IMES DISCUSSION PAPER SERIES

Identification of Monetary Policy Shocks in Japan Using Sign Restrictions within the TVP-VAR Framework

Michal Franta

Discussion Paper No. 2011-E-13

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

IMES Discussion Paper Series 2011-E-13
June 2011

Identification of Monetary Policy Shocks in Japan Using Sign Restrictions within the TVP-VAR Framework

Michal Franta*

Abstract

This paper contributes to the discussion on the functioning of the monetary policy transmission mechanism in Japan during the past three decades. It extends the methodology of time-varying parameter vector autoregressions (TVP-VAR) by employing an identification scheme based on sign restrictions. This approach allows for an explicit account of the zero lower bound on the nominal interest rate. Results suggest differences in the transmission mechanism between the quantitative easing policy period and the periods when the call rate played the role of a monetary policy instrument. Monetary policy operating through call rate movements is found to influence output more than when it targets banks' balances held at the central bank. Monetary policy operating through quantitative easing is found to influence inflation, in sharp contrast to the previous literature.

Keywords: Structural vector autoregressive model; time-varying parameters; sign restrictions; unconventional monetary policy; zero lower bound
JEL classification: C11, C15, E52

***** Czech National Bank, Economic Research Department (E-mail: michal.franta@cerge-ei.cz)

This paper was prepared in part while the author was a visiting scholar at the Institute for Monetary and Economic Studies (IMES), Bank of Japan. The author thanks Toshiaki Watanabe, Kozo Ueda, Kentarou Koyama, Shigenori Shiratsuka, Jan Libich and seminar participants at the IMES for helpful comments and suggestions. Views expressed in this paper are those of the author and do not necessarily reflect the official views of the Bank of Japan and Czech National Bank.

1. Introduction

The analysis of the Japanese monetary policy transmission mechanism within vector autoregressions (VAR) has been the subject of extensive discussions. Miyao (2002) introduces the benchmark VAR model for monetary policy analysis. Changes in the transmission mechanism over time are accounted for in Kimura et al. (2003), who estimate a VAR with time-varying coefficients, and Nakajima et al. (2009), who add the possibility of time-varying volatilities. In addition, several papers deal explicitly with the zero lower bound on the interest rate: Fujiwara (2006) uses a Markov-switching VAR model, Iwata and Wu (2006) employ a censored VAR model and Kamada and Sugo (2006) set up a monetary policy proxy that can attain negative values.

Despite the effort devoted to the description of monetary policy effects in Japan, many questions still require analysis. Unanswered questions relate to the effectiveness of the quantitative easing (QE) policy at the beginning of the 2000s, the functioning of the transmission mechanism and conventional policies near the zero lower bound during the late 1990s, the effectiveness of aggressive rate reductions at the beginning of the 1990s, and changes in the transmission mechanism during the recent economic crisis, among others.

To contribute to the discussion of such questions, a new modeling framework is developed in this paper. The reduced-form time-varying parameters VAR model *à la* Primiceri (2005) is extended to the identification of shocks based on sign restrictions. Sign restrictions allow for an intuitive way of incorporating restrictions arising from the presence of the zero lower bound on the nominal interest rate, and enable the avoidance of puzzles common in the VAR literature. Furthermore, the set of endogenous variables involves operating targets used by the Bank of Japan (BoJ) during the past three decades. Thus, both conventional and unconventional monetary policies are investigated in a flexible modeling framework, and the maximum of information is exploited.

The identification scheme based on sign restrictions represents an additional source of uncertainty. In contrast to the usual recursive assumption, sign restrictions are "weak", in the sense of the number of structural models underlying the identifying assumptions. In general, they are satisfied by many models, implying different propagation of monetary policy shocks. The recursive assumption represents just one model of data, and estimation uncertainty arises solely from the uncertainty of reduced-form parameter estimates. On the other hand, the recursive assumption is often questioned because of its strict implications for the contemporaneous effects between endogenous variables.

Estimation results provide a picture of the transmission mechanism in Japan during the past three decades. The analysis of impulse responses highlights qualitative differences in the monetary policy transmission mechanism between the QE policy period and the periods when the short-term interest rate acted as the BoJ's policy instrument. Cutting the interest rate brings about an increase in output, which reaches a peak after a year or so, and the effect on inflation is weak. On the other hand, during the QE period, the opposite is observed: output responds temporarily only but inflation increases following a change in the monetary base growth. This estimated strong effect of QE on the price level is in contrast to the previous literature. The difference arises as a consequence of our methodology, which allows for changes in variances of shocks, and of the identification approach employed.

Within the QE period, no important differences in the transmission mechanism are detected. Similar results are obtained for the period 2007–2009: the recent economic crisis does not greatly affect the transmission. Of course, changes in the size of the monetary policy shock are observed, as characterized by the changing magnitude of the immediate impact of the shock on the endogenous variables.

Furthermore, the criticism of the sign restrictions approach is addressed in this paper. First, it is demonstrated that additional uncertainty of results related to such an identification approach is of little importance and that the presented approach can provide conclusive results. The main results are significant in terms of certain percentiles of the impulse response distribution supporting the conclusions. Second, Fry and Pagan's (2007) impulse responses based strictly on orthogonal shocks are discussed. We show that the traditional approach based on simple summary statistics of impulse response distributions is preferable.

Let us point out that our investigation of the transmission mechanism is not only of interest from the Japanese perspective. Shirakawa (2010), for example, discusses several similarities in economic conditions and the policy response in Japan after the asset bubble burst at the beginning of the 1990s and the US and European economies during the recent crisis. Reconsidering the approach to monetary policy analysis based on vector autoregressions can thus be useful for other developed countries.

The rest of the paper is organized as follows. Japan's monetary policy is briefly discussed in Section 2. A reduced-form VAR with time-varying parameters is introduced in Section 3. Identification issues and the implementation of the zero lower bound are described in Section 4. Section 5 presents the results and an additional discussion related to the methodology. Section 6 concludes the paper.

2. Monetary policy in Japan

In this section, the history of monetary policy conducted by the BoJ during the past three decades is briefly reviewed, with a focus on the events that are to be captured by the modeling framework.

Since the asset bubble burst at the beginning of the 1990s, the performance of the Japanese economy has worsened significantly (for example, real GDP growth was 1.5 percent in the 1990s compared with 4.4 percent in the 1980s). To stimulate the economy, the BoJ lowered its operating target—the uncollateralized overnight call rate, gradually reaching values below 0.5 by 1995. Since 1999, the call rate has been virtually zero; the BoJ followed the zero interest rate policy (ZIRP) until August 2000. In April 1999, the BoJ introduced a commitment to the zero interest rate "until deflationary concerns are dispelled" to influence market expectations and thus stabilize interest rates.[1]

In March 2001, the BoJ implemented a new monetary policy framework based on QE.[2] The QE policy was designed to overcome deflation and realize long-lasting economic growth after the IT bubble burst. The QE policy was based on three pillars. First, the framework consisted of the introduction of a new operating target—the BoJ switched the operating target from the call rate to the outstanding balances of current accounts held by financial institutions at the bank. The target for the current account balances increased from ¥5 trillion at the beginning of 2001 to ¥30–35 trillion at the beginning of 2004 and remained unchanged until the QE policy ended in 2006.

Second, the BoJ announced a commitment to pursue the QE policy until deflationary concerns disappeared. The commitment was built on the core consumer price index (CPI), i.e., CPI excluding perishables, and was clarified in terms of duration of the nonnegative core CPI growth for a few consecutive months in October 2003. Third, the BoJ announced its willingness to increase the amount of the outright purchases of long-term Japanese government bonds in order to provide sufficient liquidity. The BoJ also purchased asset-backed securities from July 2003 to March 2006, altering the composition of its balance sheet. In March 2006, the BoJ returned to the call rate as its operating target, with the value close to zero.

3. TVP-VAR

Vector autoregressions with time-varying parameters (TVP-VAR) are increasingly popular, and not just in the field of monetary policy analysis. Although they are not based on the optimization problems of economic agents, VARs with time variation of parameters can deal with regime changes, with time variation in the lag structure of the model or with nonlinearities that emerge, for example, when the zero lower bound on the interest rate is a factor. On the other hand, having a huge number of parameters to estimate in TVP-VARs often results in high uncertainty regarding the estimates and, thus, inconclusiveness of the results.[3]

[1] The effect of the announcement is investigated in Fujiki and Shiratsuka (2002).

[2] A detailed description of the QE policy and its effectiveness can be found in Shiratsuka (2010) and Ugai (2007).

[3] An up-to-date survey of the literature on TVP-VARs and an estimated TVP-VAR for Japan can be found in Nakajima (2011b).

Our model and estimation approach follow Primiceri (2005).[4] The important feature of the model is that it allows both coefficients and shock variances to change over time.

We assume that data can be represented as a finite-order VAR with time-varying parameters, as follows:

$$A_t y_t = X_t \gamma_t + \Sigma_t \varepsilon_t, \qquad t = p+1,...,T, \tag{1}$$

where y_t is an $M \times 1$ vector of endogenous variables, $X_t = I_M \otimes \left(1, y'_{t-1},..., y'_{t-p}\right)$ is a Kronecker product of identity matrix and lagged vectors of endogenous variables (and a constant) and ε_t is an $M \times 1$ vector of uncorrelated disturbances representing structural shocks. The matrix capturing contemporaneous relations, A_t, is lower triangular, as shown below:

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

γ_t is an $M(Mp+1)\times 1$ vector of the effects of lagged endogenous variables on the system, and the matrix of standard deviations of structural shocks, Σ_t, is diagonal:

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{M,t} \end{bmatrix}.$$

The reduced form to be estimated takes the following form:

$$y_t = X_t \beta_t + A_t^{-1} \Sigma_t \varepsilon_t \qquad t = p+1,...,T, \tag{2}$$

[4] Primiceri (2005) provides an extensive discussion of the model specification, assumptions and the estimation technique used. As the reduced-form model and its estimation are not the focus of this paper, we refer the reader to Primiceri (2005) for details.

where the time-varying parameters follow random walks and a geometric random walk:

$$\beta_{i,t} = \beta_{i,t-1} + u_t^i \qquad i = 1,\dots,M^2 p + M \ , \tag{3}$$

$$\alpha_{i,t} = \alpha_{i,t-1} + v_t^i \qquad i = 1,\dots,\left(M^2 - M\right)/2 \ , \tag{4}$$

$$\log(\sigma_{i,t}) = \log(\sigma_{i,t-1}) + w_t^i \qquad i = 1,\dots,M \ . \tag{5}$$

Model innovations are assumed to be jointly normally distributed:

$$\begin{bmatrix} \varepsilon_t \\ u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} I_M & 0 & 0 & 0 \\ 0 & U & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & W \end{bmatrix}\right),$$

where vectors u_t, v_t and w_t consist of innovations as introduced in (3)–(5). The matrices *U*, *V* and *W* are positive definite. Moreover, *V* is assumed to be block diagonal, with blocks constituted by the coefficient innovations from a particular equation, i.e., we assume that innovations to contemporaneous effects are uncorrelated across equations.

The estimation of the system is carried out using the Bayesian approach. To evaluate the posterior distribution of the unobservable states $\alpha_{i,t}, \beta_{j,t}$ and $\sigma_{k,t}$ and the hyperparameters *U*, *V* and *W*, the Gibbs sampler is employed. Instead of sampling from the joint posterior of the parameter set, the Gibbs sampler provides draws from the conditional posteriors of the subsets of the whole parameter set. Therefore, after initializing parameters, a draw is taken from the distribution of the vector β (for all *t*), given the other parameters A, Σ, hyperparameters *U* and the data. Then, a draw from *U* is produced. Similarly, the sampler goes on over the other subsets of the parameter set (*A*, *V*, Σ and *W*). The cycle is repeated to produce posterior numerical evaluations of the parameters. Details of the sampling algorithm can be found in Appendix A.

To evaluate posteriors, prior distributions need to be specified. The hyperparameters *U*, *V* and *W* are distributed as an independent inverse Wishart. The priors for initial states of the other parameters (the log of a parameter in the case of volatilities) are distributed normally. Ordinary least squares estimates based on a part of the data set are used in the specification of the prior distribution. The hyperparameters and initial states for the other parameters are assumed to be independent. In summary, we assume that:

$$\beta_0 \sim N\left(\hat{\beta}_{OLS}, 4Var\left(\hat{\beta}_{OLS}\right)\right),$$

$$\alpha_0 \sim N(\hat{\alpha}_{OLS}, 4Var(\hat{\alpha}_{OLS})),$$

$$\log(\sigma_0) \sim N(\log(\hat{\sigma}_{OLS}), I_4),$$

$$U \sim IW\left(40k_U^2 Var\left(\hat{\beta}_{OLS}\right), 80\right),$$

$$V_{block} \sim IW\left((1+size_{block})k_V^2 Var\left(\hat{\alpha}_{block,OLS}\right), 1+size_{block}\right),$$

$$W \sim IW\left(5k_W^2 I_4, 5\right),$$

where the index *block* denotes blocks corresponding to a particular equation (the system contains three blocks of sizes 2, 3 and 4). Finally, $k_U = 0.01$, $k_V = 0.1$ and $k_W = 0.01$. The priors are chosen as in Primiceri (2005), except for the number of degrees of freedom for the hyperparameter *U*, which is slightly tighter to avoid problems associated with convergence of the sampling algorithm.

4. Identification

To describe the effect of a shock, the identification of the shock is a necessary starting point. As the identification of shocks is the main focus of the paper, we discuss identification issues in detail.

Regardless of whether TVP-VARs are used for an analysis of monetary policy (Primiceri, 2005, Darvas, 2009, Nakajima et al., 2009) or for other purposes (e.g., see Kirchner et al., 2010, for a fiscal policy analysis), the recursive assumption is employed.[5] On the one hand, the recursive assumption, materialized in the Cholesky decomposition of the reduced-form innovations variance–covariance matrix, is convenient from the estimation point of view (Smith and Kohn, 2002). On the other hand, the assumption that some endogenous variables do not react contemporaneously on the others is often questioned, especially when a data set of quarterly frequency is employed. In our case, for example, the assumption that a shock to the monetary base does not affect the interest rate in a given quarter seems inappropriate, particularly in the

[5] Baumeister and Benati (2010) employ TVP-VAR with shocks identified by sign restrictions to assess the effects of the compression in the yield spread during the recent crisis.

period of QE.[6] The recursive assumption leads to underestimation of structural shocks for variables that co-move with those ordered earlier.

One possible way to avoid short-run inertial restrictions in the identification process is to impose qualitative information requirements on impulse responses. Usually, the sign of the reaction to a shock is prescribed to identify the shock. Sign restrictions can be imposed on a subset of responses to a shock, leaving other responses to be determined by the data (Faust, 1998, Uhlig, 2005) or more shocks are identified (Canova and de Nicolo, 2002, Peersman, 2005). Moreover, the use of sign restrictions in this paper is justified by the fact that it enables the implementation of the zero lower bound in an intuitive way.

4.1 Sign restrictions in a TVP-VAR framework

The model shown in (2)–(5) is estimated assuming the recursive determination of the system, i.e., A_t is assumed to be a lower triangular matrix. Let Q be an orthonormal $M \times M$ matrix, i.e., $Q'Q = I_M$. Then:

$$y_t = X_t\beta_t + A_t^{-1}\Sigma_t Q'Q\varepsilon_t = X_t\beta_t + A_t^{-1}\Sigma_t Q'\tilde{\varepsilon},$$

where the variance of the new shocks $\tilde{\varepsilon}_t = Q\varepsilon$ is $\mathrm{var}(\tilde{\varepsilon}) = \mathrm{var}(Q\varepsilon) = QI_M Q' = I_M$. In other words, we constructed a new set of uncorrelated shocks with unit variances. In general, the orthonormal matrix Q' affects both the contemporaneous effects of shocks on variables and the standard deviations of shocks. Thus, impulse responses generated by the new set of shocks $\tilde{\varepsilon}_t$ change.

To cover the space of all orthonormal matrices, we use Givens rotations Q_{12}, Q_{13}, Q_{14}, Q_{23}, Q_{24} and Q_{34}, where $Q_{kl}(\theta)$ is a matrix with $\cos(\theta)$ at the positions *k,k* and *l,l*, $-\sin(\theta)$ at the position *l,k* and $\sin(\theta)$ at the position *k,l*, (*k,l*=1,…,4). Other elements on the diagonal are equal to one, whereas out of the diagonal, elements equal zero.[7]

Therefore, for example:

[6] The correlation coefficient between the uncollateralized overnight call rate and the monetary base is less than –0.8 for the period 1970–2010. This suggests (but does not prove) that there is a close two-way contemporaneous negative relationship between the two variables.

[7] We discuss the case with four endogenous variables ($M = 4$).

$$Q_{23}(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then, we define the following:

$$Q = Q_{12}(\theta_1)Q_{13}(\theta_2)Q_{14}(\theta_3)Q_{23}(\theta_4)Q_{24}(\theta_5)Q_{34}(\theta_6).$$

The parameters $\theta_m, m = 1,...,6$ are drawn randomly from a uniform distribution $U(0,\pi)$, as in Peersman (2005). Note that the order of variables implied by the assumption that A_t is lower triangular is not preserved by the Givens transform—the recursive assumption is a special case.

For each vector $\Theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$, we obtain a set of impulse responses. To compute an impulse response, we take the estimated parameters of a reduced-form model from the relevant quarter. More specifically, for the computation of the impulse response for quarter *t*, we use the estimated standard deviations and the matrix of contemporaneous effects from quarter *t* and the coefficients at lags of variables from quarters *t+1*, *t+2*,…, $t + T_{HOR}$, where T_{HOR} denotes the number of quarters for which the impulse responses are computed. In the case where $t + T_{HOR} > T$, the coefficients from the last quarter are used (which is a reasonable assumption provided we assume that coefficients of lagged variables follow a random walk).

The computation of impulse responses is worth several comments. The usual approach, including in the TVP-VAR literature, is to take the estimated coefficients as of the quarter *t* for all quarters involved in the computation of impulse responses. There are two reasons for this approach. First, for the quarters at the end of the sample, the responses cannot be computed using estimates of parameters for the whole horizon without an assumption on the evolution of the system for $t = T+1, T+2,...$; otherwise, we lose several observations that are often of the greatest interest. Second, the usual argument is that estimated parameters do not change much over time and, thus, the assumption of a static system is a reasonable approximation. Nevertheless, because one of our crucial issues is the zero lower bound on the interest rate, we need to use as much available information about the state of the system as possible.

In other words, the usual identification scheme based on the Cholesky decomposition of the variance–covariance matrix imposes an inertial restriction only on the quarter in which the shock occurs. Sign restrictions assume something about the future of the system and, thus, the future should be modeled appropriately.

4.2 Imposed restrictions and the zero lower bound

Several attempts have been made in the literature to explicitly account for the zero lower bound on the call rate—a monetary policy instrument used by the BoJ. Iwata and Wu (2006) set up a censored VAR, where the latent variable captures the monetary policy stance and equals the interest rate if it exceeds zero. The latent variable is used to identify monetary policy shocks even when it attains values of less than zero. Kamada and Sugo (2006) implement the zero lower bound by setting up a monetary policy proxy that depends on a lending rate and the lending attitude of financial institutions. In such a case, the problem of a bound for the call rate seems to be shifted to the problem of a bound for lending rates. Fujiwara (2006) estimates a Markov-switching VAR with two states to examine whether reaching the zero lower bound brings about a break in parameters, and to discuss the effects of such a break on the monetary policy transmission. Nakajima (2011a) estimates a TVP-VAR with nominal interest rates considered to be a censored variable.

To summarize, in the papers mentioned above, the zero lower bound is dealt with from the very beginning of the estimation—the variants of the reduced-form VAR are estimated with respect to the zero lower bound. In this paper, the zero lower bound is imposed at a later stage. The reduced-form TVP-VAR model is flexible enough to deal with nonlinearities arising near the zero lower bound and, thus, the model is estimated without any restriction. The only restriction is a data set with a call rate that is nonnegative, although close to zero.

The implementation of the zero lower bound for the call rate draws on the sign restrictions approach. In addition to sign restrictions on responses, we impose a restriction that the call rate can only move down a certain small amount in a certain number of periods, i.e., the rate cannot be negative. Thus, the zero lower bound is imposed not only for the ZIRP period but also for other periods when the interest rate is close to zero. Table 1 reports the benchmark restrictions imposed.

We distinguish four periods in relation to the monetary policy conducted in Japan. For all periods, we assume that, following an accommodative monetary policy shock, industrial production does not decrease immediately. Moreover, some persistence and a certain lag are assumed and, thus, we impose the nonnegativity constraint for two quarters. Following Kamada and Sugo (2006), we do not impose any restriction on inflation. The monetary base growth is assumed to be positive after an accommodative monetary policy shock and lasts one quarter.

The differences in the identification schemes are mainly driven by the restrictions on the call rate—the distance of the call rate from zero. During the QE period, the fall in the call rate cannot exceed one basis point. In the ZIRP period and the "very low call rate" period, the decrease cannot be larger than five and 50 basis points, respectively. The lower bounds approximately reflect the average value of the rate for the respective period and are prescribed for four quarters.

Table 1. Identification of monetary policy shock: Benchmark restrictions

Description	Period	Industrial production	Call rate	Monetary base growth
Quantitative easing	2001Q2–2006Q1	≥0 for 2Q	≤0 for 1Q ≥–0.01 for 4Q	>0 for 1Q
ZIRP (call rate operating target)	1999Q2–2000Q3	≥0 for 2Q	≤0 for 2Q ≥–0.05 for 4Q	>0 for 1Q
Very low call rate	1995Q1–1999Q1 2000Q4–2001Q1 2006Q2–2010Q3	≥0 for 2Q	≤0 for 2Q ≥–0.5 for 4Q	>0 for 1Q
Other	1981Q1–1994Q4	≥0 for 2Q	≤0 for 2Q	>0 for 1Q

Three of the considered periods are characterized by the call rate as the operating target. During the QE policy period, the outstanding current account balances at the BoJ served as the operating target. The distinction is materialized in the duration of the nonpositive constraint on the call rate. We assume that interest rate smoothing is present in the rule followed by the central bank and, thus, the accommodative shock lasts two quarters when the call rate is the operating target and one quarter when the central bank targets outstanding balances.

More precisely, we illustrate the issue using the following simple money market model (the model is similar to that of Iwata and Wu, 2006):

$$R_t = \beta'_1 \varepsilon_t^Y + \beta_2 \varepsilon_t^s + \beta_3 \varepsilon_t^d$$

$$\Delta m_t = \alpha'_1 \varepsilon_t^Y - \alpha_2 \varepsilon_t^s + \alpha_3 \varepsilon_t^d ,$$

where R_t denotes the call rate and ε_t^Y , ε_t^s and ε_t^d represent a vector of innovations of macroeconomic variables, an exogenous monetary policy shock and an exogenous money demand shock, respectively. If the call rate is the policy instrument, then $\beta_3 = 0$. If a central bank is smoothing the interest rate, a supply shock affects the call rate for a period longer than a quarter. Next, in the case where the BoJ targets the monetary base growth Δm_t ($\alpha_3 = 0$), there is no reason for a gradual change of the interest rate and the prescribed restriction lasts for one quarter only.

The transmission channel of the QE policy works through the effect of the increase in the current account balances on long-term interest rates. The lowered spreads then provide a boost to economic activity. Even though long-term rates and spreads are not included in the set of endogenous variables, a monetary policy shock can still be identified through the effect on the monetary base, the call rate and industrial production. The shock is identified as an exogenous movement in endogenous variables such that the monetary base growth increases, the call rate decreases and industrial production increases (for two periods). The possibility of inertia in the industrial production and call rate is included, i.e., a change in the monetary base that leaves the industrial production and call rate unaffected satisfies the set of sign restrictions.

In addition to the change in the target of current account balances, other aspects of the QE policy mentioned in Section 2 can be identified as monetary policy shocks. First, the announcement of the commitment to pursue the policy until deflationary concerns disappear can affect endogenous variables in the directions satisfying the sign restrictions. Similarly, the pure credit easing conducted by the BoJ in the early 2000s can be identified as a monetary policy shock. However, because the structural shocks capture unexplained variation of the model variables, particular sources of a shock cannot be identified separately.

Finally, it is necessary to check whether other types of shock possibly captured by the endogenous variables do not satisfy the sign restrictions prescribed for identification of a monetary policy shock or whether the signs are exactly the opposite. A demand shock implies co-movement of the industrial production and call rate in the same direction and, thus, is distinguishable from the monetary policy shock that leads to the opposite reaction of the two endogenous variables. It is assumed that a supply shock does not affect the monetary base growth immediately when a central bank targets the monetary base and, thus, it is also distinguishable from a monetary policy shock, which affects the monetary base immediately after the shock occurs.

5. Results

Based on the previous literature (mainly Miyao, 2002, and Kimura et al., 2003), four endogenous variables are included in the analysis: the index of industrial production (=10 in 2005), CPI inflation, the call rate and the growth rate of the monetary base.[8] Following Miyao (2005), the call rate series before 1985 is constructed using the collateralized overnight rate, adjusted by the mean difference between the collateralized and uncollateralized overnight rates after 1985. We take quarterly averages.

The data are quarterly, covering the period 1971Q1–2010Q3. The sample 1971Q1–1980Q4 is used to calibrate priors for initial states of coefficients and hyperparameters; the resulting sample covers the period 1981Q1–2010Q3.

[8] Sources: The index of industrial production is taken from the Ministry of Economy, Trade and Industry, the CPI from the Official Statistics of Japan and the call rate and monetary base from the Bank of Japan. The monetary base comprises average amounts outstanding, adjusted for the reserve requirement ratio changes. CPI is the core index adjusted for consumption tax. Data are seasonally adjusted.

The model and additional procedures are coded in Matlab.[9] Two lags of the endogenous variables are used, as is usual in the literature.[10] Based on the convergence diagnostics presented in Appendix B, a sample of 7,500 iterations of the Gibbs sampler is used, discarding the first 1,000 draws for convergence. The implementation of the identifying restrictions is as follows. For each draw (iteration) from the conditional posteriors of coefficients and volatilities, a draw from the uniform distribution is taken to generate a Givens rotation and the two are combined to compute the impulse responses. The procedure draws from the uniform distribution until the combined draw generates impulse responses satisfying the sign restrictions and a lower bound constraint (if imposed), or if more than 40 draws from the uniform distribution are made. The accepted draw is then saved. The procedure moves to the next draw from the conditional posteriors of parameters. After generating the set of impulse responses, those that satisfy the prescribed restrictions are taken and a summary statistics is produced.

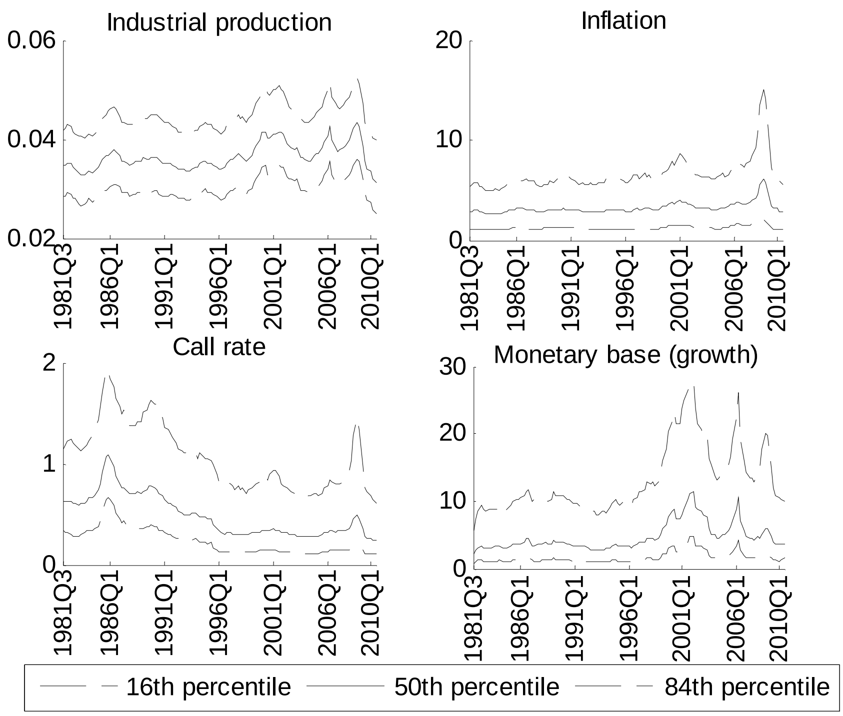

Figure 1. Posterior estimates of the standard deviations of reduced-form residuals.

[9] The estimation of the reduced-form TVP-VAR is based on the code by Gary Koop and Dimitris Korobilis available at http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html. See Koop and Korobilis (2010) for details.

[10] One lag is not enough to capture the dynamics of the system. On the other hand, three lags imply a large number of additional parameters to estimate.

The discussion of the results starts with the estimation of the volatility of reduced-form innovations. The model allows for time variation in the volatility and, thus, the changing impact of exogenous factors can be investigated. Next, the monetary policy shock is identified. As both operating targets are involved in the analysis, the monetary policy shock can be viewed as a vector in the two-dimensional system: an identified shock is a combination of a shock to the monetary base and the call rate. On the other hand, direct comparison of shock sizes is difficult because of the incomparability of the two operating targets. The discussion of the results then proceeds to the transmission mechanism, i.e., the way an identified shock affects endogenous variables and how the transmission mechanism changes over time.

Figure 1 shows the selected quantiles of posterior estimates of the reduced-form residual standard deviations. The figure can serve as a first check of the reduced-form model settings, for example, whether the prior time-variation imposed on coefficients is reasonable and does not falsely account for exogenous shocks. Figure 1 suggests, for example, that, since 2008Q1, the volatility of shocks to inflation has increased remarkably. The rise in inflation volatility has reflected the global oil price volatility. During the QE period, exogenous shocks to the monetary base growth played an important role. In addition to monetary policy, other factors (the Year 2000 Problem, the Pay-off Problem) could contribute to the rise in the variance of the monetary base. Figure 1 reports the volatility of reduced-form innovations, which need not necessarily say anything about the structural shocks.

Figure 2 focuses on a structural shock—an accommodative monetary policy shock—identified by the restrictions on impulse responses described in Table 1, i.e., the structural monetary policy shock is defined as an unexplained positive movement of industrial production, the call rate and the monetary base for a certain number of periods. The figure shows the posterior of contemporaneous effects of the one-standard-deviation monetary policy shock on endogenous variables.

First, note that, in contrast to the recursive assumption, the monetary policy shock cannot be related directly to the variation of a particular endogenous variable because the operating target of the BoJ changed in 2001 and 2006. The change is indicated by the vertical lines. The monetary policy shocks are reflected mainly in the monetary base growth during the QE period (2001–2006), whereas other periods are characterized by the call rate as an indicator of exogenous movements in the monetary policy. Second, note that the monetary policy shock $\left(\varepsilon_t^i\right)$ does not affect shocks to contemporaneous relations (v_t^i) because the variance–covariance matrix of all model innovations is assumed to be block diagonal. The interpretation of a structural shock as an exogenous change uncorrelated with other innovations thus remains.

Figure 2 depicts nonsystematic monetary policy. Such shocks can be the result of, for example, setting monetary policy rates differently from the rule, the effect of variables exogenous to our system[11] or announcements by the BoJ. The figure suggests that, during the QE period, monetary policy shocks affect the monetary base growth but not the call rate. For the call rate, the imposed lower bounds are clearly visible. Furthermore, there is some evidence of the positive contemporaneous effect of monetary policy shocks to inflation during the QE period that could be a consequence of the announcements regarding the duration of loose monetary policy. Note that, with the exception of inflation, the response of the other variables is partly driven by the imposed restrictions.

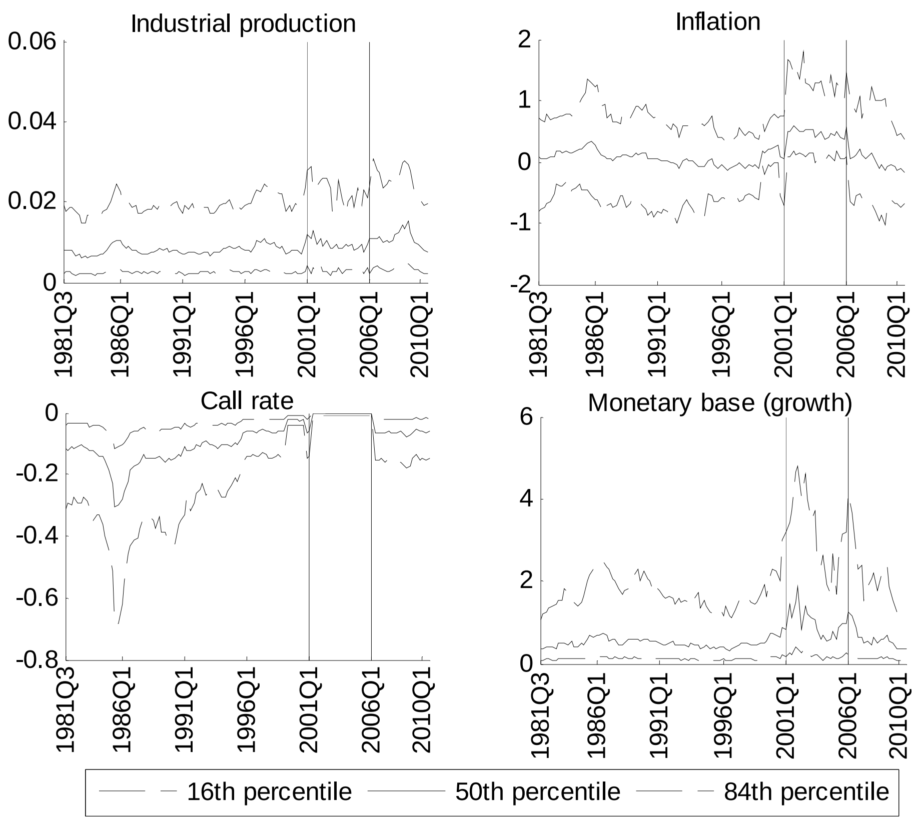

Figure 2. Contemporaneous effects of an accommodative monetary policy shock. Note: Estimation based on 2,000 draws with first 500 discarded.

The effect of nonsystematic monetary policy is captured by impulse response functions. In the setting with time-varying parameters, responses differ for each quarter. We describe the transmission mechanism from several points of view. First, we take an overall perspective, looking at the responses in 1986Q1, 1994Q1 (after the asset bubble burst) and 2002Q1 (the QE

[11] For example, the system does not include the exchange rate, which played a significant role in some periods. Computational reasons force us to have a system with four endogenous variables only.

period).[12] Second, we focus in detail on the ZIRP and QE policy by examining the responses in 2000Q1, 2002Q1 and 2003Q4. The latter two dates are chosen in order to investigate the effect of the BoJ's commitment to pursue QE policy until CPI inflation reached zero or increased in year-on-year terms. This announcement was made in 2003Q4, and we compare the transmission of the shock with 2002Q1. Finally, we discuss changes in the transmission mechanism during the recent economic crisis, focusing on responses in 2007Q1, 2008Q1 and 2009Q1.

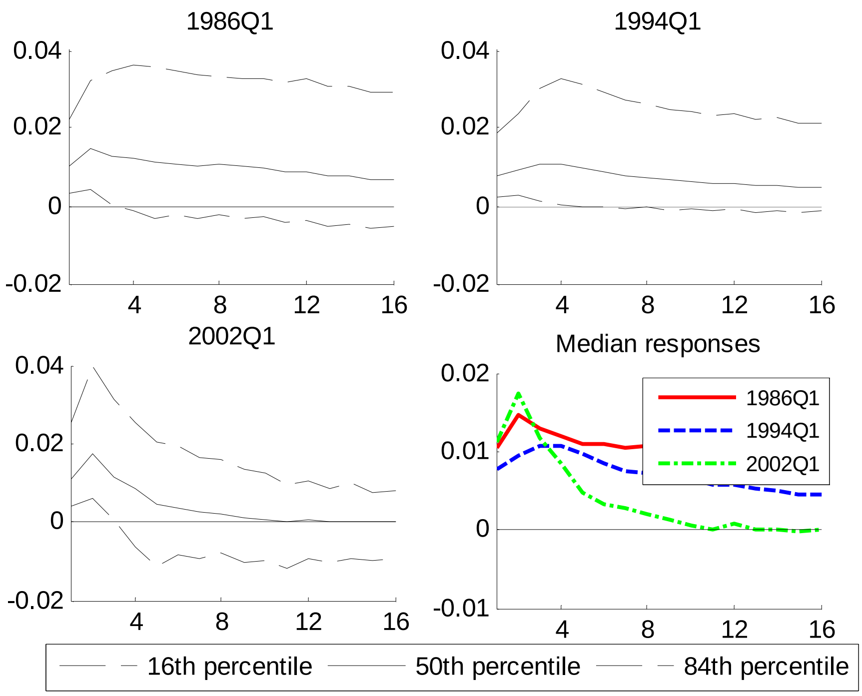

Figure 3. The effect of a monetary policy shock on industrial production in 1986Q1, 1994Q1 and 2002Q1, and the respective median responses.

Impulse responses reflect the effect of a shock identified by sign restrictions and impose lower bounds on responses for a particular period. The normalization of responses makes little sense when there is a zero lower bound and other possible nonlinearities. The interpretation of the size of responses should draw on the effect of the shock on the operating target currently in use. The uncertainty is captured by the selected percentiles of the posterior distribution of responses; this is discussed in detail in the next section.

Figure 3 suggests that industrial production responds similarly to a monetary policy shock in the first year after the shocks in 1986Q1, 1994Q1 and 2002Q1. The exception is the behavior of the index of industrial production after a year for 2002Q1. Monetary policy shocks seem not to be

[12] The same dates are investigated by Nakajima et al. (2009), so we retain these dates to allow for direct comparison.

effective in avoiding a fall in real output during the QE period, i.e., the increase in the monetary base did not stimulate the real economy for more than a few quarters. Such a result is in accordance with the literature (see Ugai, 2007). In 1986Q1 and 1994Q1, the standard profile of industrial production, with a peak after approximately one year, is observable.

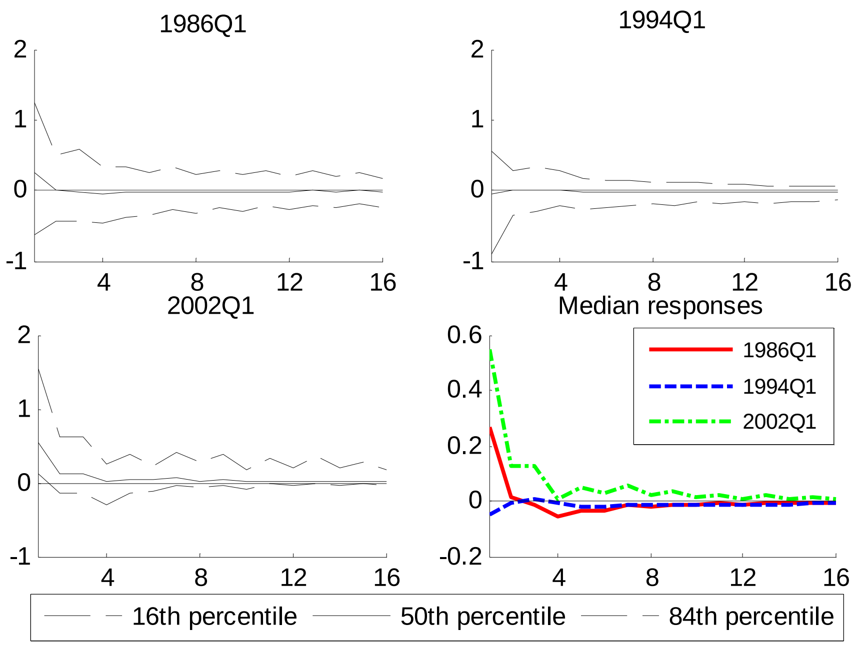

Figure 4. The effect of a monetary policy shock on inflation in 1986Q1, 1994Q1 and 2002Q1, and the respective median responses.

Regarding inflation, the QE period is characterized by the positive effect of a monetary policy shock (Figure 4c). In contrast, Kimura et al. (2003) and Nakajima et al. (2009) do not find any strong effect of the unexpected monetary base growth on inflation. The conclusion in Kimura et al. (2003) could be a consequence of the constant volatility assumption. As noted by Primiceri (2005), such an assumption implies constant contemporaneous relations between variables, which could lead to the underestimation of the effect. Although Nakajima et al. (2009) allow for the changing volatility, their finding could be driven by the recursive assumption: they assume inertia of inflation and the call rate in the quarter of the unexpected monetary base rise. Regarding the other dates, inflation responds negatively to the shock in 1994Q1 and positively in 1984Q1. However, the middle 68 percent of the posterior distribution contains zero, so we conclude that the effect of an accommodative monetary policy shock on inflation is very weak in 1986Q1 and 1994Q1.

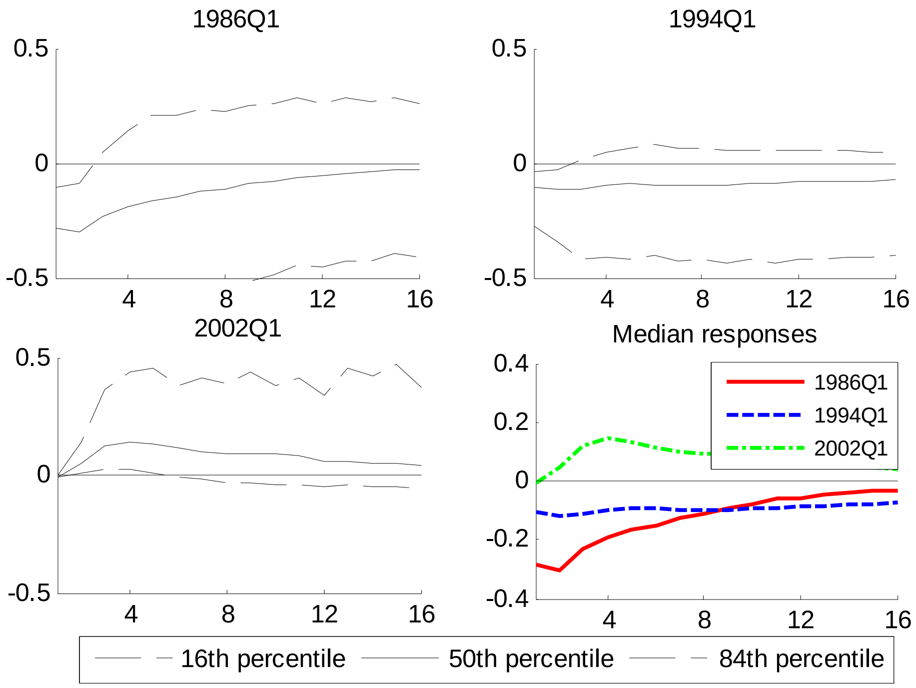

Figure 5. The effect of a monetary policy shock on the call rate in 1986Q1, 1994Q1 and 2002Q1, and the respective median responses.

To compare the size of the monetary policy shock in a particular quarter, Figures 5 and 6 present the responses of the call rate and the monetary base growth to a monetary policy shock. The call rate was the operating target for the first two dates, whereas the balances of current accounts were for the third date. Monetary policy shocks bring about negative/positive movements of the operating targets. Figure 5 also demonstrates the imposed lower bound on the call rate in 2002Q1 when the rate was virtually zero. Note that the positive median response of the call rate in 2002Q1 is a consequence of the constraint on the call rate, rather than a profile suggested by data. The distribution of the response is forced into a narrow interval above zero and, thus, the median is also above zero. Finally, it is not surprising that the greatest impact of the monetary policy shock on the monetary base growth occurred during the QE period, in 2002Q1 (Figure 6).

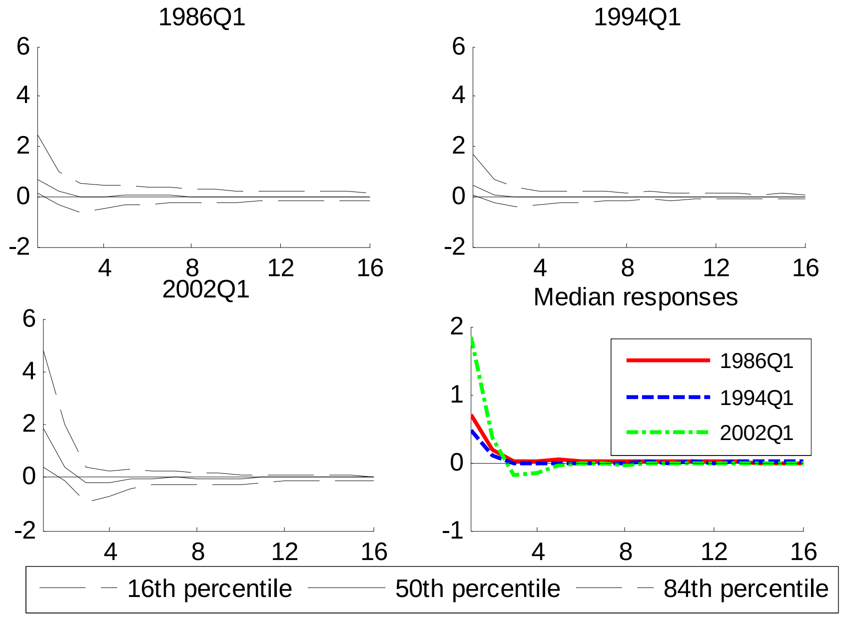

Figure 6. The effect of a monetary policy shock on monetary base growth in 1986Q1, 1994Q1 and 2002Q1, and the respective median responses.

Figure 7 captures the ZIRP and QE periods. It reflects the change of the operating target and the zero lower bound on the call rate, as discussed above. In a comparison of the effects of an unexpected shock for 2002Q1 and 2003Q4 (when the announcement about the policy commitment was made), no major difference is detected. The effect of the announcement is not revealed by comparing the responses between the two periods, i.e., the effect does not work through the way the monetary policy shock is transmitted by the endogenous variables considered in the model. The effect, however, need not necessarily be reflected by changes in the transmission mechanism. It can be identified as a monetary policy shock by the model and is thus reflected in the size of the shock. In such a case, the model is not able to distinguish between the effect of the announcement and other sources of shocks and, thus, it cannot evaluate the effectiveness of the commitment itself.

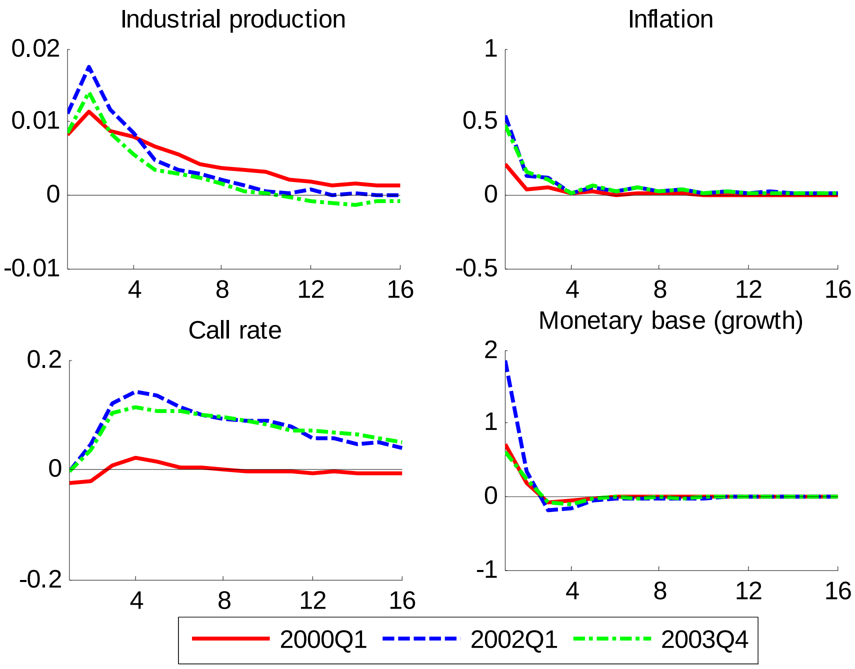

Figure 7. Median responses of endogenous variables to the monetary policy shock for 2000Q1, 2002Q1 and 2003Q4.

Finally, Figure 8 compares the effects of a monetary policy shock for 2007Q1, 2008Q1 and 2009Q1. It suggests that the transmission mechanism has not changed during the recent crisis. The negative impact of the shock on inflation is not relevant because the distribution of the response is spread around zero almost symmetrically. The figure suggests proper functioning of the transmission mechanism, reflected by the fact that the relative magnitudes of responses correspond to the relative magnitudes of the call rate changes.

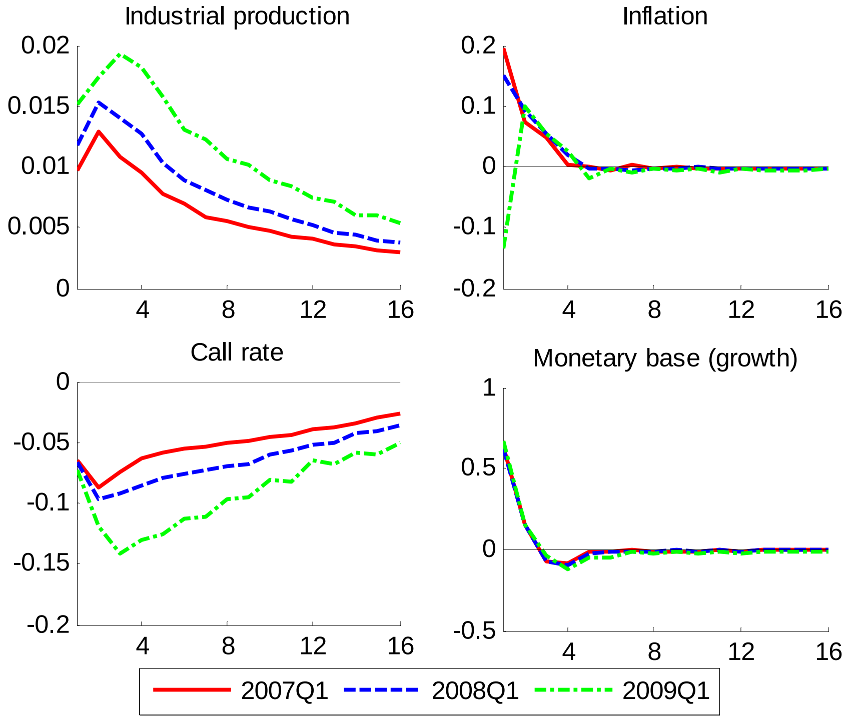

Figure 8. Median responses of endogenous variables to the monetary policy shock for 2007Q1, 2008Q1 and 2009Q1.

5.1 Several caveats

The use of sign restrictions remains a controversial issue. In particular, Fry and Pagan (2007) address some very important points. They assert that the median of the posterior of impulse responses identified by sign restrictions summarizes information *across models.*[13] Every Givens rotation generating a set of impulse responses represents just one underlying structural model. To attempt to capture the uncertainty added by the averaging of the effects of a shock across models, Figure 9 presents the responses of an accommodative monetary policy shock in 1994Q1 identified by sign restrictions and by the recursive assumption. The recursive assumption determines the model underlying the reduced-form TVP-VAR uniquely, i.e., the responses do not include uncertainty related to different structural models. One can observe that the increase in uncertainty when using sign restrictions is not large (it is highest for industrial production).

[13] Uhlig (2005) avoids the problem of mixing responses from different models by taking the response (model) resulting from a minimization problem. However, the ad hoc selection of the function to minimize is also subject to criticism.

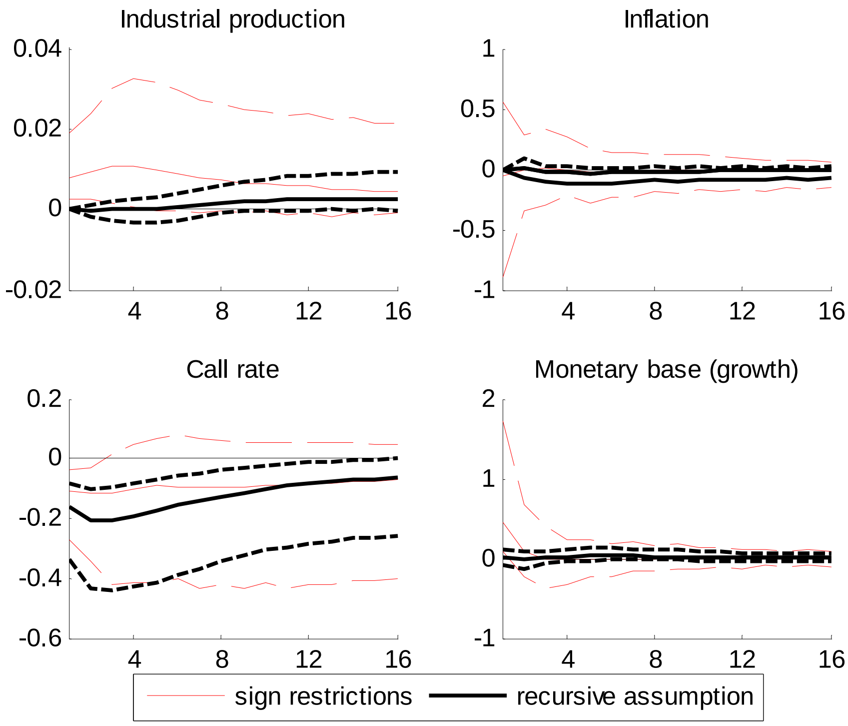

Figure 9. The comparison of responses for two identification schemes, 1994Q1.

Another issue related to the problem of many models satisfying sign restrictions is that a particular median response consists of the responses based on different models. Fry and Pagan (2007) argue that median responses provide "information about shocks coming from different models". This can cause a problem when, for example, the assumption of uncorrelated shocks is necessary, as in the variance decomposition exercise. The problem is not as serious in our case, because we take the agnostic approach and identify one shock by imposing a sign on the responses of (not necessarily all) endogenous variables. For our purposes, it suffices to keep in mind the two sources of uncertainty involved in our results.

Fry and Pagan (2007) suggest presenting the impulse response that is the closest (in terms of the sum of squares of the normalized response) to the median of impulse responses for all endogenous variables to guarantee the existence of one model producing such responses. The practical issue we encountered when implementing the suggested procedure is the high sensitivity of the resulting responses. Even for a large number of iterations of the Gibbs sampler, responses can differ for different runs of the estimation procedure (Figure 10). We therefore retain the original approach.[14]

[14] Note that different runs of the estimation procedure do not affect the median response, as computed in a standard way.

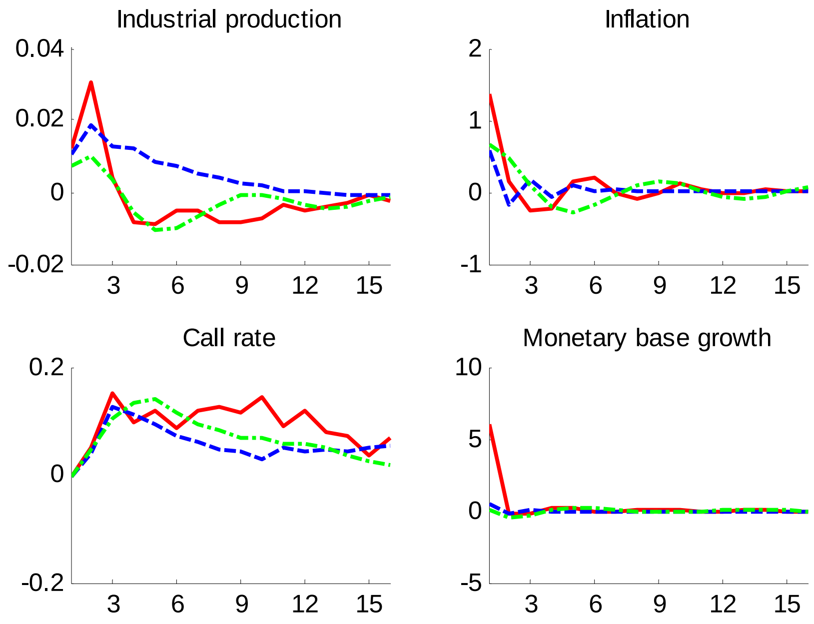

Figure 10. Sensitivity of Fry and Pagan's (2007) responses, 2002Q1.

Finally, the last point of criticism draws on Paustian (2006), who shows that sign restrictions identify a structural shock correctly only if a sufficient number of restrictions is imposed and structural shocks are large enough. In this paper, restrictions are imposed on three out of four variables. Moreover, the size of the monetary policy shocks seems to be large, especially in the QE period (Figure 2).

5.2 Robustness analysis

We start with a sensitivity analysis focusing on the number of lags. Impulse responses are very similar for three lags of endogenous variables. If we estimate the system with just one lag of endogenous variables, the relative relationship of the responses between the considered dates does not change. The magnitudes (even the sign of the response), however, often differ. One lag is not sufficient to capture the dynamics of endogenous variables and, thus, two lags are viewed as a minimum to obtain reliable results.

Regarding the priors, the choice of the prior for the hyperparameter U is crucial, as noted by Primiceri (2005). We chose a slightly tighter prior because our system contains four endogenous variables, in comparison with the three-variable system of Primiceri. This is in accordance with Nakajima et al. (2009), who also estimate a four-variable system. Primiceri's prior on U leads to problems with convergence that are not resolved even if we increase the number of runs of the

MCMC algorithm considerably. This fact is not surprising because U contains 1,296 parameters, and a tighter prior seems to be the only way to avoid unrealistic estimates of the parameters. The overall prior information imposed was tested by using uninformative priors. The responses of the monetary base growth and the call rate do not change. The response of inflation and industrial production involve greater uncertainty and their results are not very conclusive.

Finally, the effect of restrictions on the responses is worth discussing. First, the restrictions are based on assumptions about the behavior of the system, and high sensitivity to changes in restrictions is expected. On the other hand, some assumptions—the duration of restrictions, for example—do not necessarily change the results much. We tested several sets of restrictions that differ in terms of the decision regarding whether to apply any restriction at all (e.g., on inflation). Such assumptions turned out to be crucial for the resulting responses. In the previous section, therefore, we provided a clear reasoning for the concrete set of restrictions used. The reasoning is based on economic theory and the previous literature employing sign restrictions for Japan.

6. Conclusion

To analyze monetary policy in Japan during the past three decades, the paper extends the methodology of TVP-VARs to the identification of shocks based on sign restrictions. Two points related to this extension should be highlighted. First, it is important to compute impulse responses with regard to the actual movements of the system, i.e., using time-varying parameters estimated for dates involved in the computation of the response. Assuming a static system could lead to biased results. Second, the zero lower bound on the nominal interest rate is introduced in a very intuitive way. In comparison with other approaches that are currently being explored, the lower bound is implemented at the stage of identification of the shocks, not during the estimation of the reduced-form version of the model.

In this paper, the pros and cons of sign restrictions implemented in a TVP-VAR setting are clearly stated and, thus, the paper contributes to the current debates on using sign restrictions in applied macroeconomics. Moreover, the paper provides additional evidence of the effectiveness of monetary policy conducted in Japan during parts of the past three decades.

References

Baumeister, C. and L. Benati (2010). Unconventional Monetary Policy and the Great Recession. Estimating the Impact of a Compression in the Yield Spread at the Zero Lower Bound. European Central Bank Working Paper Series, 1258.

Canova, F. and G. de Nicolo (2002). Monetary disturbances matter for business fluctuations in the G7. *Journal of Monetary Economics*, 49, 1131–1159.

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

Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5, Ch. 57, Amsterdam, Netherlands: Elsevier, 3569–3649.

Cogley, T. and T.J. Sargent (2005). Drift and volatilities: Monetary policies and outcomes on the post WWII US. *Review of Economic Studies*, 8, 262–302.

Darvas, Z. (2009). Monetary Transmission in Three Central European Economies: Evidence From Time-Varying Coefficient Vector Autoregressions. De Nederlandsche Bank Working Paper Series, 208.

Faust, J. (1998). The robustness of identified VAR conclusions about money. *Carnegie–Rochester Series on Public Policy* 49, 207–224.

Fry, R. and A. Pagan (2007). Some Issues in Using Sign Restrictions for Identifying Structural VARS. National Centre for Economic Research Working Paper Series, 14.

Fujiki, H. and S. Shiratsuka (2002). Policy duration effect under the zero interest rate policy in 1999–2000: Evidence from Japan’s money market data. *Monetary and Economic Studies*, 20(1), 1–31.

Fujiwara, I. (2006). Evaluating monetary policy when nominal interest rates are almost zero. *Journal of Japanese and International Economies*, 20, 434–454.

Iwata, S. and S. Wu (2006). Estimating monetary policy effects when interest rates are close to zero. *Journal of Monetary Economics*, 53, 1395–1408.

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

Kamada, K. and T. Sugo (2006). Evaluating Japanese Monetary Policy under the Non-Negativity Constraint on Nominal Short-Term Interest Rates. Bank of Japan Working Paper Series, 06–E–17.

Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65, 361–393.

Kimura, T., H. Kobayashi, J. Muranaga and H. Ugai (2003). The effect of the increase in the monetary base of Japan's economy at zero interest rates: An empirical analysis. In: Bank of International Settlements (ed.), *Monetary Policy in a Changing Environment*, vol. 19, Bank for International Settlements, Basel, 276–312.

Kirchner, M., J. Cimadomo and S. Hauptmeier (2010). Transmission of Government Spending Shocks in the Euro Area: Time Variation and Driving Forces. Tinbergen Institute Discussion Paper, TI 2010–021/2.

Koop, G. and D. Korobilis (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. Rimini Centre for Economic Analysis, mimeo.

LeSage, J.P. (1999). *Applied Econometrics Using MATLAB*. Toledo: University of Toledo.

Miyao, R. (2002). The effects of monetary policy in Japan. *Journal of Money, Credit and Banking*, 34, 376–392.

Miyao, R. (2005). Use of the money supply in the conduct of Japan's monetary policy: Re-examining the time-series evidence. *The Japanese Economic Review*, 56(2), 165–187.

Nakajima, J. (2011a). Monetary Policy Transmission under Zero Interest Rates: An Extended Time-Varying Parameter Vector Autoregression Approach. Institute for Monetary and Economic Studies Discussion Paper Series, 2011-E-8.

Nakajima, J. (2011b). Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications. Institute for Monetary and Economic Studies Discussion Paper Series, 2011-E-9.

Nakajima, J., M. Kasuya and T. Watanabe (2009). Bayesian Analysis of Time-Varying Parameter Vector Autoregressive Model for the Japanese Economy and Monetary Policy. Institute for Monetary and Economic Studies Discussion Paper Series, 2009-E-13.

Paustian, M.O. (2006). When Do Sign Restrictions Work? Available at SSRN: http://ssrn.com/abstract=932267.

Peersman, G. (2005). What caused the early millennium slowdown? Evidence based on autoregressions. *Journal of Applied Econometrics*, 20, 185–207.

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

Raftery, A.E. and S. Lewis (1992). How many iterations in the Gibbs sampler? In: J.M. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics*, Oxford: Oxford University Press, 297–346.

Shirakawa, M. (2010). Uniqueness or similarity? Japan's post-bubble experience in monetary policy studies. Keynote address at second International Journal of Central Banking Fall Conference, Institute for Monetary and Economic Studies, Bank of Japan, Tokyo.

Shiratsuka, S. (2010). Size and composition of the central bank balance sheet: Revisiting Japan's experience of the quantitative easing policy. *Monetary and Economic Studies*, Institute for Monetary and Economic Studies, Bank of Japan, 28, 79–106.

Smith, M. and R. Kohn (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97, 1141–1153.

Ugai, H. (2007). Effects of the quantitative easing policy: A survey of empirical analyses. *Monetary and Economic Studies*, 25(1), 1–48.

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

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard and Pitt (1997). *Biometrika,* 91, 246–248.

Appendix A: Sampling algorithm

Our TVP-VAR is estimated using the Gibbs sampler, which produces a sample from the joint posterior distribution of the parameters. The basic steps consist of drawing lagged effects coefficients (β_t), contemporaneous relations coefficients (α_t), volatilities (σ_t) and hyperparameters (*U*, *V* and *W*). The steps are discussed in turn. The algorithm largely follows the one described in Primiceri (2005). Some minor changes are based on Cogley and Sargent (2005).

A.1 Drawing lagged effects coefficients

Given the data, contemporaneous relations, volatilities and hyperparameters, equation (2) is linear (in β_ts), with Gaussian innovations with a known covariance matrix. Together with the state equation (3), it constitutes a state space form that can be estimated. Thus, the Kalman filter and a backward recursion are used to sample the lagged effects coefficients. The algorithm of Carter and Kohn (1994) is employed.

A.2 Drawing contemporaneous relations coefficients

Equation (2) can be rewritten as:

$$A_t\left(y_t - X_t\beta_t\right) = \Sigma_t\varepsilon_t .$$

Given the data and lagged effects coefficients, $y_t - X_t\beta_t$ is observable. Moreover, as A_t is a lower triangular matrix with the ones on the main diagonal, we obtain:

$$\tilde{y}_t \equiv y_t - X_t\beta_t = Z_t\alpha_t + \Sigma_t\varepsilon_t ,$$

where

$$Z_t = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -\tilde{y}_{1,t} & 0 & \cdots & 0 \\ 0 & -\tilde{y}_{[1,2],t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\tilde{y}_{[1,\dots,n],t} \end{bmatrix},$$

where $\tilde{y}_{[1,\dots,i],t}$ is the row vector $\left[\tilde{y}_{1,t},\dots,\tilde{y}_{i,t}\right]$. Assuming the block diagonal structure of *V*, the structure of Z_t allows the application of the Kalman filter and the backward recursion equation by equation using the algorithm of Carter and Kohn (1994) (the innovations are Gaussian). Note that (4) represents the state equation of the system.

A.3 Drawing volatilities

To draw volatilities two approaches are taken: the algorithm used in Primiceri (2005) and that used in Cogley and Sargent (2005). Cogley and Sargent (2005) assume the diagonality of the hyperparameter *W*.

Given the data, contemporaneous relations coefficients and lagged effects coefficients, $A_t\left(y_t - X_t\beta_t\right)$ is observable and (2) can be rewritten as follows:

$$\hat{y}_t \equiv A_t\left(y_t - X_t\beta_t\right) = \Sigma_t\varepsilon_t\,. \tag{A1}$$

The nonlinearity of the system can be dealt with by taking squares and the logarithm of every element of the system. Following Primiceri (2005), we add an offset constant $\bar{c} = 0.001$ to $\hat{y}_t$ to avoid the term $A_t^2\left(y_t - X_t\beta_t\right)^2$ being equal to zero and, thus, the nonexistence of the logarithm. That is, we obtain:

$$\log\left(\hat{y}_{i,t}^2 + \bar{c}\right) = 2\log\left(\sigma_{i,t}\right) + \log\left(\varepsilon_{i,t}^2\right). \tag{A2}$$

The resulting system (with the state equation (5)) takes the linear but non-Gaussian state space form: the innovations in the measurement equation, $e_{i,t} \equiv \log\left(\varepsilon_{i,t}^2\right)$, are distributed as $\log\chi^2(1)$. Volatilities are sampled for all variables jointly using the approximate multivariate state space model based on (A2) and (5). Kim et al. (1998) is followed to transform the system into a Gaussian one based on the mix of normal approximations of the $\log\chi^2$ distribution. Therefore, we use a mixture of seven normal densities, with the parameters (component probability, mean and variance) from Kim et al. (1998). The variance covariance matrix of innovations in (A2) is diagonal. Thus, the same independent mix of normal densities can be used to approximate elements of the innovations vector. The indicators of the normal densities used are sampled independently using the discrete density stated in Kim et al. (1998).[15]

To shed some light on the role of the ad hoc offset constant $\bar{c}$ used in Primiceri (2005), and also as a robustness exercise, the sampling of volatilities based on Jacquier et al. (1994) is tested. We follow the particular use of the algorithm by Cogley and Sargent (2005). Equation (A1) suggests that the orthogonalized residuals are observed. To apply the algorithm by Jacquier et al. (1994), the diagonality of *W* is assumed. Then, the univariate algorithm is applied element by element on the orthogonalized residuals.

Basically, the date-by-date blocking scheme is adopted to compute the following:

$$f\left(h_{it} \mid h_{-it}, \hat{y}_i^T, W_{ii}\right) = f\left(h_{it} \mid h_{it-1}, h_{it+1}, \hat{y}_i^T, W_{ii}\right). \tag{A3}$$

[15] An alternative to the mixture sampler introduced in Kim et al. (1998) is the block sampler introduced in Watanabe and Omori (2004). The block sampler is an exact method, but requires the assumption on the diagonality of *W*.

Jacquier et al. (1994) derive the formula for (A3) and suggest a Metropolis step instead of a Gibbs step, i.e., they employ a proposal density from which to take draws and compute the acceptance probability. The acceptance probability bound is set equal to 0.5 in this paper. The estimation algorithm using the procedure from Jacquier et al. (1994) generates very similar results to Primiceri's (2005) approach and, thus, we stick to the more general algorithm by Primiceri (2005).

A.4 Drawing hyperparameters

Given the data, contemporaneous relations coefficients, lagged effects coefficients and volatilities, the innovations are observable. Moreover, assuming priors on *U* and blocks of *V* and *W* being distributed as an inverse-Wishart distribution, the posterior distribution is also inverse-Wishart, and, therefore, drawing hyperparameters is straightforward.

In the case of the diagonality assumption on *W*, the situation is analogous: the diagonal elements have an inverse-gamma distribution.[16]

[16] When assuming the *W* diagonal, the prior is set to $W_{ii} \sim IGamma(5k_V^2, 5)$.

Appendix B: Convergence diagnostics[17]

The algorithm estimates 5,382 coefficients (six for contemporaneous effects, 36 for lagged effects + constants and four for standard deviations for each of the 117 periods) and 1,348 hyperparameters. Because of space constraints, we present the convergence statistics of coefficients for one quarter only (1994Q1) and all hyperparameters. Three statistics are presented to assess the convergence of the MCMC algorithm. First, autocorrelation of the chain at lag = 10 for hyperparameters is presented in Figure B1. Low autocorrelation suggests the independence of draws and, thus, the efficiency of the algorithm. Only a few hyperparameters related to the time-variation of the contemporaneous relations coefficients are highly correlated.

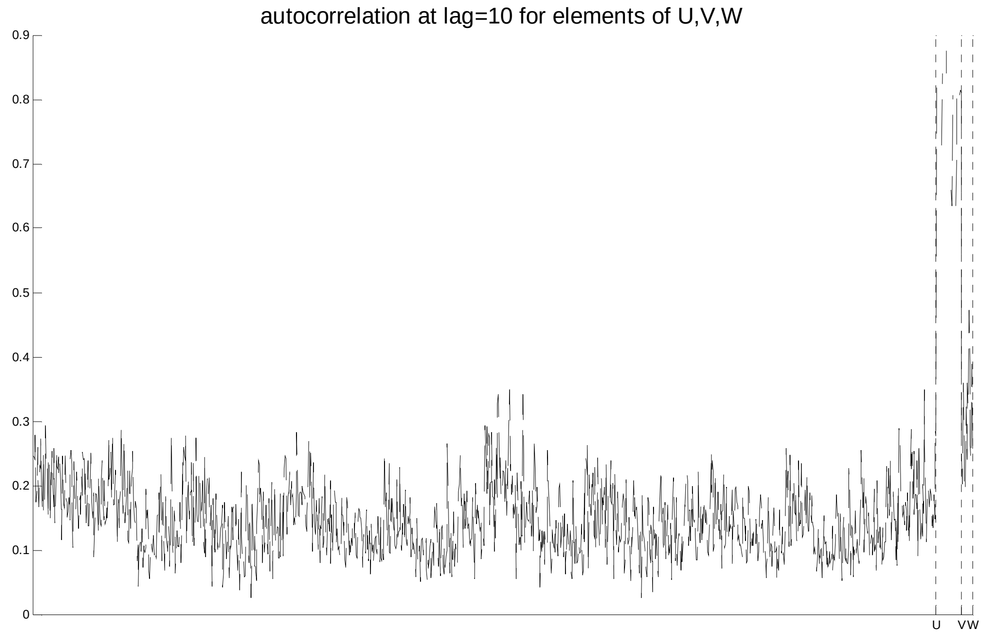

Figure B1. Autocorrelation of the chain for the hyperparameters.

Figures B2 and B3 present inefficiency factors, and Raftery and Lewis (1992) estimate the total number of runs necessary to achieve a certain precision. The inefficiency factor (Chib, 2001) is an estimate of $\left(1+2\sum_{k=1}^{\infty}\rho_k\right)$, where ρ_k is the k-th autocorrelation of the chain. It is the inverse of the relative numerical efficiency measure. The relative numerical efficiency suggests the number of draws needed to achieve the same numerical accuracy as if the draws were taken from an i.i.d.

[17] Codes from Econometrics Toolbox are used. For details, see LeSage (1999).

sample directly from the posterior distribution. The estimate is carried out using an estimate of the spectral density at frequency zero (4% tapered window). Values of inefficiency factors around and below 20 are viewed as satisfactory. Again, some problems are detected for hyperparameter *V*.

Finally, according to the Raftery and Lewis (1992) estimate, the number of runs suggested for hyperparameters is smaller than the number of runs used in the paper (7,500). The precision is tested for the 0.025 and 0.975 quantiles of marginal posteriors, with the minimum probability of achieving the precision set to 0.95 and the desired accuracy to 0.025.

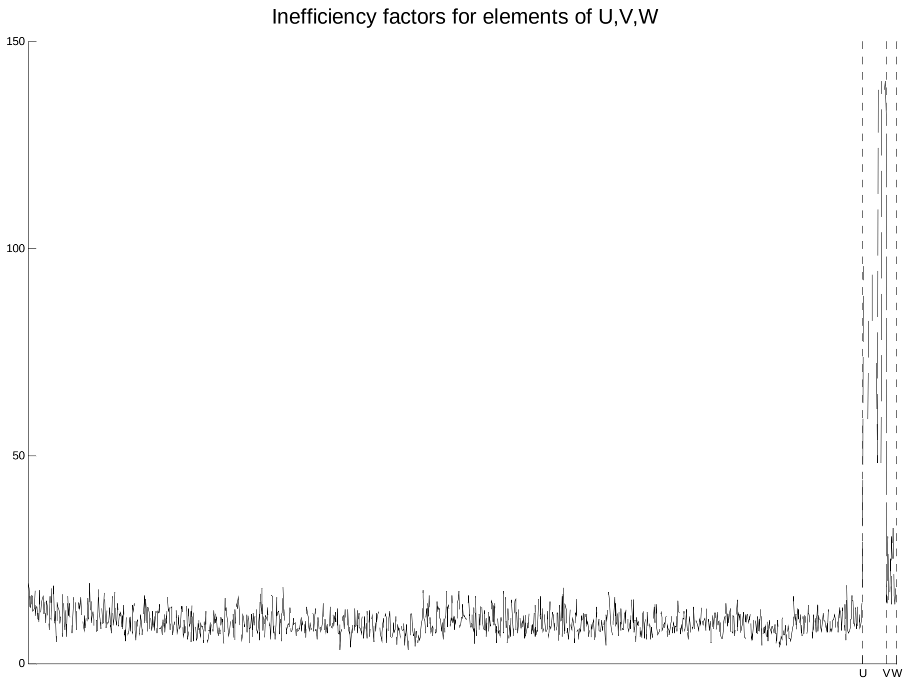

Figure B2. Inefficiency factors of the chain for the hyperparameters.

The same three statistics are presented for coefficients in Figures B3 and B4.

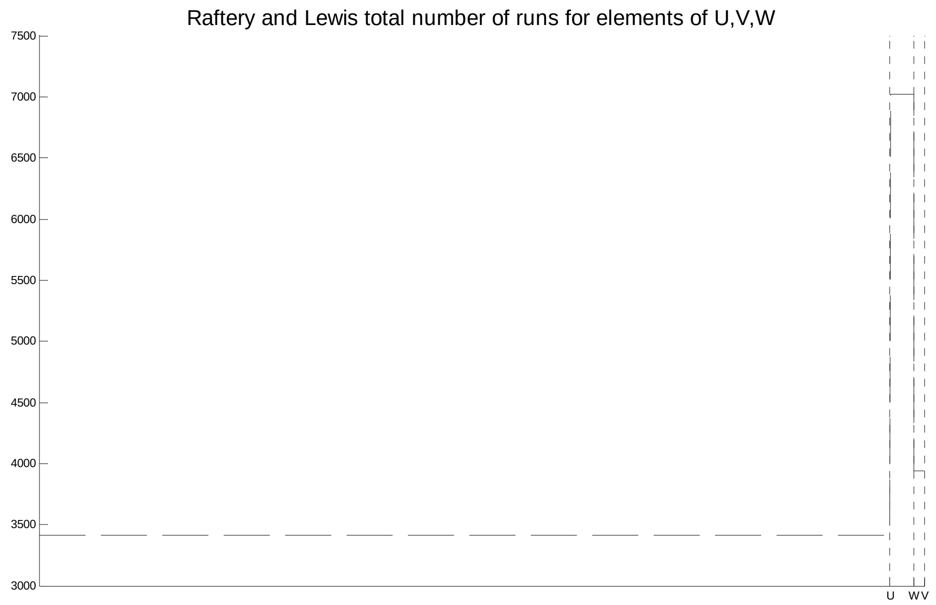

Figure B3. Raftery and Lewis statistics for the hyperparameters.

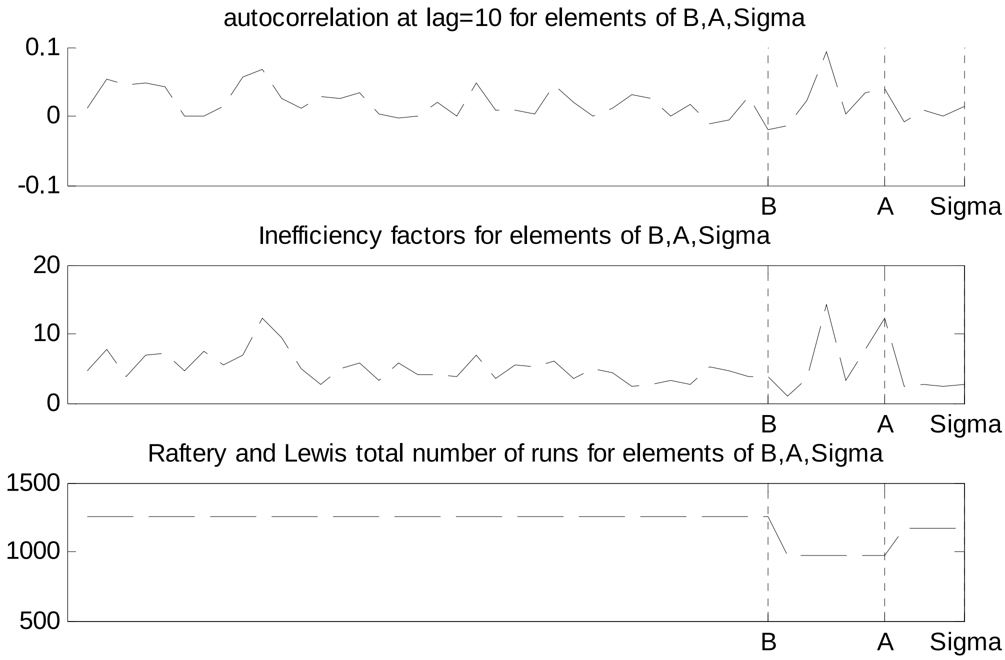

Figure B4. Autocorrelation, inefficiency factors and Raftery and Lewis statistics for coefficients, 1994Q1.