefficients, and the posterior mean is estimated around the averaged level of time-varying coefficients over time. Figure 3 [2] plots the estimates based on the same time-invariant model with structural breaks. To detect a possible break, the CUSUM of squares test proposed by Brown, Durbin, and Evans (1975) is applied to divide the sample period into three parts ($t = 1$–19, 20–81, 82–100). Then, the constant coefficient and constant volatility model is estimated for each subsample period.[8] In the first and second subsample periods, the posterior 95 percent credible intervals are wide, primarily due to the high volatility of the disturbance. In the third subsample period, the posterior means seem to follow the average level of the time-varying coefficient over each subsample period, and the 95 percent credible intervals are narrower. However, the true states are not traced well.

Figure 3 [3] exhibits the estimation results for the TVP regression with constant volatility. The posterior means seem to follow the true states of the time-varying coefficients to some extent. However, for α_{1t}, some true values do not drop in the 95 percent credible intervals. On the other hand, for α_{2t}, the intervals are too wide to capture the movement of the true value. The constant volatility model neglects the behavior of the volatility change and lacks the accuracy of estimates for α_{it}. The estimates of the TVP regression with stochastic volatility, which is the true model, are plotted

8. Modeling structural changes is one of the central issues of recent econometrics (see, e.g., Perron [2006]). As well as the time-varying coefficients and stochastic volatility, structural changes can assess possible changes in the underlying data generation process. Whether or not a *true* model has a structural break or time-varying parameters such as the one in this paper, both models are intended to capture it by approximating its behavior in each case.

Figure 3 Estimation Results of α on the TVP Regression Model for the Simulated Data



Note: True value (solid line), posterior mean (bold line), and 95 percent credible intervals (dashed line). The true model is the time-varying coefficient and stochastic volatility [4].

in Figure 3 [4]. The posterior means trace the movement of the true values and the 95 percent credible intervals tend to be narrower overall than the constant volatility model, and almost include the true values.

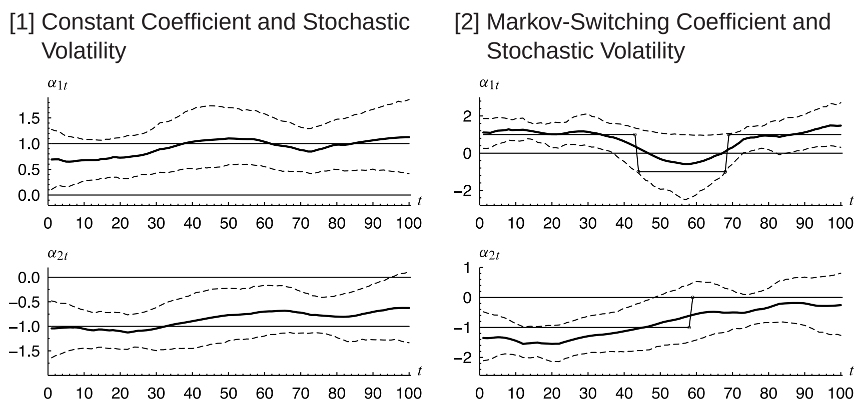
The simulation analysis here refers to a profound issue of identifying the source of the shock. Focusing on the third case, the estimated constant variance (σ) of the disturbance is smaller in the first-half period and larger in the second half than the true state of stochastic volatility, because the constant variance captures the average level of volatility. For the first-half period, the 95 percent credible intervals are almost as wide as the stochastic volatility model, although the posterior mean is less accurate with respect to the distance between the estimated posterior means and true values, because the shock to the disturbance is estimated to be smaller than the true state and the rest of the shock is drawn up to the drifting α_{it} in a misspecified way. On the other hand, for the second-half period, the posterior mean of the constant volatility model is relatively accurate compared to the first-half period, but the 95 percent credible intervals are wider than the stochastic volatility model, because the constant volatility is over-estimated and the vagueness remains in the drifting α_{it}.

D. Other Models

In addition, other interesting simulations in which the true model is not the TVP regression form with time-varying coefficient and stochastic volatility are examined. First, data are simulated from the TVP regression model with constant coefficient and stochastic volatility. The true values are the same as the previous simulation study, except $\alpha_{1t} = 1$ and $\alpha_{2t} = -1$, for all $t = 1, \ldots, n$. The TVP regression model with time-varying coefficient and stochastic volatility is estimated to examine how the time-varying coefficient follows the time-invariant true state. The estimation results of $(\alpha_{1t}, \alpha_{2t})$ are shown in Figure 4 [1]. Though the estimates of the posterior means are not perfectly time-invariant, they are moving near the true states, and the 95 percent credible intervals include the true value throughout the sample periods.

Second, data are simulated from the TVP regression model with stochastic volatility, but with the time-varying coefficients $(\alpha_{1t}, \alpha_{2t})$ modeled to have the Markov-switching structural change. Much of the literature considers the Markov-switching type of time-varying parameters in macroeconomic issues. We assume that α_{1t} and α_{2t} have two regimes $(\alpha_{1t}^{(0)}, \alpha_{1t}^{(1)}) = (1, -1)$ and $(\alpha_{2t}^{(0)}, \alpha_{2t}^{(1)}) = (-1, 0)$, respectively. The coefficients $(\alpha_{1t}, \alpha_{2t})$ switch independently with the transition probabilities $p(\alpha_{it} = \alpha_{it}^{(j)} \mid \alpha_{i,t-1} = \alpha_{i,t-1}^{(j)}) = 0.98$, for $i = 1, 2$ and $j = 0, 1$. The TVP regression model with time-varying coefficient (of the original form) and stochastic volatility is estimated to examine how the time-varying coefficient follows the Markov-switching structural

Figure 4 Estimation Results of α on the TVP Regression Model for the Simulated Data



Note: True value (solid line), posterior mean (bold line), and 95 percent credible intervals (dashed line). The true models are (1) constant coefficient and stochastic volatility, and (2) Markov-switching coefficient and stochastic volatility. The TVP regression model with time-varying coefficient and stochastic volatility is fitted.

change. Figure 4 [2] plots the estimation results of the coefficients. The true states of α_{1t} and α_{2t} have two breaks and one break, respectively. For both coefficients, the 95 percent credible intervals include the true values. Around the structural breaks, the posterior means of the coefficients follow the true states to some extent, although their movements would not be so responsive, especially for α_{2t}. The degree of adjustment to the structural change depends on the size of the volatility of the disturbance in regression. The posterior estimates tend to smooth the true states of the coefficients.

The simulations in this section are just one case of generated data for each setting. However, the estimation results show the flexibility and the applicability of the TVP regression models, which would help us to understand the importance of the time-varying parameters in the regression models.

IV. Time-Varying Parameter VAR with Stochastic Volatility

This section extends the estimation algorithm for a univariate TVP estimation model to a multivariate TVP-VAR model.

A. Model

To introduce the TVP-VAR model, we begin with a basic structural VAR model defined as

$$A\boldsymbol{y}_t = F_1\boldsymbol{y}_{t-1} + \cdots + F_s\boldsymbol{y}_{t-s} + \boldsymbol{u}_t, \quad t = s+1, \ldots, n, \tag{6}$$

where $\boldsymbol{y}_t$ is the $k \times 1$ vector of observed variables, and $A, F_1, \ldots, F_s$ are $k \times k$ matrices of coefficients. The disturbance $\boldsymbol{u}_t$ is a $k \times 1$ structural shock and, we assume that $\boldsymbol{u}_t \sim N(0, \Sigma\Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_k \end{pmatrix}.$$

We specify the simultaneous relations of the structural shock by recursive identification, assuming that A is lower-triangular,

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{k1} & \cdots & a_{k,k-1} & 1 \end{pmatrix}.$$

We rewrite model (6) as the following reduced form VAR model:

$$\boldsymbol{y}_t = B_1\boldsymbol{y}_{t-1} + \cdots + B_s\boldsymbol{y}_{t-s} + A^{-1}\Sigma\varepsilon_t, \quad \varepsilon_t \sim N(0, I_k),$$

where $B_i = A^{-1}F_i$, for $i = 1, \ldots, s$. Stacking the elements in the rows of the B_i's to form $\boldsymbol{\beta}$ ($k^2 s \times 1$ vector), and defining $X_t = I_k \otimes (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-s})$, where $\otimes$ denotes the Kronecker product, the model can be written as

$$\boldsymbol{y}_t = X_t \boldsymbol{\beta} + A^{-1} \Sigma \varepsilon_t. \tag{7}$$

Now, all parameters in equation (7) are time-invariant. We extend it to the TVP-VAR model by allowing the parameters to change over time.

Consider the TVP-VAR model stochastic volatility specified by

$$\boldsymbol{y}_t = X_t \boldsymbol{\beta}_t + A_t^{-1} \Sigma_t \varepsilon_t, \quad t = s+1, \ldots, n, \tag{8}$$

where the coefficients $\boldsymbol{\beta}_t$, and the parameters A_t and Σ_t are all time varying.[9] There are many ways to model the process for these time-varying parameters.[10] Following Primiceri (2005), let $\boldsymbol{a}_t = (a_{21}, a_{31}, a_{32}, a_{41}, \ldots, a_{k,k-1})'$ be a stacked vector of the lower-triangular elements in A_t and $\boldsymbol{h}_t = (h_{1t}, \ldots, h_{kt})'$ with $h_{jt} = \log \sigma_{jt}^2$, for $j = 1, \ldots, k$, $t = s+1, \ldots, n$. We assume that the parameters in (8) follow a random walk process as follows:

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + u_{\beta t}, \quad \boldsymbol{a}_{t+1} = \boldsymbol{a}_t + u_{at}, \quad \boldsymbol{h}_{t+1} = \boldsymbol{h}_t + u_{ht},$$

$$\begin{pmatrix} \varepsilon_t \\ u_{\beta t} \\ u_{at} \\ u_{ht} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & O & O & O \\ O & \Sigma_\beta & O & O \\ O & O & \Sigma_a & O \\ O & O & O & \Sigma_h \end{pmatrix}\right),$$

for $t = s+1, \ldots, n$, where $\boldsymbol{\beta}_{s+1} \sim N(\mu_{\beta_0}, \Sigma_{\beta_0})$, $\boldsymbol{a}_{s+1} \sim N(\mu_{a_0}, \Sigma_{a_0})$ and $\boldsymbol{h}_{s+1} \sim N(\mu_{h_0}, \Sigma_{h_0})$.

Several remarks are required for the specification of the TVP-VAR model. First, the assumption of a lower-triangular matrix for A_t is recursive identification for the VAR system. This specification is simple and widely used, although an estimation of structural models may require a more complicated identification to extract implications for the economic structure, as pointed out by Christiano, Eichenbaum, and Evans (1999) and other studies. In this paper, the estimation algorithm is explained in the model with recursive identification for simplicity, although the estimation procedure is applicable for the model with non-recursive identification by a slight modification of the variable in the MCMC algorithm.

Second, the parameters are not assumed to follow a stationary process such as AR(1), but the random walk process. As mentioned before, because the TVP-VAR model has a number of parameters to estimate, we had better decrease the number of parameters by assuming the random walk process for the innovation of parameters. Most of studies that use the TVP-VAR model assume the random walk process for

9. Time-varying intercepts are incorporated in some literature on the TVP-VAR models. This case requires only the modification of defining $X_t := I_k \otimes (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-s})$.

10. Hereafter, we use the "TVP-VAR model" to indicate that model with stochastic volatility for simplicity.

parameters. Note that the extension of the estimation algorithm to the case of stationary process is straightforward.

Third, the variance and covariance structure for the innovations of the time-varying parameters are governed by the parameters, Σ_β, Σ_a, and Σ_h. Most of the articles assume that Σ_a is a diagonal matrix. In this paper, we further assume that Σ_h is also a diagonal matrix for simplicity. The experience of several estimations indicates that this diagonal assumption for Σ_h is not sensitive for the results, compared to the non-diagonal assumption.

Fourth, when the TVP-VAR model is implemented in the Bayesian inference, the priors should be carefully chosen because the TVP-VAR model has many state variables and their process is modeled as a non-stationary random walk process. The TVP-VAR model is so flexible that the state variables can capture both gradual and sudden changes in the underlying economic structure. On the other hand, allowing time variation in every parameter in the VAR model may cause an over-identification problem. As mentioned by Primiceri (2005), the tight prior for the covariance matrix of the disturbance in the random walk process avoids the implausible behaviors of the time-varying parameters. The time-varying coefficient ($\boldsymbol{\beta} = (\boldsymbol{\beta}_{s+1}, \ldots, \boldsymbol{\beta}_n)$) requires a tighter prior than the simultaneous relations ($\boldsymbol{a} = (\boldsymbol{a}_{s+1}, \ldots, \boldsymbol{a}_n)$) and the volatility ($\boldsymbol{h} = (\boldsymbol{h}_{s+1}, \ldots, \boldsymbol{h}_n)$) of the structural shock for the variance of the disturbance in their time-varying process. The structural shock we consider in the model unexpectedly hits the economic system, and its size would fluctuate more widely over time than the possible change in the autoregressive system of the economic variables specified by VAR coefficients. In most of the related literature, a tighter prior is set for Σ_β and a rather diffuse prior for Σ_a and Σ_h. A prior sensitivity analysis would be necessary to check the robustness of the empirical result with respect to the prior tightness.

Finally, the prior of the initial state of the time-varying parameters is specified. When the time-series model is a stationary process, we often assume the initial state following a stationary distribution of the process (for instance, $h_1 \sim N(0, \sigma_\eta^2/(1-\phi^2))$ in the TVP regression model). However, our time-varying parameters are random walks; thus, we specify a certain prior for $\boldsymbol{\beta}_{s+1}$, $\boldsymbol{a}_{s+1}$, and $\boldsymbol{h}_{s+1}$. We have two ways to set the prior. First, following Primiceri (2005), we set a prior of normal distribution whose mean and variance are chosen based on the estimates of a constant parameter VAR model computed using the pre-sample period. It is reasonable to use the economic structure estimated from the pre-sample period up to the initial period of the main sample data. Second, we can set a reasonably flat prior for the initial state from the standpoint that we have no information about the initial state *a priori*.[11]

B. Estimation Methodology

The estimation procedure for the TVP-VAR model is illustrated by extending several parts of the algorithm for the TVP regression model. Let $\boldsymbol{y} = \{\boldsymbol{y}_t\}_{t=1}^n$, and $\omega = (\Sigma_\beta, \Sigma_a, \Sigma_h)$. We set the prior probability density as $\pi(\omega)$ for ω. Given the data $\boldsymbol{y}$, we draw samples from the posterior distribution, $\pi(\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{h}, \omega \mid \boldsymbol{y})$, by the following MCMC algorithm:

11. Koop and Korobilis (2010) provide a comprehensive discussion on the methodology for the TVP-VAR model, including the issues about the prior specifications.

(1) Initialize $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{h}$, and ω.
(2) Sample $\boldsymbol{\beta} \mid \boldsymbol{a}, \boldsymbol{h}, \Sigma_\beta, \boldsymbol{y}$.
(3) Sample $\Sigma_\beta \mid \boldsymbol{\beta}$.
(4) Sample $\boldsymbol{a} \mid \boldsymbol{\beta}, \boldsymbol{h}, \Sigma_a, \boldsymbol{y}$.
(5) Sample $\Sigma_a \mid \boldsymbol{a}$.
(6) Sample $\boldsymbol{h} \mid \boldsymbol{\beta}, \boldsymbol{a}, \Sigma_h, \boldsymbol{y}$.
(7) Sample $\Sigma_h \mid \boldsymbol{h}$.
(8) Go to (2).

The details of the procedure are illustrated as follows.

1. Sample $\boldsymbol{\beta}$

To sample $\boldsymbol{\beta}$ from the conditional posterior distribution, the state space model with respect to $\boldsymbol{\beta}_t$ as the state variable is written as

$$\boldsymbol{y}_t = X_t \boldsymbol{\beta}_t + A_t^{-1} \Sigma_t \varepsilon_t, \quad t = s+1, \dots, n,$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + u_{\beta t}, \quad t = s, \dots, n-1,$$

where $\boldsymbol{\beta}_s = \mu_{\beta_0}$, and $u_{\beta s} \sim N(0, \Sigma_{\beta_0})$. We run the simulation smoother with the correspondence of the variables to equation (4) as follows:

$$X_t \beta = 0_k, \quad Z_t = X_t, \quad G_t = (A_t^{-1} \Sigma_t, O_{k_\beta}),$$

$$T_t = I_{k_\beta}, \quad H_t = (O_k, \Sigma_\beta^{1/2}), \quad H_0 = (O_k, \Sigma_{\beta_0}^{1/2}),$$

where k_β is the number of rows of $\boldsymbol{\beta}_t$.

2. Sample $\boldsymbol{a}$

To sample $\boldsymbol{a}$ from the conditional posterior distribution, the expression of the state space form with respect to $\boldsymbol{a}_t$ is a key to implementing the simulation smoother. Specifically,

$$\hat{\boldsymbol{y}}_t = \hat{X}_t \boldsymbol{a}_t + \Sigma_t \varepsilon_t, \quad t = s+1, \dots, n,$$

$$\boldsymbol{a}_{t+1} = \boldsymbol{a}_t + u_{at}, \quad t = s, \dots, n-1,$$

where $\boldsymbol{a}_s = \mu_{a_0}$, $u_{as} \sim N(0, \Sigma_{a_0})$, $\hat{\boldsymbol{y}}_t = \boldsymbol{y}_t - X_t \boldsymbol{\beta}_t$, and

$$\hat{X}_t = \begin{pmatrix} 0 & \cdots & & & & & & 0 \\ -\hat{\boldsymbol{y}}_{1t} & 0 & 0 & \cdots & & & & \vdots \\ 0 & -\hat{\boldsymbol{y}}_{1t} & -\hat{\boldsymbol{y}}_{2t} & 0 & \cdots & & & \\ 0 & 0 & 0 & -\hat{\boldsymbol{y}}_{1t} & \cdots & & & \\ \vdots & & & & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & & & 0 & -\hat{\boldsymbol{y}}_{1t} & \cdots & -\hat{\boldsymbol{y}}_{k-1,t} \end{pmatrix},$$

for $t = s+1, \ldots, n$. We run the simulation smoother to sample $\boldsymbol{a}$ with the correspondences:

$$X_t\beta = 0_k, \quad Z_t = \hat{X}_t, \quad G_t = (\Sigma_t, O_{k_a}),$$

$$T_t = I_{k_a}, \quad H_t = (O_k, \Sigma_a^{1/2}), \quad H_0 = (O_k, \Sigma_{a_0}^{1/2}),$$

where k_a is the number of rows of $\boldsymbol{a}_t$.

3. Sample $\boldsymbol{h}$

As for stochastic volatility $\boldsymbol{h}$, we make the inference for $\{h_{jt}\}_{t=s+1}^n$ separately for j $(= 1, \ldots, k)$, because we assume Σ_h and Σ_{h_0} are diagonal matrices. Let y_{it}^* denote the i-th element of $A_t\hat{\boldsymbol{y}}_t$. Then, we can write:

$$y_{it}^* = \exp(h_{it}/2)\varepsilon_{it}, \quad t = s+1, \ldots, n,$$

$$h_{i,t+1} = h_{it} + \eta_{it}, \quad t = s, \ldots, n-1,$$

$$\begin{pmatrix} \varepsilon_{it} \\ \eta_{it} \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & v_i^2 \end{pmatrix}\right),$$

where $\eta_{is} \sim N(0, v_{i0}^2)$, and v_i^2 and v_{i0}^2 are the i-th diagonal elements of Σ_h and Σ_{h_0}, respectively, and η_{it} is the i-th element of u_{ht}. We sample $(h_{i,s+1}, \ldots, h_{in})$ using the multi-move sampler developed in Section B of the Appendix.

4. Sample $\boldsymbol{\omega}$

Sampling Σ_β from its conditional posterior distribution is the same way to sample Σ in the TVP regression model. Sampling the diagonal elements of Σ_a and Σ_h is also the same way to sample σ_η in the TVP regression model. When the prior is the inverse gamma distribution, so is the conditional posterior distribution.

C. Literature

The econometric analysis using the VAR model was originally developed by Sims (1980). Numerous studies have been investigated in this context, and it has become a standard econometric tool in macroeconomics literature (see, e.g., Leeper, Sims, and Zha [1996] and Christiano, Eichenbaum, and Evans [1999] for a broader survey of the literature).

Since the late 1990s, the time-varying components have been incorporated into the VAR analysis. A salient analysis using the VAR model with time-varying coefficients was developed by Cogley and Sargent (2001). They estimate a three-variable VAR model (inflation, unemployment, and nominal short-term interest rates), focusing on the persistence of inflation and the forecasts of inflation and unemployment for postwar U.S. data. The dynamics of policy activism are also discussed based on their time-varying VAR model. Among the discussions of their results, Sims (2001) and Stock (2001) questioned the assumption of the constant variance ($\boldsymbol{a}$ and $\boldsymbol{h}$ in our notation) for the VAR's structural shock, and were concerned that the results for the drifting coefficients of Cogley and Sargent (2001) might be exaggerated due to the neglect of a possible variation of the variance.[12] Replying to them, Cogley and Sargent (2005) incorporated stochastic volatility into the VAR model with time-varying coefficients.[13]

Primiceri (2005) proposes the TVP-VAR model that allows all parameters ($\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{h}$) to vary over time, and estimate a three-variable VAR model (the same variables as Cogley and Sargent [2001]) for the U.S. data.[14] The empirical results reveal that the responses of the policy interest rates to inflation and unemployment exhibit a trend toward more aggressive behavior in recent decades, and this has a negligible effect on the rest of the economy.

After Primiceri (2005)'s introduction of the TVP-VAR model, several papers have analyzed the time-varying structure of the macroeconomy in specific ways. Benati and Mumtaz (2005) estimate the TVP-VAR model for the U.K. data by imposing sign restrictions on the impulse responses to assess the source of the "Great Stability" in the United Kingdom as well as uncertainty for inflation forecasting (see also Benati [2008]). Baumeister, Durinck, and Peersman (2008) estimate the TVP-VAR model for the euro area data to assess the effects of excess liquidity shocks on macroeconomic variables. D'Agostino, Gambetti, and Giannone (2010) examine the forecasting performance of the TVP-VAR model over other standard VAR models. Nakajima, Kasuya, and Watanabe (2009) and Nakajima, Shiratsuka, and Teranishi (2010) estimate the TVP-VAR model for the Japanese macroeconomic data. An increasing number of studies have examined the TVP-VAR models to provide empirical evidence of the dynamic structure of the economy (see e.g., Benati and Surico [2008], Mumtaz and Surico [2009], Baumeister and Benati [2010], and Clark and Terry [2010]). Given such previous literature, we will show an empirical application of the TVP-VAR model to Japanese data, with emphasis on the role of stochastic volatility in the estimation.

V. Empirical Results for the Japanese Economy

As mentioned above, this section applies the TVP-VAR model, developed so far, to Japanese macroeconomic variables, with emphasis on the role of stochastic volatility in the estimation.[15]

A. Data and Settings

A three-variable TVP-VAR model is estimated for quarterly data from the period 1977/Q1 to 2007/Q4, thereby examining the time-varying nature of macroeconomic dynamics over the three decades of the sample period. To this end, two sets of variables

12. Cogley and Sargent (2005) state, "If the world were characterized by constant θ [coefficients of the VAR] and drifting R [variance of the VAR], and we fit an approximating model with constant R and drifting θ, then it seems likely that our estimates of θ would drift to compensate for misspecification of R, thus exaggerating the time variation in θ."
13. Uhlig (1997) originally developed the VAR model with stochastic volatility.
14. In Cogley and Sargent (2005), it is assumed that the simultaneous relations, a, of the structural shock remain time-invariant.
15. Similar studies for Japanese macroeconomic data are analyzed by Nakajima, Kasuya, and Watanabe (2009) and Nakajima, Shiratsuka, and Teranishi (2010). See the previous section for literature on the empirical studies of the TVP-VAR models using other countries' data.

Table 2 Estimation Results of Selected Parameters in the TVP-VAR Model for the Variable Set of (p, x, b)

Parameter	Mean	Stdev.	95 percent interval	CD	Inefficiency
$(\Sigma_\beta)_1$	0.0531	0.0123	[0.0341, 0.0824]	0.165	3.97
$(\Sigma_\beta)_2$	0.0567	0.0129	[0.0361, 0.0866]	0.253	10.25
$(\Sigma_a)_1$	0.5575	0.4392	[0.1487, 1.7505]	0.511	45.58
$(\Sigma_a)_2$	0.6148	0.5439	[0.1633, 1.9004]	0.383	60.34
$(\Sigma_h)_1$	0.4453	0.2452	[0.1302, 1.0847]	0.382	33.64
$(\Sigma_h)_2$	0.1300	0.0808	[0.0304, 0.3377]	0.526	43.37

Note: The estimates of Σ_β and Σ_a are multiplied by 100.

are examined: (p, x, b) and (p, x, i), where p is the inflation rate; x is the output; b is the medium-term interest rates; and i is the short-term interest rates.[16]

The number of the VAR lags is four,[17] and we assume that Σ_β is a diagonal matrix in this study for simplicity. Some experiences indicate that this assumption is not sensitive for the results, compared to the non-diagonal assumption. The following priors are assumed for the i-th diagonals of the covariance matrices:

$$(\Sigma_\beta)_i^{-2} \sim \text{Gamma}(40, 0.02), \quad (\Sigma_a)_i^{-2} \sim \text{Gamma}(4, 0.02),$$

$$(\Sigma_h)_i^{-2} \sim \text{Gamma}(4, 0.02).$$

For the initial state of the time-varying parameter, rather flat priors are set; $\mu_{\beta_0} = \mu_{a_0} = \mu_{h_0} = 0$, and $\Sigma_{\beta_0} = \Sigma_{a_0} = \Sigma_{h_0} = 10 \times I$. To compute the posterior estimates, we draw $M = 10{,}000$ samples after the initial 1,000 samples are discarded. Table 2 and Figure 5 report the estimation results for selected parameters of the TVP-VAR model for the variable set (p, x, b). The results show that the MCMC algorithm produces posterior draws efficiently.

B. Empirical Results

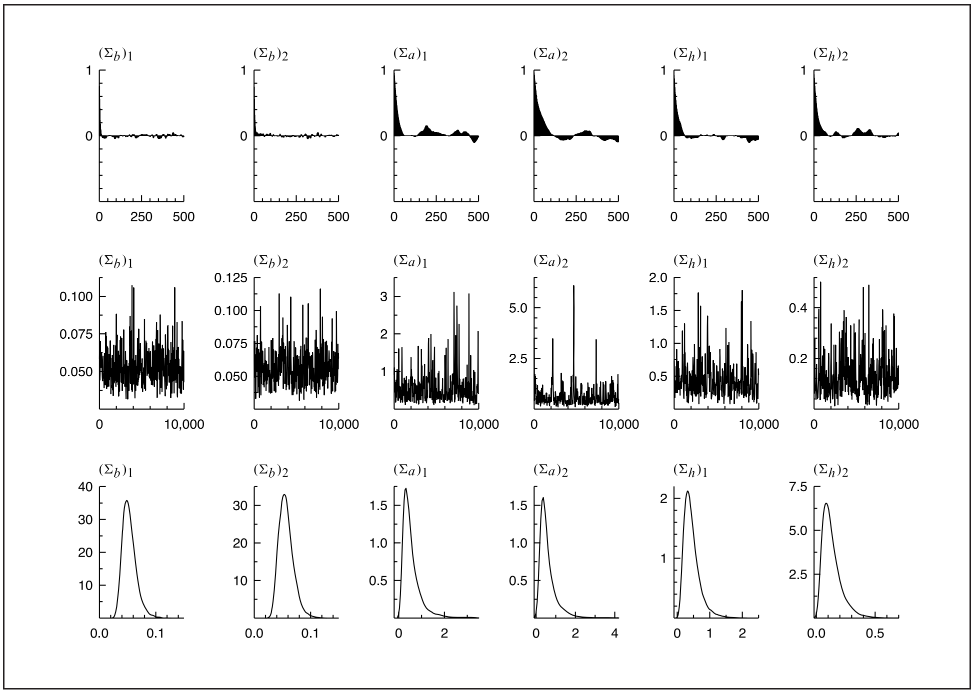
1. Estimation results for the first set of variables: (p, x, b)

First, the variable set of (p, x, b) is estimated. Figure 6 plots the posterior estimates of stochastic volatility and the simultaneous relation. The time-series plots consist of the posterior draws on each date. As for the simultaneous relation, which is specified by the lower triangular matrix A_t, the posterior estimates of the free elements in A_t^{-1}, denoted

16. The inflation rate is taken from the consumer price index (CPI, general excluding fresh food, log-difference, the effects of the increase in the consumption tax removed, and seasonally adjusted). The output gap is a series of deviations of GDP from its potential level, calculated by the BOJ. The medium-term bond interest rates are a yield of five-year Japanese government bonds. Up to 1988/Q1, the five-year interest-bearing bank debenture, and from 1988/Q2 a series of the generic index of Bloomberg, is used. The short-term interest rates are the overnight call rate. Except for the output gap, the monthly data are arranged to a quarterly base by monthly average. For both the interest rates, the (log-scale) difference of the original series from the trend of the HP filter, that is, an interest rate gap from the trend, is computed for the variable of the estimation.

17. The marginal likelihood is estimated for different lag lengths (up to six) and the number of lags is determined based on the highest marginal likelihood (see Nakajima, Kasuya, and Watanabe [2009] for the computation of the marginal likelihood).

Figure 5 Estimation Results of Selected Parameters in the TVP-VAR Model for the Variable Set of (p, x, b)



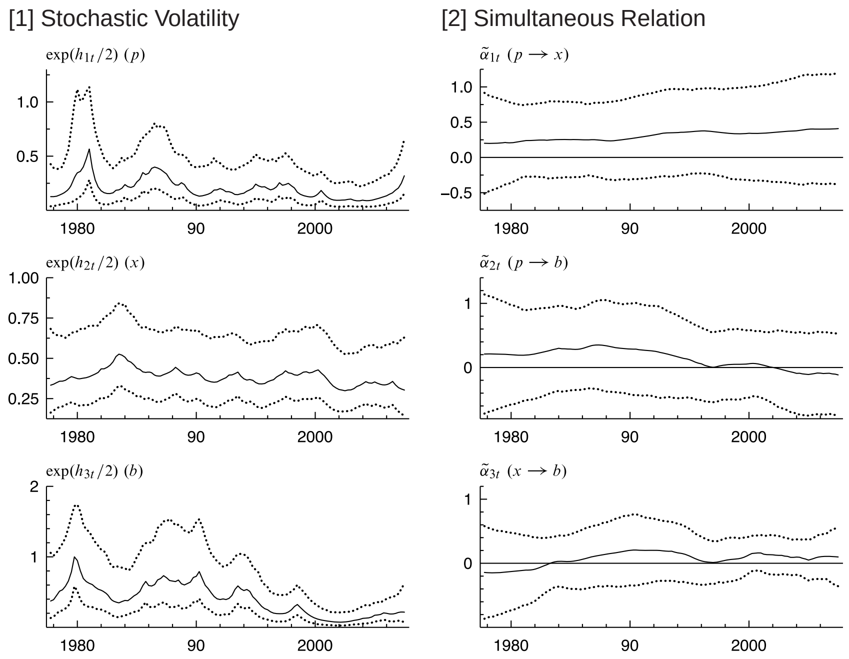
Note: Sample autocorrelations (top), sample paths (middle), and posterior densities (bottom). The estimates of Σ_β and Σ_a are multiplied by 100.

$\tilde{a}_{it}$, are plotted. This implies the size of the simultaneous effect of other variables to one unit of the structural shock based on the recursive identification.

Stochastic volatility of inflation (p) exhibits a spike around 1980 due to the second oil shock, and shows a general downward trend thereafter, with some cyclical ups and downs around this downward trend. In particular, it remains low and stable during the first half of the 2000s, when the Japanese economy experiences deflation. Stochastic volatility of output (x) remains slightly high in the early 1980s and the late 1990s. Nakajima, Kasuya, and Watanabe (2009) report that the estimated stochastic volatility of the structural shock for industrial production becomes higher in the second half of the 1990s and the beginning of the 2000s, compared to the 1980s. However, stochastic volatility of the output gap in our analysis based on GDP shows relatively moderate movements in the 1990s to 2000s. Stochastic volatility of the medium-term interest rates (b) declines significantly in the mid-1990s, when the BOJ reduces the overnight interest rates close to zero. It declines further in the late 1990s, and remains very low and stable in the late 1990s to mid-2000s, when the BOJ carries out the zero interest rate policy from 1999 to 2000 and the quantitative easing policy from 2001 to 2006.

The time-varying simultaneous relation is one of the characteristics in the TVP-VAR model. The simultaneous relation of the output to the inflation shock ($p \to x$)

Figure 6 Posterior Estimates for (1) Stochastic Volatility of the Structural Shock, $\sigma_{it} = \exp(h_{it}/2)$, and (2) Simultaneous Relation, $\tilde{a}_{it}$, for the Variable Set of (p, x, b)



Note: Posterior mean (solid line) and 95 percent credible intervals (dotted line).

stays positive, and remains almost constant over the sample period. By contrast, the simultaneous relations of the interest rates to the inflation shock $(p \to b)$ and the output shock $(x \to b)$ vary over time.

The impulse response is a basic tool to see the macroeconomic dynamics captured by the estimated VAR system. For a standard VAR model whose parameters are all time-invariant, the impulse responses are drawn for each set of two variables. By contrast, for the TVP-VAR model, the impulse responses can be drawn in an additional dimension, that is, the responses are computed at all points in time using the estimated time-varying parameters. In this case, we have several ways to simulate the impulse response based on the parameter estimates of the TVP-VAR model. Considering the comparability over time, we propose to compute the impulse responses by fixing an initial shock size equal to the time-series average of stochastic volatility over the sample period, and using the simultaneous relations at each point in time. To compute the recursive innovation of the variable, the estimated time-varying coefficients are used from the current date to future periods. Around the end of the sample period, the coefficients are set constant in future periods for convenience. A three-dimensional plot can be drawn for the time-varying impulse responses, although a time series of impulse responses for selected horizons or impulse responses for selected periods are often exhibited in the literature.

Figure 7 shows the impulse responses of the constant VAR model and the time-varying responses for the TVP-VAR model. The latter responses are drawn in a time-series manner by showing the size of the impulses for one-quarter and one- to three-year horizons over time. The time-varying nature of the macroeconomic dynamics between the variables is shown in the impulse responses, and the shape of the impulse response in the constant VAR model is associated with the average level of the response in the TVP-VAR model to some extent.

The impulse responses of output to a positive inflation shock ($\varepsilon_p \to x$) are estimated as being insignificantly different from zero using the constant-parameter VAR model, although it is remarkable that the impulse responses vary significantly over time once the TVP-VAR model is used: the impulse responses stay negative from the 1980s to the early 1990s, and they turn positive in the mid-1990s. Basic economic theory tells us that an inflation shock affects output negatively in the medium to long term, which is consistent with the negative impulse responses observed in the first half of the sample period. The positive impulse responses observed in the second half of the sample period imply the possibility of a deflationary spiral, that is, mutual reinforcement between deflation and recession. The impulse responses of inflation to a positive output shock ($\varepsilon_x \to p$) decline rapidly in the early 1980s, and remain around zero thereafter. This observation can be regarded as empirical evidence of the flattened Phillips curve. The impulse responses of output to a positive interest rate shock ($\varepsilon_b \to x$) stay negative in the 1980s, but approach very closely to zero in the mid-1990s, when nominal short-term interest rates are close to zero, and have remained around zero since then.

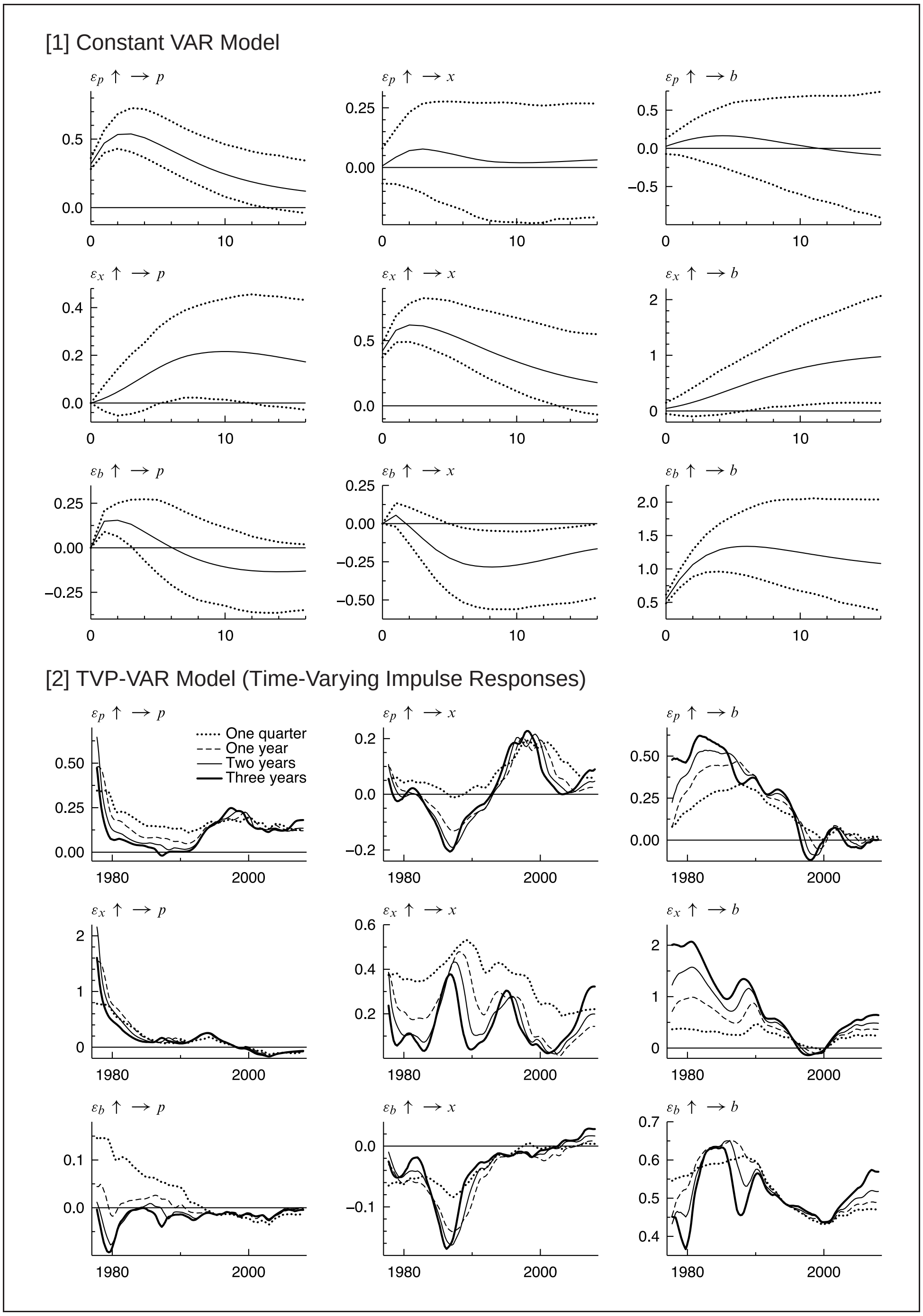
2. Estimation results for the second set of variables: $(\boldsymbol{p}, \boldsymbol{x}, \boldsymbol{i})$

Next, the variable set of (p, x, i) is estimated. Figure 8 plots the results of stochastic volatility and simultaneous relations. The stochastic volatilities of inflation and output seem to be similar to the previous analysis, and stochastic volatility of short-term interest rates (i) implies the changing variance of the monetary policy shock. Two major hikes in the interest rate volatility are observed around 1981 and 1986, and the volatility stays quite low from 1995 under virtually zero interest rate circumstances.

Regarding the simultaneous relations, the effects of inflation on output $(p \to x)$ and on interest rates $(p \to i)$ seem clearer than the previous specification. The simultaneous effects of inflation on the short-term interest rate shock diminish from the mid-1980s. At the same time, the simultaneous effects of output on interest rates $(x \to i)$ become significantly positive temporarily in the mid-1990s, but decline to zero thereafter. These observations suggest the possibility that monetary policy responses are constrained by the zero lower bound (ZLB) of nominal interest rates from the mid-1990s.

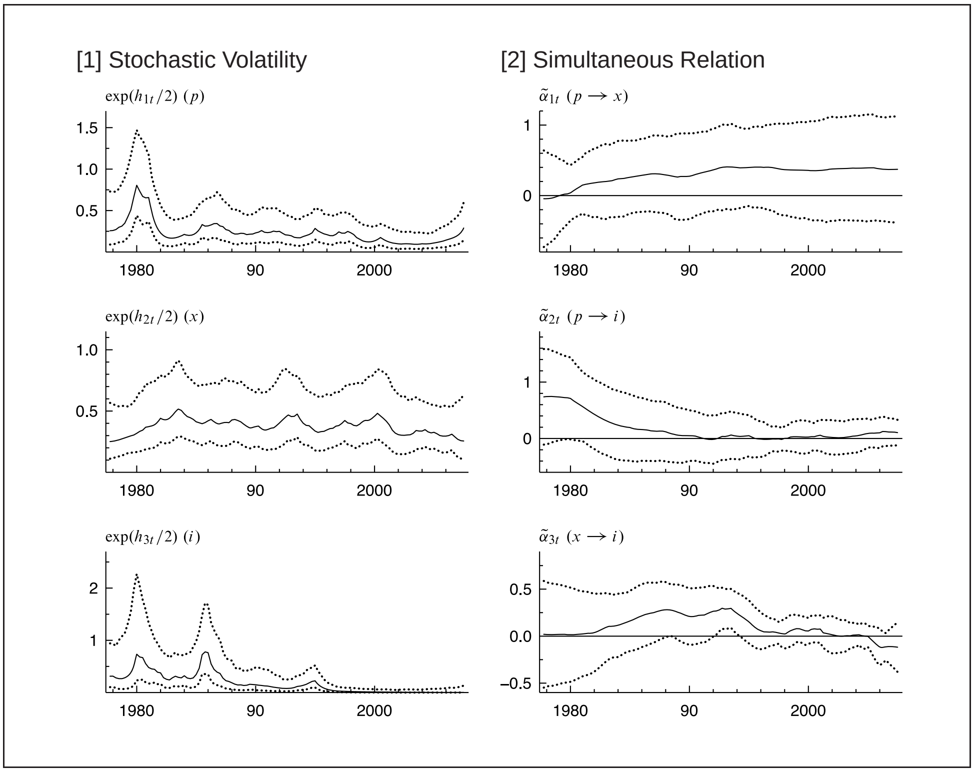
Figure 9 shows the impulse responses of estimation results for the variables set of (p, x, i). The impulse responses between inflation (p) and output (x) are similar to the previous specification. Regarding the response related to short-term interest rates, the impulse responses of inflation to a positive short-term interest rate shock ($\varepsilon_i \to p$) differ significantly from the previous specifications. The price puzzle in the 1980s becomes less evident, but time-series movements of the impulse responses become more volatile, especially from the mid-1980s.

Figure 7 Impulse Responses of (1) Constant VAR and (2) TVP-VAR Models for the Variable Set of (p, x, b)



Note: Posterior mean (solid line) and 95 percent intervals (dotted line) for the constant VAR model. Time-varying responses for one-quarter (dotted line), one-year (dashed line), two-year (solid line), and three-year (bold line) horizons for the TVP-VAR model.

Figure 8 Posterior Estimates for (1) Stochastic Volatility of the Structural Shock, $\sigma_{it} = \exp(h_{it}/2)$, and (2) Simultaneous Relation, $\tilde{a}_{it}$, for the Variable Set of (p, x, i)



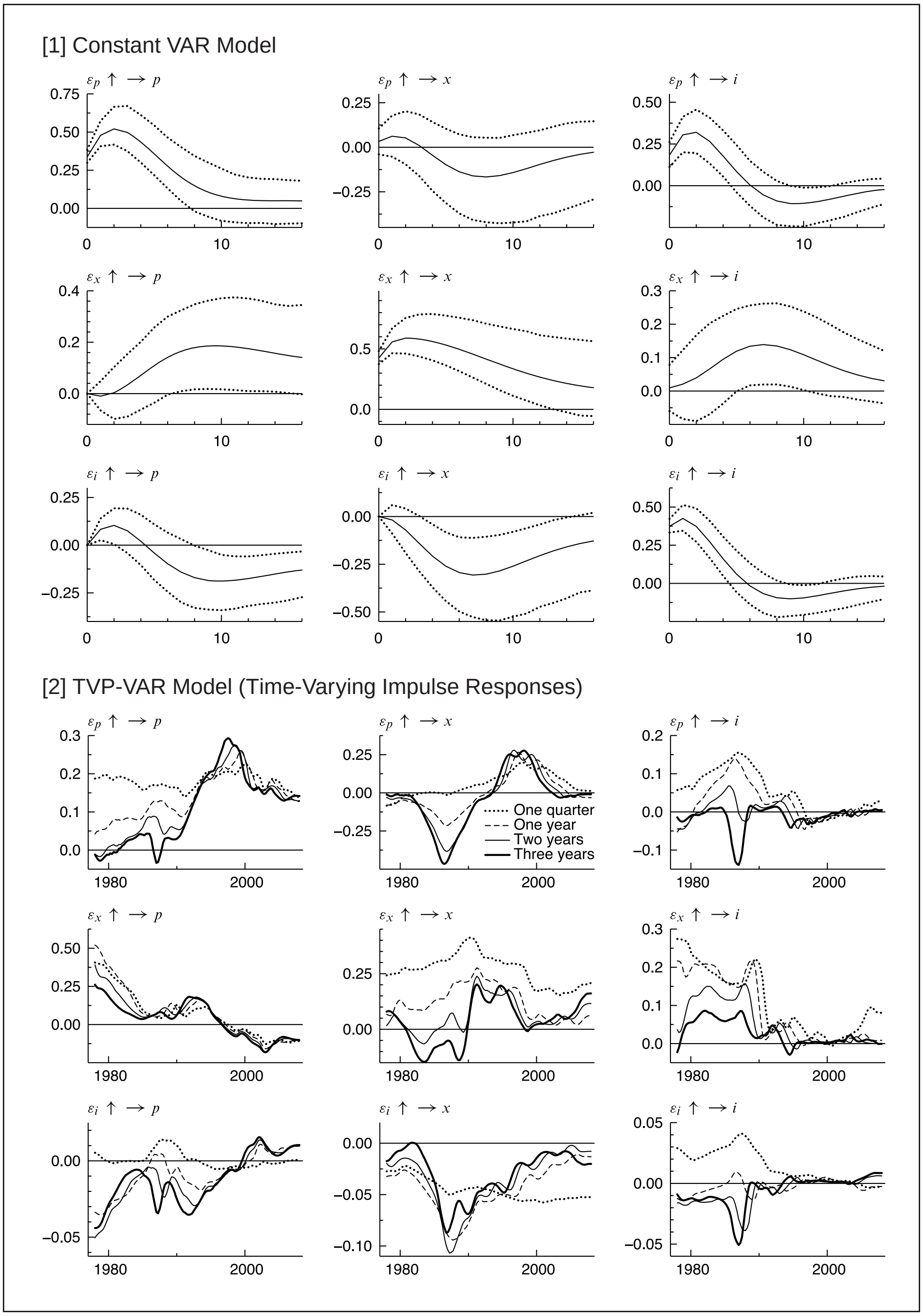
Note: Posterior mean (solid line) and 95 percent credible intervals (dotted line).

VI. Concluding Remarks

This paper provided an overview of the empirical methodology of the TVP-VAR model with stochastic volatility as well as its application to the Japanese data. The simulation exercises of the TVP regression model revealed the importance of incorporating stochastic volatility into the TVP regression models. The empirical applications using the Japanese data showed the time-varying nature of the dynamic relationships between macroeconomic variables.

Some words of caution are in order regarding the empirical application of the TVP-VAR model to data including an extremely low level of interest rates due to the ZLB of nominal interest rates. Nominal interest rates cannot become negative in the real world, although the ZLB of nominal interest rates is not assumed explicitly in the standard specification of the TVP-VAR model, as developed in this paper. Under the ZLB of nominal interest rates, structural shocks should not be observed in the VAR system. It is natural that stochastic volatility of the short-term interest rates is estimated to be very low in the related periods and that the time-varying impulse response of interest rates to some shocks of economic variables is equal to zero. However, other impulse responses

Figure 9 Impulse Responses of (1) Constant VAR and (2) TVP-VAR Models for the Variable Set of (p, x, i)



Note: Posterior mean (solid line) and 95 percent intervals (dotted line) for the constant VAR model. Time-varying responses for one-quarter (dotted line), one-year (dashed line), two-year (solid line), and three-year (bold line) horizons for the TVP-VAR model.

related to the interest rates in Figure 9 are not zero but fluctuating for the involved periods in which the short-term interest rates never change. To solve this problem, Nakajima (2011) proposes a TVP-VAR model with the ZLB of nominal interest rates and presents empirical findings using Japanese economic data.

The technique of the TVP-VAR model has been recently extended to the factor-augmented VAR (FAVAR, originally proposed by Bernanke, Boivin, and Eliasz [2005]) models. The MCMC algorithm illustrated in this paper can be straightforwardly applied to the estimation of the TVP-FAVAR model. Several studies show the empirical evidence of the TVP-FAVAR models (e.g., Korobilis [2009] and Baumeister, Liu, and Mumtaz [2010]). The TVP-VAR model has great potential as a very flexible toolkit to analyze the evolving structure of the modern economy.

APPENDIX

A. Joint Posterior Distribution for the TVP Regression Model

Given data y, we obtain the joint posterior distribution of (θ, α, h) as

$$
\begin{aligned}
&\pi(\theta, \alpha, h \mid y) \\
&\propto \pi(\theta) \times \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\gamma e^{h_t/2}}} \exp\left\{-\frac{(y_t - x_t'\beta - z_t'\alpha_t)^2}{2\gamma e^{h_t}}\right\} \\
&\quad \times \prod_{t=1}^{n-1} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\alpha_{t+1} - \alpha_t)'\Sigma^{-1}(\alpha_{t+1} - \alpha_t)\right\} \\
&\quad \times \frac{1}{(2\pi)^{k/2}|\Sigma_0|^{1/2}} \exp\left\{-\frac{1}{2}(\alpha_1'\Sigma_0^{-1}\alpha_1)\right\} \\
&\quad \times \prod_{t=1}^{n-1} \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left\{-\frac{(h_{t+1} - \phi h_t)^2}{2\sigma_\eta^2}\right\} \times \frac{\sqrt{1-\phi^2}}{\sqrt{2\pi}\sigma_\eta} \exp\left\{-\frac{(1-\phi^2)h_1^2}{2\sigma_\eta^2}\right\}.
\end{aligned}
$$

B. Multi-Move Sampler for the TVP Regression Model

In this paper, the multi-move sampler is applied to draw samples from the conditional posterior density of stochastic volatility in the TVP regression model. This appendix shows the algorithm of the multi-move sampler following Shephard and Pitt (1997) and Watanabe and Omori (2004). We rewrite the model as

$$y_t^* = \exp(h_t/2)e_t, \quad t = 1, \ldots, n,$$

$$h_{t+1} = \phi h_t + \eta_t, \quad t = 0, \ldots, n-1,$$

$$\begin{pmatrix} e_t \\ \eta_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right), \quad t = 1, \ldots, n,$$

where $y_t^* = (y_t - x_t'\beta - z_t'\alpha_t)/\sqrt{\gamma}$, $h_0 = 0$, and $\eta_0 \sim N(0, \sigma_\eta^2/(1-\phi^2))$. For sampling a typical block $(h_r, \ldots, h_{r+d})$ from its joint posterior density (note that $r \geq 1$, $d \geq 1$, $r + d \leq n$), we consider the draw of

$$\begin{aligned}(\eta_{r-1}, \ldots, \eta_{r+d-1}) &\sim \pi(\eta_{r-1}, \ldots, \eta_{r+d-1} \,|\, \omega) \\ &\propto \prod_{t=r}^{r+d} \frac{1}{e^{h_t/2}} \exp\left(-\frac{y_t^{*2}}{2e^{h_t}}\right) \times \prod_{t=r-1}^{r+d-1} f(\eta_t) \times f(h_{r+d}),\end{aligned} \tag{A.1}$$

where

$$f(\eta_t) = \begin{cases} \exp\left\{-\dfrac{(1-\phi^2)\eta_0^2}{2\sigma_\eta^2}\right\} & (\text{if } t = 0), \\ \exp\left(-\dfrac{\eta_t^2}{2\sigma_\eta^2}\right) & (\text{if } t \geq 1), \end{cases}$$

$$f(h_{r+d}) = \begin{cases} \exp\left\{-\dfrac{(h_{r+d+1} - \phi h_{r+d})^2}{2\sigma_\eta^2}\right\} & (\text{if } r + d < n), \\ 1 & (\text{if } r + d = n), \end{cases}$$

and $\omega = (h_{r-1}, h_{r+d+1}, \beta, \gamma, \phi, \sigma_\eta, \alpha, y)$. The posterior draw of $(h_r, \ldots, h_{r+d})$ can be obtained by running the state equation with the draw of $(\eta_{r-1}, \ldots, \eta_{r+d-1})$ given h_{r-1}.

We sample $(\eta_{r-1}, \ldots, \eta_{r+d-1})$ from the density (A.1) using the acceptance-rejection MH (AR-MH) algorithm (see, e.g., Tierney [1994] and Chib and Greenberg [1995]) with the following proposal distribution. Our construction of the proposal density begins with the second-order Taylor expansion of

$$g(h_t) \equiv -\frac{h_t}{2} - \frac{y_t^{*2}}{2e^{h_t}},$$

around a certain point $\hat{h}_t$, which is discussed later, namely,

$$\begin{aligned} g(h_t) &\approx g(\hat{h}_t) + g'(\hat{h}_t)(h_t - \hat{h}_t) + \frac{1}{2}g''(\hat{h}_t)(h_t - \hat{h}_t)^2 \\ &\propto \frac{1}{2}g''(\hat{h}_t)\left\{h_t - \left(\hat{h}_t - \frac{g'(\hat{h}_t)}{g''(\hat{h}_t)}\right)\right\}^2. \end{aligned}$$

We have

$$g'(\hat{h}_t) = -\frac{1}{2} + \frac{y_t^{*2}}{2e^{h_t}}, \quad g''(\hat{h}_t) = -\frac{y_t^{*2}}{2e^{h_t}}.$$

We use the proposal density formed as

$$q(\eta_{r-1}, \ldots, \eta_{r+d-1} \,|\, \omega) \propto \prod_{t=r}^{r+d} \exp\left\{-\frac{(h_t^* - h_t)^2}{2\sigma_t^{*2}}\right\} \times \prod_{t=r-1}^{r+d-1} f(\eta_t),$$

where

$$\sigma_t^{*2} = -\frac{1}{g''(\hat{h}_t)}, \quad h_t^* = \hat{h}_t + \sigma_t^{*2} g'(\hat{h}_t), \tag{A.2}$$

for $t = r, \ldots, r+d-1$, and $t = r+d$ (when $r+d = n$). For $t = r+d$ (when $r+d < n$),

$$\sigma_{r+d}^{*2} = \frac{1}{-g''(\hat{h}_{r+d}) + \phi^2/\sigma_\eta^2}, \tag{A.3}$$

$$h_{r+d}^* = \sigma_{r+d}^{*2}\{g'(\hat{h}_{r+d}) - g''(\hat{h}_{r+d})\hat{h}_{r+d} + \phi h_{r+d+1}/\sigma_\eta^2\}. \tag{A.4}$$

The choice of this proposal density is derived from its correspondence to the state space model

$$h_t^* = h_t + \zeta_t, \quad t = r, \ldots, r+d,$$

$$h_{t+1} = \phi h_t + \eta_t, \quad t = r-1, \ldots, r+d-1, \tag{A.5}$$

$$\begin{pmatrix} \zeta_t \\ \eta_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_t^{*2} & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right), \quad t = r, \ldots, r+d,$$

with $\eta_{r-1} \sim N(0, \sigma_\eta^2)$, when $r \geq 2$, and $\eta_0 \sim N(0, \sigma_\eta^2/(1-\phi^2))$. Given ω, we draw a candidate point of $(\eta_{r-1}, \ldots, \eta_{r+d-1})$ for the AR-MH algorithm by running the simulation smoother over the state-space representation (A.5).

Now we find $(\hat{h}_r, \ldots, \hat{h}_{r+d})$, for which it is desirable to be near the mode of the posterior density for an efficient sampling. We loop the following steps several times enough to reach near the mode:

(1) Initialize $(\hat{h}_r, \ldots, \hat{h}_{r+d})$.
(2) Compute $(h_r^*, \ldots, h_{r+d}^*)$, $(\sigma_r^{2*}, \ldots, \sigma_{r+d}^{2*})$ by (A.2) and (A.4).
(3) Run the moment smoother using the current $(h_r^*, \ldots, h_{r+d}^*)$, $(\sigma_r^{2*}, \ldots, \sigma_{r+d}^{2*})$ on (A.5) and obtain $\hat{h}_t^* \equiv E(h_t \,|\, \omega)$ for $t = r, \ldots, r+d$.
(4) Replace $(\hat{h}_r, \ldots, \hat{h}_{r+d})$ by $(\hat{h}_r^*, \ldots, \hat{h}_{r+d}^*)$.
(5) Go to (2).

Note that the $E(h_t\,|\omega)$ is the product in the simulation smoother as $\Lambda_t r_t$ with $\varepsilon_t = 0$. We divide $(h_1, \ldots, h_n)$ into $K + 1$ blocks, say, $(h_{k_{i-1}+1}, \ldots, h_{k_i})$ for $i = 1, \ldots, K + 1$ with $k_0 = 0$ and $k_{K+1} = n$, and sample each block recursively. One remark should be made about the determination of $(k_1, \ldots, k_K)$. The method called stochastic knots (Shephard and Pitt [1997]) proposes $k_i = \mathrm{int}[n(i + U_i)/(K + 2)]$, for $i = 1, \ldots, K$, where U_i is a random sample from the uniform distribution $U[0, 1]$. We randomly choose $(k_1, \ldots, k_K)$ for every iteration of MCMC sampling for a flexible draw of $(h_1, \ldots, h_n)$.

Baumeister, C., and L. Benati, "Unconventional Monetary Policy and the Great Recession—Estimating the Impact of a Compression in the Yield Spread at the Zero Lower Bound," Working Paper No. 1258, European Central Bank, 2010.

———, E. Durinck, and G. Peersman, "Liquidity, Inflation and Asset Prices in a Time-Varying Framework for the Euro Area," Working Paper Research No. 142, National Bank of Belgium, 2008.

———, P. Liu, and H. Mumtaz, "Changes in the Transmission of Monetary Policy: Evidence from a Time-Varying Factor-Augmented VAR," Working Paper No. 401, Bank of England, 2010.

Benati, L., "The Great Moderation in the United Kingdom," *Journal of Money, Credit and Banking,* 40, 2008, pp. 121–147.

———, and H. Mumtaz, "The 'Great Stability' in the U.K.: Good Policy or Good Luck?" mimeo, Bank of England, 2005.

———, and P. Surico, "Evolving U.S. Monetary Policy and the Decline of Inflation Predictability," *Journal of the European Economic Association,* 6 (2–3), 2008, pp. 643–646.

Bernanke, B., J. Boivin, and P. Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics,* 120 (1), 2005, pp. 387–422.

Black, F., "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association*, 1976, pp. 177–181.

Brown, R. L., J. Durbin, and J. M. Evans, "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society,* B37 (2), 1975, pp. 149–163.

Chib, S., "Markov Chain Monte Carlo Methods: Computation and Inference," in J. J. Heckman and E. Leamer, eds. *Handbook of Econometrics,* Volume 5, Amsterdam: North-Holland, pp. 3569–3649, 2001.

———, and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician,* 49 (4), 1995, pp. 327–335.

———, and ———, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory,* 12 (3), 1996, pp. 409–431.

Christiano, L. J., M. Eichenbaum, and C. L. Evans, "Monetary Policy Shocks: What Have We Learned and to What End?" in J. B. Taylor and M. Woodford, eds. *Handbook of Macroeconomics,* Volume 3, Amsterdam: Elsevier Science B. V., 1999, pp. 65–148.

Clark, T., and S. Terry, "Time Variation in the Inflation Passthrough of Energy Prices," *Journal of Money, Credit and Banking,* 42 (7), 2010, pp. 1419–1433.

Cogley, T., and T. J. Sargent, "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual,* 16, 2001, pp. 331–373.

———, and ———, "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics,* 8, 2005, pp. 262–302.

D'Agostino, A., L. Gambetti, and D. Giannone, "Macroeconomic Forecasting and Structural Change," ECB Working Paper No. 1167, European Central Bank, 2010.

de Jong, P., and N. Shephard, "The Simulation Smoother for Time Series Models," *Biometrika,* 82 (2), 1995, pp. 339–350.

Doornik, J., *Ox: Object Oriented Matrix Programming,* London: Timberlake Consultants Press, 2006.

Durbin, J., and S. J. Koopman, *Time Series Analysis by State Space Methods,* Oxford: Oxford University Press, 2001.

———, and ———, "Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika,* 89 (3), 2002, pp. 603–616.

Gamerman, D., and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,* Second Edition, Boca Raton: Chapman & Hall/CRC, 2006.

Geweke, J., "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo *et al.,* eds. *Bayesian Statistics,* Volume 4, New York: Oxford University Press, 1992, pp. 169–188.

———, *Contemporary Bayesian Econometrics and Statistics,* Wiley, 2005.

Ghysels, E., A. C. Harvey, and E. Renault, "Stochastic Volatility," in C. R. Rao and G. S. Maddala, eds. *Statistical Methods in Finance,* Amsterdam: North-Holland, 2002, pp. 119–191.

Harvey, A. C., *Time Series Models,* Second Edition, Hemel Hempstead: Harvester Wheatsheaf, 1993.

Kim, S., N. Shephard, and S. Chib, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies,* 65 (3), 1998, pp. 361–393.

Koop, G., *Bayesian Econometrics,* Hemel Hempstead: Wiley-Interscience, 2003.

———, and D. Korobilis, "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends in Econometrics,* 3, 2010, pp. 267–358.

Korobilis, D., "Assessing the Transmission of Monetary Policy Shocks Using Dynamic Factor Models," mimeo, 2009.

Lancaster, T., *Introduction to Modern Bayesian Econometrics,* Hemel Hempstead: Wiley-Blackwell, 2003.

Leeper, E. M., C. A. Sims, and T. Zha, "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity,* 2, 1996, pp. 1–63.

Mumtaz, H., and P. Surico, "Time-Varying Yield Curve Dynamics and Monetary Policy," *Journal of Applied Econometrics,* 24 (6), 2009, pp. 895–913.

Nakajima, J., "Monetary Policy Transmission under Zero Interest Rates: An Extended Time-Varying Parameter Vector Autoregression Approach," IMES Discussion Paper No. 2011-E-8, Institute for Monetary and Economic Studies, Bank of Japan, 2011.

———, M. Kasuya, and T. Watanabe, "Bayesian Analysis of Time-Varying Parameter Vector Autoregressive Model for the Japanese Economy and Monetary Policy," IMES Discussion Paper No. 2009-E-13, Institute for Monetary and Economic Studies, Bank of Japan, 2009.

———, S. Shiratsuka, and Y. Teranishi, "The Effects of Monetary Policy Commitment: Evidence from Time-Varying Parameter VAR Analysis," IMES Discussion Paper No. 2010-E-6, Institute for Monetary and Economic Studies, Bank of Japan, 2010.

———, and Y. Teranishi, "The Evolution of Loan Rate Stickiness across the Euro Area," IMES Discussion Paper No. 2009-E-10, Institute for Monetary and Economic Studies, Bank of Japan, 2009.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima, "Stochastic Volatility with Leverage: Fast Likelihood Inference," *Journal of Econometrics,* 140 (2), 2007, pp. 425–449.

Perron, P., "Dealing with Structural Breaks," in K. Patterson and T. C. Mills, eds. *Palgrave Handbook of Econometrics,* Volume 1, Palgrave Macmillan, 2006, pp. 278–352.

Primiceri, G. E., "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies,* 72 (3), 2005, pp. 821–852.

Sekine, T., "Time-Varying Exchange Rate Pass-Through: Experiences of Some Industrial Countries," BIS Working Paper No. 202, Bank for International Settlements, 2006.
———, and Y. Teranishi, "Inflation Targeting and Monetary Policy Activism," IMES Discussion Paper No. 2008-E-13, Institute for Monetary and Economic Studies, Bank of Japan, 2008.
Shephard, N., ed. *Stochastic Volatility: Selected Readings,* Oxford: Oxford University Press, 2005.
———, and M. Pitt, "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika,* 84 (3), 1997, pp. 653–667.
Sims, C. A., "Macroeconomics and Reality," *Econometrica*, 48 (1), 1980, pp. 1–48.
———, "Comment on Sargent and Cogley's 'Evolving Post-World War II U.S. Inflation Dynamics,'" *NBER Macroeconomics Annual,* 16, 2001, pp. 373–379.
Stock, J. H., "Discussion of Sargent and Cogley's 'Evolving Post-World War II U.S. Inflation Dynamics,'" *NBER Macroeconomics Annual,* 16, 2001, pp. 379–387.
Tierney, L., "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics,* 22, 1994, pp. 1701–1728.
Uhlig, H., "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica,* 65 (1), 1997, pp. 59–73.
Watanabe, T., and Y. Omori, "A Multi-move Sampler for Estimating Non-Gaussian Time Series Models: Comments on Shephard and Pitt (1997)," *Biometrika,* 91 (1), 2004, pp. 246–248.
West, M., and P. J. Harrison, *Bayesian Forecasting and Dynamic Models,* Second Edition, New York: Springer-Verlag, 1997.