Working Paper Series

Small Scale Bayesian VAR Modeling of the Japanese Macro Economy Using the Posterior Information Criterion and Monte Carlo Experiments

Munehisa Kasuya* **and Tomoki Tanemura****

Working Paper 00-4

Research and Statistics Department

Bank of Japan

C.P.O BOX 203 TOKYO

100-8630 JAPAN

*e-mail : munehisa.kasuya@boj.or.jp

** e-mail : tomoki.tanemura@boj.or.jp

Small Scale Bayesian VAR Modeling of the Japanese Macro Economy Using the Posterior Information Criterion and Monte Carlo Experiments *

Munehisa Kasuya
Bank of Japan

Tomoki Tanemura
Bank of Japan

February 2000

Abstract

We construct Bayesian vector autoregressive (BVAR) models optimized by the Posterior Information Criterion (PIC), in which hyperparameters are data-determined in the same way as the lag length and trend order. We also assess the performance of the selected models by one-step ahead forecasts using historical data and Monte Carlo experiments. The results suggest that the selected models have a superior performance in forecasting as compared with ordinary VAR models.

Keywords: Bayesian vector autoregression, Posterior Information Criterion, forecasting, model selection
JEL Classification: C51, C52, E17

* We are grateful to Michael McAleer for helpful suggestions and comments. Views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Japan or the Research and Statistics Department.

1 Introduction

Bayesian vector autoregressive (BVAR) models are used for forecasting empirical time series[1]. Bayesian VAR models reduce the tendency of unrestricted VAR models to be over-parameterized, by placing prior distributions over the parameters of the unrestricted VAR models. Litterman(1980,1986) introduced a class of priors for VAR models that induce a random walk mean for the coefficients and have a parsimonious set of hyperparameters, so-called tightness parameters, that govern their variance[2]. These priors are known as Minnesota priors.

BVAR models are closely related to reduced rank regression models (RRR, or Vector Error Correction Models), which are popular models in cointegration analysis. This is because RRR models, formulated in terms of levels and differences, are theoretically based on VAR models in terms of simple levels in the same way as are BVAR models[3].

BVAR models with Minnesota priors are not fully data-determined models. These models require the prespecification of several parameters, including the lag length of the VAR, the degree of any deterministic trend that is to be included, and the setting of the hyperparameters that govern the Minnesota prior variances[4]. We determine the values of the hyperparameters using the Posterior Information Criterion (PIC), which is an information criterion such as AIC[5] and BIC[6]. When the data are stationary and ergodic, PIC is asymptotically equivalent to BIC[7]. However, when the data are nonstationary, PIC imposes a greater penalty than BIC on the presence of additional nonstationary regressors. Phillips and Ploberger(1994) show that PIC generally outperforms BIC for both stationary and nonstationary

[1] Bayesian VAR models are also frequently used by economists of central banks (see Crone and McLaughlin (1999), Sims and Zhao (1998), Stark (1998), and Wong and Jolly (1994)).

[2] See Sims (1980), Doan, Litterman and Sims (1984) and Sims (1993) for some applications of Bayesian VAR models.

[3] We do not adopt RRR models. This is because rank tests tend to have low power, as shown in Kasuya and Ueda (2000), and we avoid the possible bias of rank tests.

[4] One option for selecting tightness hyperparameters is to use root mean square errors, or Theil-U statistics, of inter-sample forecast, as suggested by Litterman(1986). However, absolute accuracy of inter-sample forecast is only weakly associated with the validity of the models. We also conducted Monte Carlo experiments for assessing this method, the results of which show the models optimized by PIC have a superior performance in forecasting as compared with the models selected by the inter-sample forecast method.

[5] See Akaike(1969,1977).

[6] See Schwarz(1978).

[7] See Phillips and Ploberger(1994) and Philips and Ploberger (1996) for PIC.

data by Monte Carlo experiments [8] [9].

We construct Bayesian vector autoregressive (BVAR) models optimized by PIC, in which the hyperparameters are data-determined in the same way as the lag length and trend order. We assess the performance of the selected models by one-step ahead forecasts using historical data, and by Monte Carlo experiments. The results suggest that the selected models have a superior performance as compared with ordinary VAR models.

The paper is organized as follows. In Section 2, we construct Bayesian VAR models of the Japanese economy, using PIC. In Section 3, we assess the performance of the Bayesian VAR by one-step ahead forecasts using historical data and Monte Carlo experiments. Section 4 concludes.

2 Bayesian VAR model

Consider a vector autoregression model with k lags and a deterministic trend of degree l:

$$y_t = \Sigma_{i=1}^{k} A_i y_{t-i} + \Sigma_{j=0}^{l} c_j t^j + \epsilon_t, \tag{1}$$

where y_t and ϵ_t are m-vector processes, t is a time variable, A_i and c_j are parameters, and $\epsilon_t \sim iidN(0, \Sigma)$ is a zero mean stationary process. When $l = -1$ there is no intercept in the model, when $l = 0$ there is a fitted intercept, and when $l = 1$ there is a fitted linear trend.

We can rewrite equation (1) in terms of levels and differences as follows:

$$\Delta y_t = \Phi_1 y_{t-1} + \Sigma_{i=2}^{k} \Phi_i \Delta y_{t-i+1} + \Sigma_{j=0}^{l} d_j t^j + \epsilon_t, \tag{2}$$

where $\Phi_1 = \Sigma_{i=1}^{k} A_i - I$ and $\Phi_i = -\Sigma_{j=i}^{k} A_j (i \geq 2)$, that is,

$$\Delta y_{it} = c_i' x_t + \epsilon_{it}, \quad var(\epsilon_{it}) = \sigma_i^2 \quad (i = 1, ..., m). \tag{3}$$

Let $\pi(c_i) \equiv N(\bar{c}_i, V_{c_i})$ be a prior density for the elements of c_i. The posterior distribution of c_i is $N(\tilde{c}_{in}, \tilde{V}_{nc_i})$ with mean vector

$$\tilde{c}_{in} = [V_{c_i}^{-1} + \sigma_i^{-2} X_n' X_n]^{-1} [V_{c_i}^{-1} \bar{c}_i + (\sigma_i^{-2} X_n' X_n) \hat{c}_{in}], \tag{4}$$

[8] PIC uses the sample *Fisher information* matrix as a penalty rather than a simple parameter count such as BIC. BIC may be regarded as a specialization of PIC to the case of stationary regressors. See Phillips and Ploberger (1996) for details.

[9] AIC and BIC cannot determine the hyperparameters. However, Phillips(1996) proposed the modified PIC for Bayesian VAR (PIC^{BVARM}) that can select hyperparameters. We use this criterion. See Phillips(1992,1994,1995b) for some applications of single-equation Bayesian models optimized by PIC. See Phillips(1995a) for an applications of Bayesian VAR models with Minnesota priors, optimized by PIC. See Kawasaki (1992) for an application of Bayesian VAR models of the Japanese economy with Minnesota priors.

and covariance matrix

$$\tilde{V}_{nc_i} = [V_{c_i}^{-1} + \sigma_i^{-2} X_n' X_n]^{-1}, \tag{5}$$

where

$$\hat{c}_{in} = (X_n' X_n)^{-1} X_n' \Delta y_i \tag{6}$$

is the MLE/OLS estimator of the unrestricted vector of coefficients c, $X_n' = [x_1, ..., x_n]$ and $\Delta y = vec(\Delta Y')$.

The Minnesota priors have mean $\bar{c} = 0$ [10]. The covariance matrix V_c of the Minnesota prior for c is diagonal, with elements constituted as follows:

$$\begin{aligned} var[(\Phi_a)_{ij}] \quad &= (\lambda/a)^2 \qquad if\ i = j, \\ &= (\theta\lambda\hat{\sigma}_i/a\hat{\sigma}_i)^2 \quad if\ i \neq j, \end{aligned} \tag{7}$$

for the lag a coefficient matrix $\Phi_a (a = 1, ..., k)$; and

$$var[(d_b)_i] = \infty \tag{8}$$

for the trend degree b deterministic coefficient $(b = 0, ..., l)^{11}$.

With this prior, the inverse of the variance matrix V_c is:

$$V_{c_i}^{-1} = diag(..., (1/var[(\Phi_a)_{ij}]), ..; ..., (1/var[(d_b)_i]), ...). \tag{9}$$

Then we select $(k, l; \lambda, \theta)$ to optimize the PIC:

$$PIC^{BVARM(k,l;\lambda,\theta)} = ln|\tilde{\Sigma}_{nM}| + (1/n) ln \left(|\tilde{B}_{nM}| / |\tilde{B}_{n_0 M}| \right), \tag{10}$$

where

$$\tilde{B}_{nM} = V_{cM}^{-1} + \Sigma_{nM}^{-1} \otimes X_n' X_n, \tag{11}$$

$$V_{cM}^{-1} = diag(V_{c_1}^{-1}, ..., V_{c_m}^{-1}), \tag{12}$$

with

$$(\tilde{\Sigma}_{nM})_{ij} = (1/n) \Sigma_{t=1}^{n} (\Delta y_{it} - \tilde{c}_{in}' x_t)(\Delta y_{it} - \tilde{c}_{jn}' x_t). \tag{13}$$

The data we use are eight quarterly time series of the Japanese economy from 1973/2Q to 1999/3Q: real GDP, GDP deflator, M2+CD, CPI, nominal

[10] This is because the first lag coefficient of unity is already in the differenced dependent variable Δy.

[11] The framework here, as in Phillips(1996), is a little different from that of Litterman(1986) because the model is formulated in terms of levels and differences rather than simply levels.

exchange rates, government bond yields, real fixed investment, and unemployment rates[12]. All series except for nominal exchange rates and interest rates are seasonally adjusted. Estimation is based on 99 observations from 1974:1 to 1999:3, with 7 presample observations used for determining the optimum lag length of the testing method[13] [14](see Appendix A for details.).

The first 64 of 99 observations are used for initial estimation. We estimate Bayesian VAR models and make one-step ahead *ex post* forecasts recursively by increasing the observations from 65 to 99. The values of $(\hat{k}, \hat{l}, \hat{\lambda}, \hat{\theta})$ may be revised on a period-by-period basis as new data become available.

Figures 1-8 show the actual data, forecasts by BVAR.

3 Forecast Performance

In order to assess forecast performance of Bayesian VAR models, we also estimate the ordinary VAR and make one-step ahead *ex post* forecasts[15] in the same way as Bayesian VAR models. Table 1 suggests that root mean square errors (RMSE) of one-step ahead forecasts by Bayesian VAR models are about 5% lower than those of the ordinary VAR models. Although this procedure is a popular method for evaluating forecast performance, the absolute accuracy of inter-sample forecast is only weakly associated with the validity of the models. This is because mis-specified models could forecast well or good models could forecast poorly[16]. Therefore, we also provide Monte Carlo experiments to evaluate the Bayesian models. By using a estimated data generating process, random samples of 207 observations are created for y_{it}, with ϵ_t normally distributed. The first 107 observations are used to initialize the process, leaving 99 observations for experiments in estimation, and the last observation for forecasting. The number of observations for

[12]The selection of economic variables in this paper follows that of Sims(1993), except for the stock price index. Omitting the Japanese stock price index is because it has a volatile phase, a so-called bubble process, in the late 1980s. However, we also constructed a nine-variable model such as Sims(1993), the results of which suggest the Bayesian VAR model optimized by PIC has a better performance than the ordinary VAR model.

[13]All variables are transformed to natural logarithms. Although interest rates are not transformed to logarithms in the usual procedure, we transformed government bond yields to logarithms. This is because of stabilizing variance and avoiding negative estimates of government bond yields. In the late 1990s, interest rates of Japan continued to be very low.

[14]We set the maximum lag length = 6, and the maximum trend degree = 1.

[15]We determined the lag length by likelihood ratio tests.

[16]See Chong and Hendry(1986) for a critique of evaluation methods.

estimation is the same as the data used in our analysis. The Monte Carlo experiments consist of estimating the Bayesian VAR and forecasting the last observation for 5000 trials of the process.

Results of forecast experiments are shown in Figures 9-16, in which forecast errors of the Bayesian VAR models have smaller variances than those of the ordinary VAR models. Table 2 shows standard errors in forecasts by the Bayesian and ordinary VAR models.

4 Implications

We construct Bayesian vector autoregression (BVAR) models optimized by the Posterior Information Criterion (PIC), in which the hyperparameters are data-determined in the same way as the lag length and trend order. We also assess the performance of the selected models by one-step ahead forecasting and Monte Carlo experiments. The results suggest that the selected models have a better forecasting performance than the ordinary VAR models.

According to the generalized Rissanen's theorem proved by Ploberger and Phillips(1998,1999), if the other conditions are the same, the estimation of models with stationary regressors would lead to a smaller loss of information than stochastic trend estimation, and would be better from the viewpoint of parsimony and forecasting. Neither the Bayesian nor classical techniques can be exceptions to this theorem. Model selection by PIC takes into account the penalty of the presence of nonstationary regressors. In this sense, the method may also be regarded as accommodating the theorem.

References

[1] Akaike, H. (1969). "Fitting autoregressive models for prediction," *Annals of the Institute of Statistical Mathematics,* Vol.21, pp.243-247.

[2] Akaike, H. (1977). "On entropy maximization principle," in P. R. Krishnarah ed., *Applications of Statistics,* Amsterdam: North-Holland, pp.27-41.

[3] Chong, Tock Y. and David F. Hendry (1986). "Econometric Evaluation of Linear Macro-Economic Models," *Review of Economic Studies,* Vol.80, pp.671-690.

[4] Crone, Theodore M. and Michael P. McLaughlin (1999). " A Bayesian VAR Forecasting Model for the Philadelphia Metropolitan Area," Working Paper No. 99-7, Federal Reserve Bank of Philadelphia.

[5] Doan, T., R. Litterman and C. Sims (1984). "Forecasting and Conditional Projection using Realistic Prior Distributions", *Econometric Reviews*, Vol.3, pp.1-144.

[6] Kasuya, Munehisa and Kozo Ueda (2000). "Testing the Purchasing Power Parity Hypothesis: Re-examination by Additional Variables, Tests with Known Cointegrating Vectors, Monte Carlo Critical Values, and Fractional Cointegration," Research and Statistics Department Working Paper 003, Bank of Japan.

[7] Kawasaki, Yoshinori(1992). "Bayesian Vector Autoregression," Bank of Japan, *Institute for Monetary and Economic Studies*, vol.10, pp.79-101.

[8] Litterman, Robert B. (1980). "A Bayesian Procedure for Forecasting with Vector Autoregression," Working paper, Department of Economics, Massachusetts Institute of Technology.

[9] Litterman, Robert B. (1986). "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," Journal of Business and Economic Statistics, Vol.4, pp.5-15.

[10] Phillips, Peter C. B. (1992). "Bayes Method for Trending Multiple Time Series with an Empirical Application to the U.S. Economy," Cowles Foundation Discussion Paper No. 1025, Yale University.

[11] Phillips, Peter C. B. (1994). “Bayesian Models and Forecasts of Australian Macroeconomic Time Series,” in Colin P. Hargreaves ed., *Non-stationary Time Series and Cointegration,* Oxford: Oxford University Press.

[12] Philips, Peter C. B. (1995a). “Automated Forecasts of Asia-Pacific Economic Activity,” *Asia-Pacific Economic Review,* Vol.1, No. 1, pp.92-102.

[13] Philips, Peter C. B. (1995b). “Bayesian Model Selection and Prediction with Empirical Applications,” *Journal of Econometrics,* Vol. 69, pp.289-331.

[14] Phillips, Peter C. B. (1996). “Econometric Model Determination,” *Econometrica*, Vol.64, pp.763-812.

[15] Phillips, Peter C. B. and W. Ploberger (1994). “Posterior Odds Testing for a Unit Root with Data-based selection,” *Econometric Theory,* Vol.10, pp.774-808.

[16] Phillips, Peter C. B. and W. Ploberger (1996). “An Asymptotic Theory of Bayesian Inference for Time Series,” *Econometrica,* Vol.84, p.381-412.

[17] Ploberger, Werner and Peter C. B. Phillips (1998). “Rissanen’s Theorem and Econometric Time Series,” Cowles Foundation Discussion Paper No.1197, Yale University.

[18] Ploberger, Werner and Peter C. B. Phillips (1999). “Empirical Limits for Time Series Econometric Models,” Cowles Foundation Discussion Paper No. 1220, Yale University.

[19] Schwarz, G. (1977). “ Estimating the Dimension of a Model,” *Annals of Statistics,* Vol. 12, pp.1425-1433.

[20] Sims, Christopher A.(1980). “Macroeconomics and Reality”, *Econometrica*, Vol.48, No.1, pp.1-48.

[21] Sims, Christopher A. (1993). “A Nine Variable Probabilistic Macroeconomic Forecasting Model,” in Stock, James H. and Mark W. Watson, eds., *Business Cycles, Indicators, and Forecasting* , NBER Studies in Business Cycles, Vol.28, Chicago and London: University of Chicago Press, pp.179-204.

[22] Sims, Christopher A. and Tao Zha (1998). “Bayesian Methods for Dynamic Multivariate Models,” *International Economic Review,* Vol. 39, No. 4, pp.949-968.

[23] Stark, Tom (1998). “ A Bayesian Vector Error Correction Model of the U.S. Economy,” Working Paper No. 98-12, Federal Reserve Bank of Philadelphia.

[24] Wong, J. K. and P. L. Jolly (1994) “A Bayesian Vector Autoregression Model of Inflation”, *New Zealand Economic Papers*, Vol.28, No. 2, pp.117-131.

A Data

CPI	:	Consumer Price Index, excluding freshfood, seasonally adjusted,
$M2CD$	:	Nominal money supply, seasonally adjusted,
GDP	:	Real GDP, seasonally adjusted,
DEF	:	GDP deflator, seasonally adjusted,
$RBND$	:	Government bond yields(10 year bonds),
$NOMEX$	:	Nominal exchange rate,
IFA	:	Real fixed investment, seasonally adjusted.
U	:	Unemployment rate.

(notes)

(1)period: 1973/2Q-1999/3Q

(2)BOJ, Statistics Bureau and Statistics Center, Economic Planning Agency.

Table 1: Root Mean Square Error

model	GDP	M2CD	CPI	DEF	RBND	NOMEX	IFA	U
BVAR	0.90	0.52	0.26	0.21	192.74	40.37	8.05	17.73
VAR	0.98	0.53	0.32	0.24	197.86	42.85	8.22	20.96

[Notes]

CPI : Consumer Price Index(excluding perishables),

$M2CD$: Nominal money supply,

GDP : Real GDP,

DEF : GDP deflator,

$RBND$: Government bond yields(10 year bonds),

$NOMEX$: Nominal exchange rate,

IFA : Real fixed investment,

U : Unemployment rate.

Figure 1. Inter-sample Frecasts by Bayesian VAR Models
(GDP)

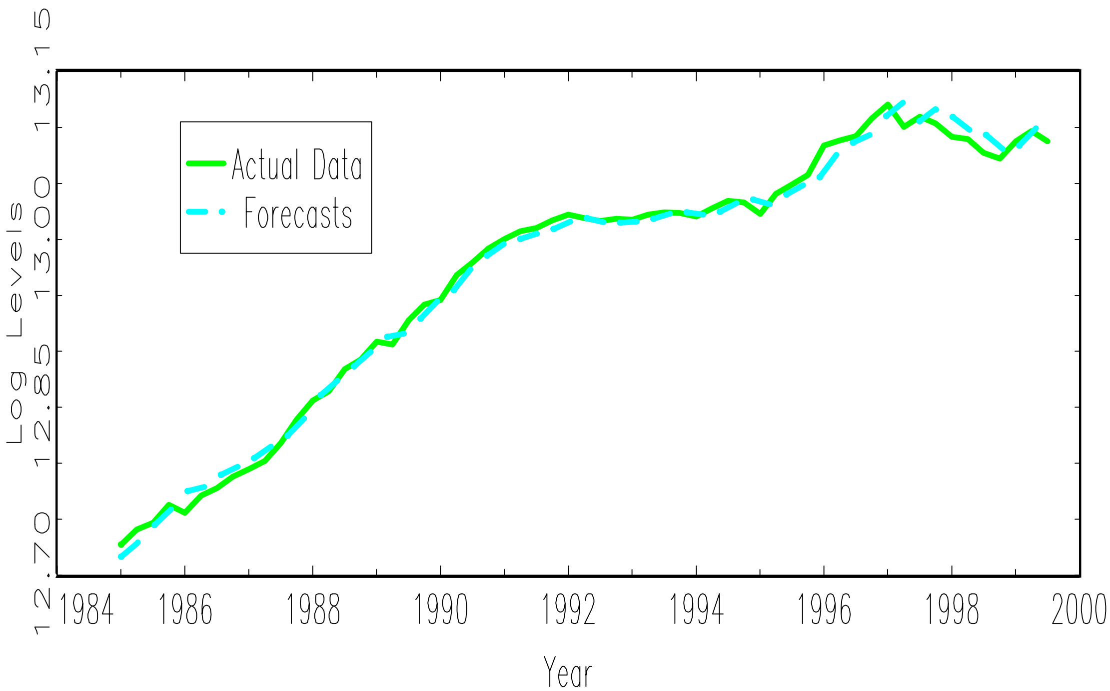

Figure 2. Inter-sample Forecasts by Bayesian VAR Models
(M2+CD)

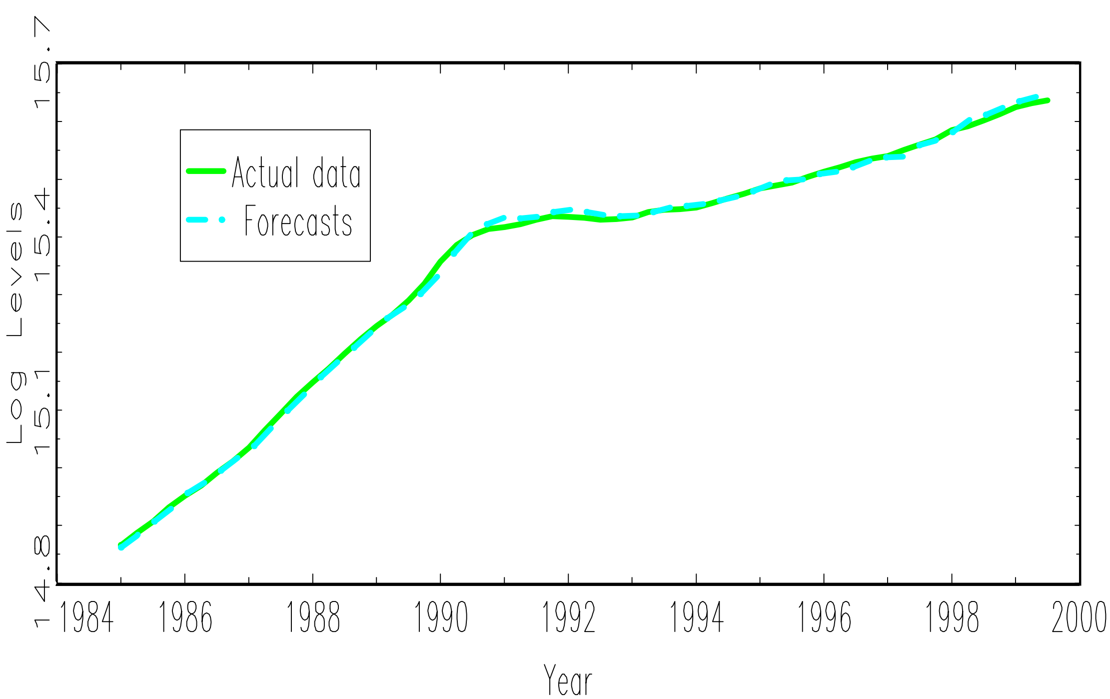

Figure 3. Inter-sample Forecasts by Bayesian VAR Models
(CPI)

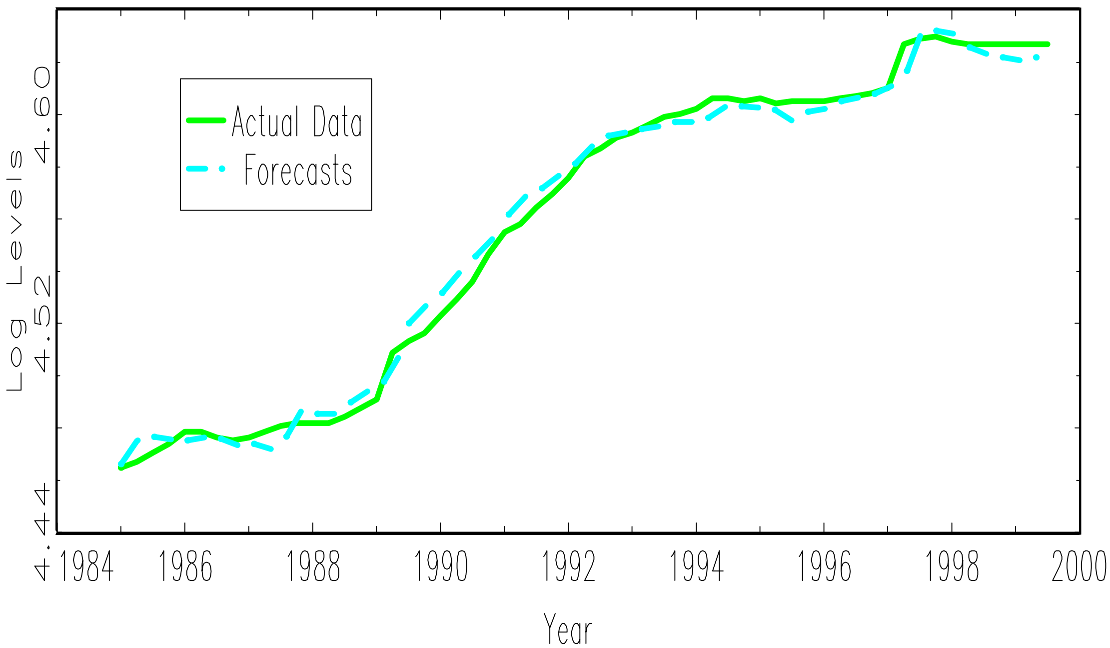

Figure 4. Inter-sample Forecasts by Bayesian VAR Models
(GDP Deflator)

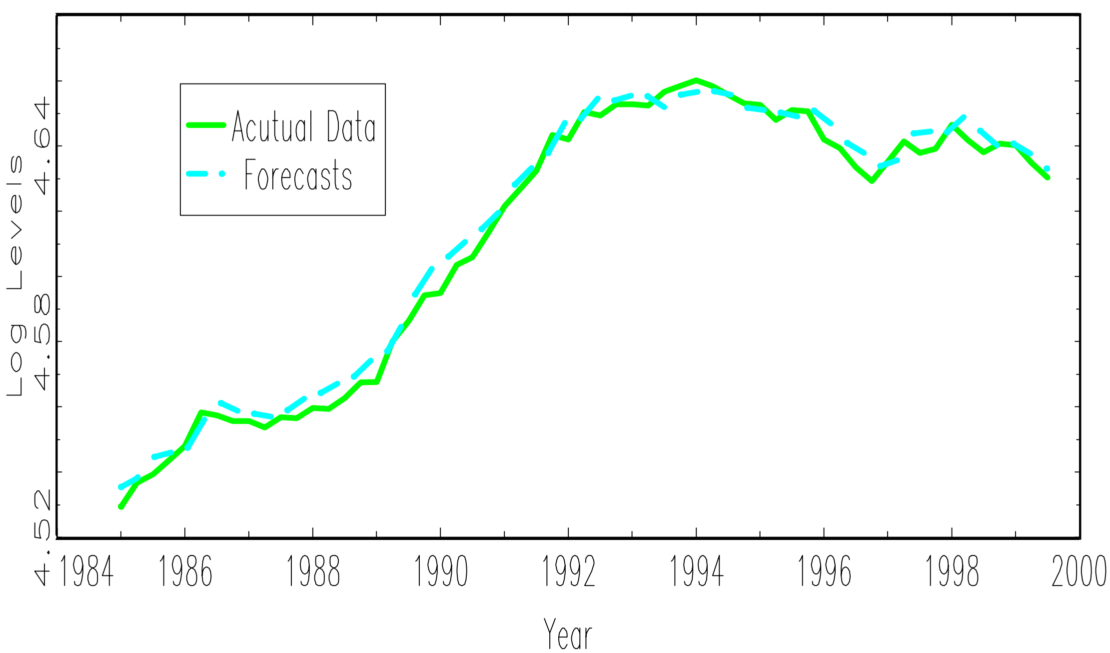

Figure 5. Inter-sample Forecasts by Bayesian VAR Models
(Governmnent Bond Yields)

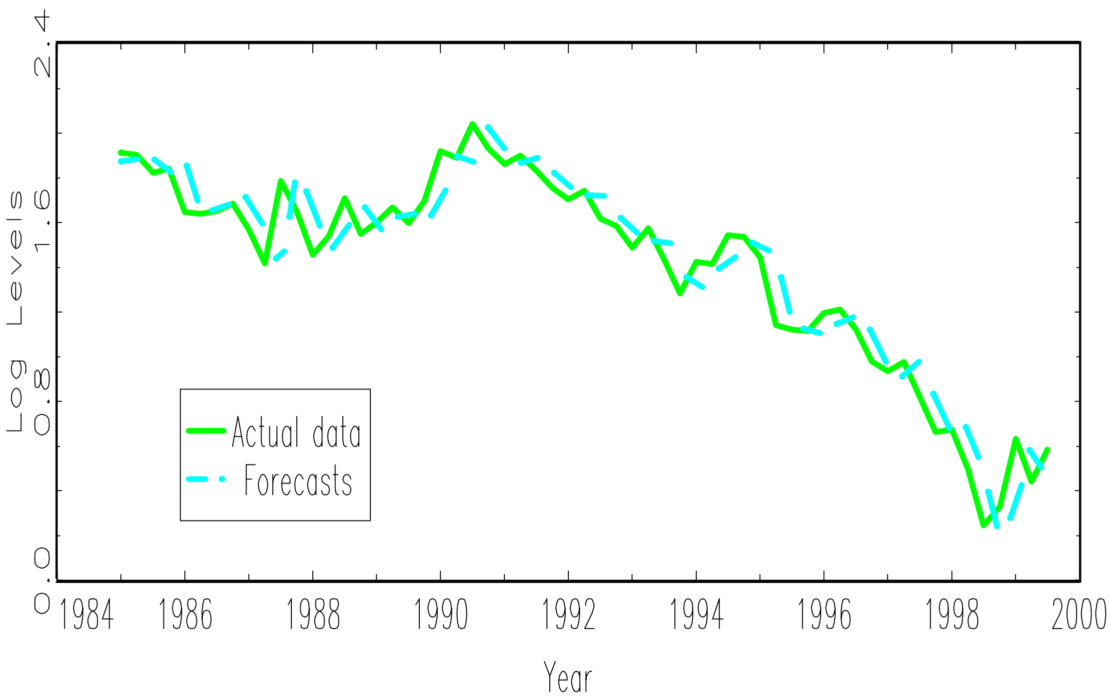

Figure 6. Inter-sample Forecasts by Bayesian VAR
(Real Fixed Investment)

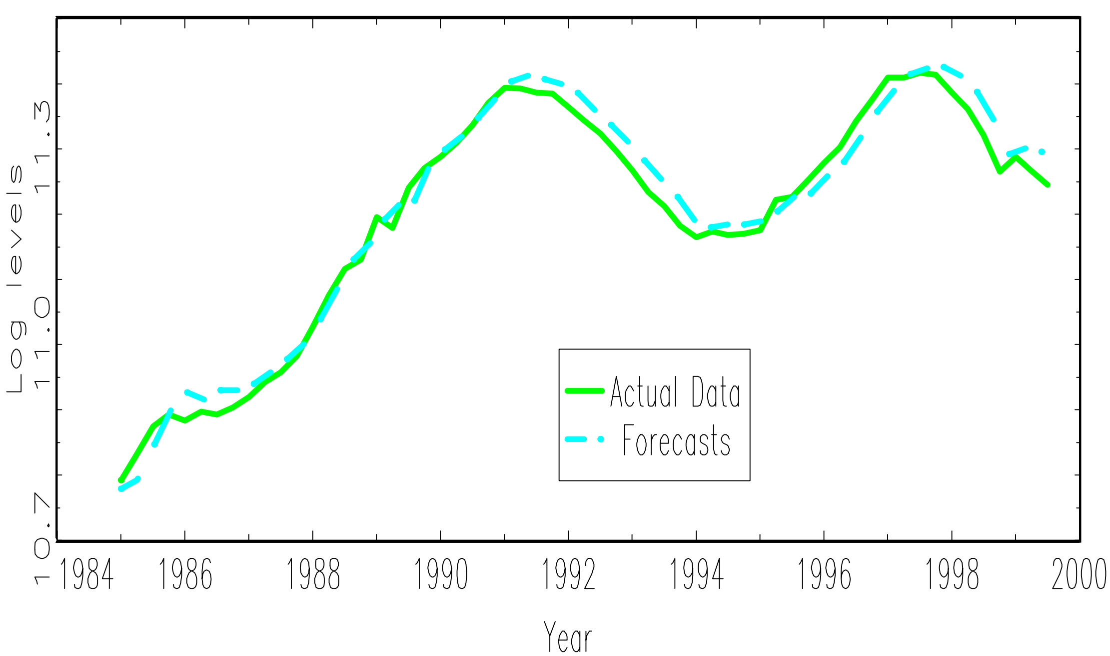

Figure 7. Inter-sample Forecasts by Bayesian VAR Models
(Nominal Exchange Rates)

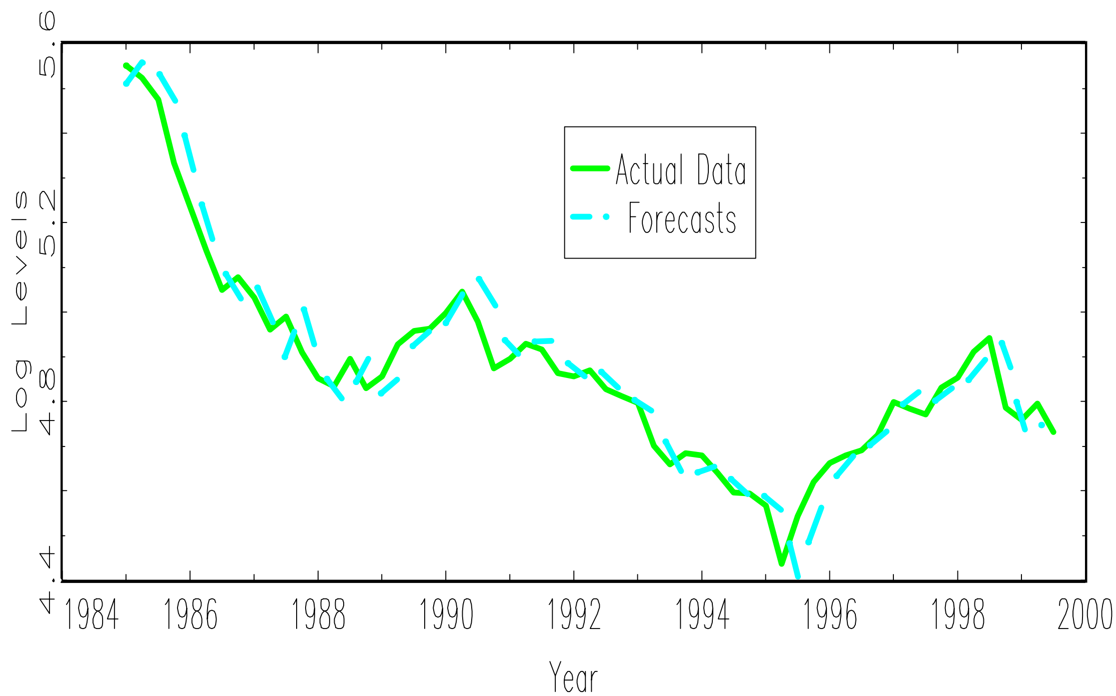

Figure 8. Inter-sample Forecasts by Bayesian VAR Models
(Unemployment Rates)

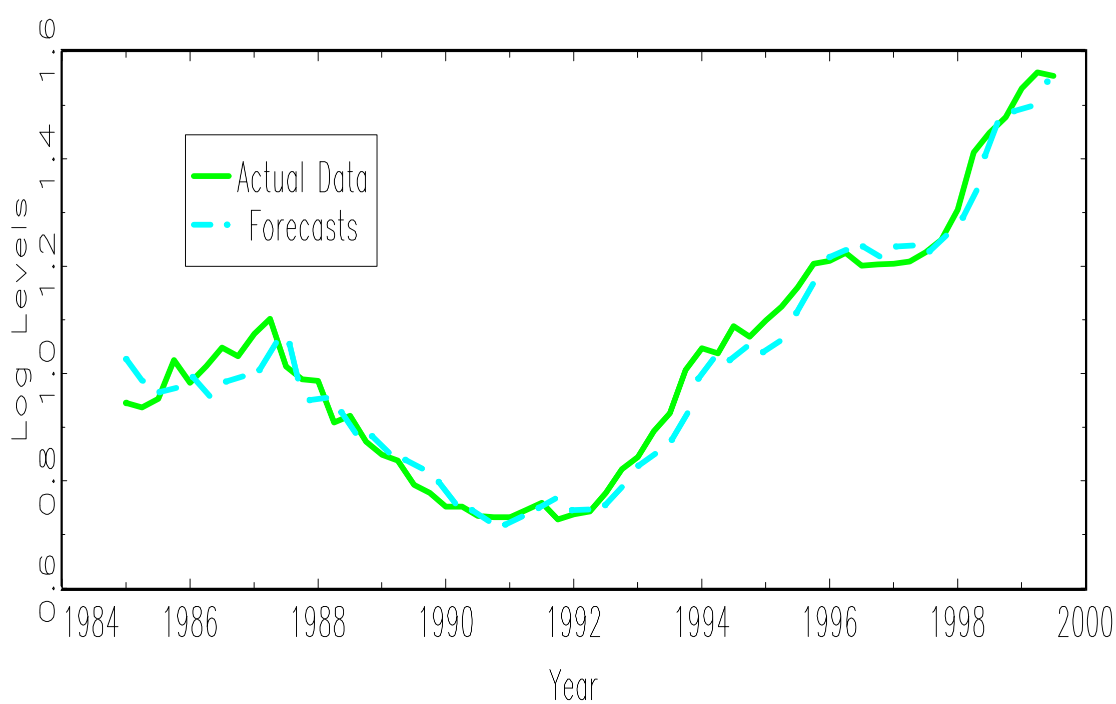

Figure 9. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(GDP)

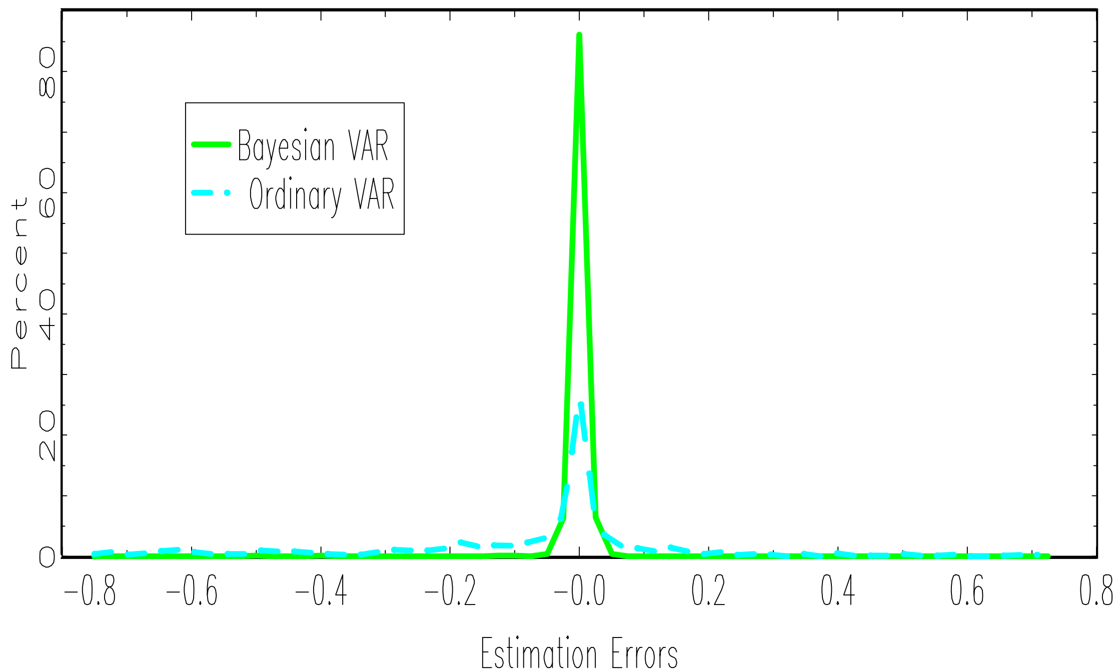

Figure 10. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(M2+CD)

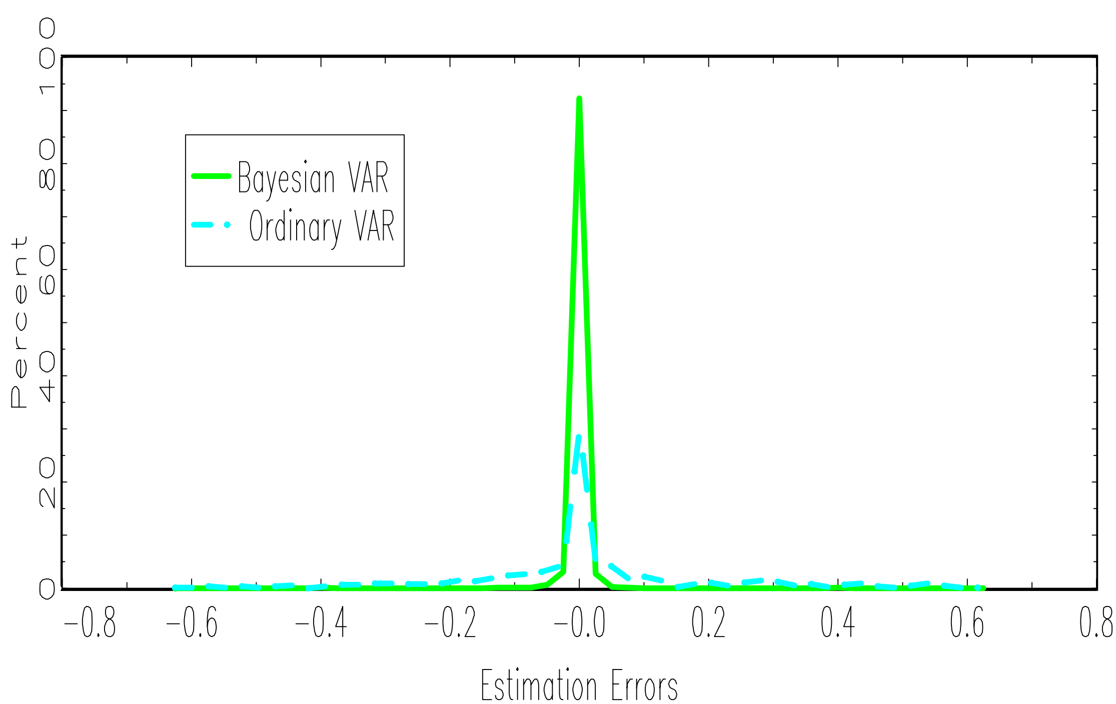

Figure 11. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(CPI)

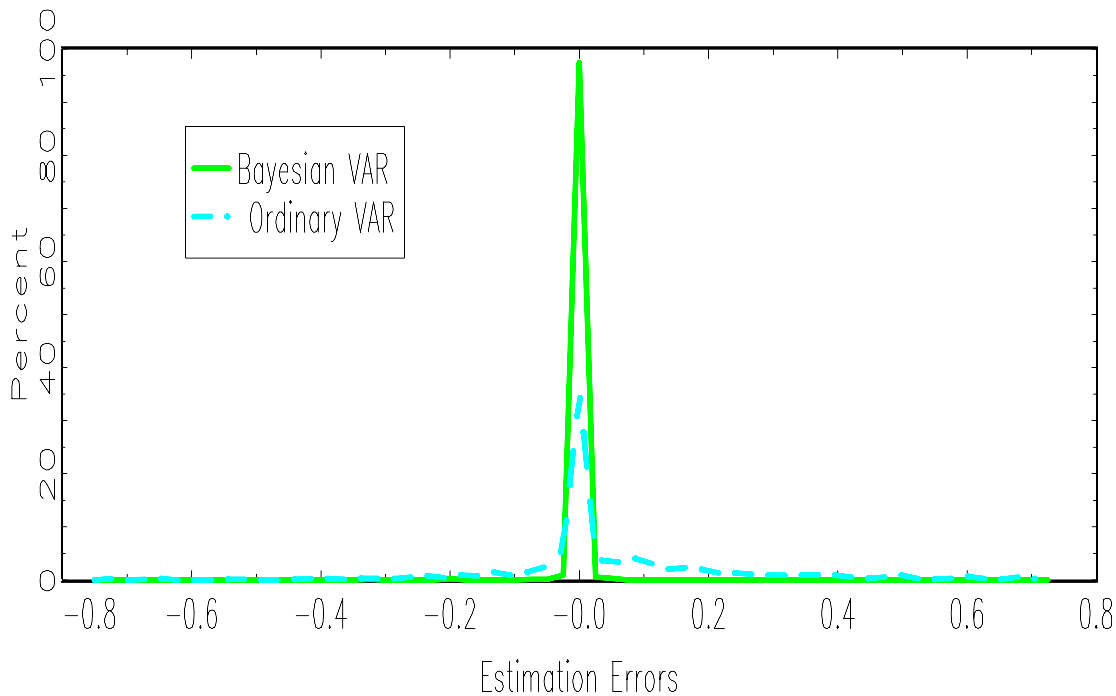

Figure 12. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(GDP Deflator)

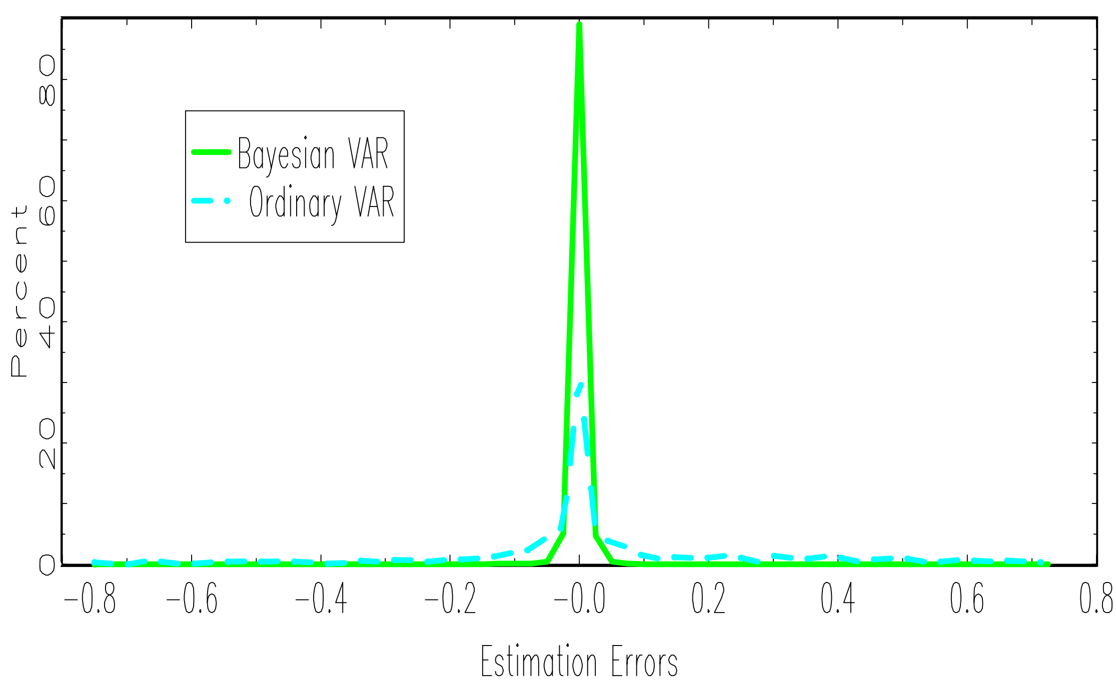

Figure 13. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(Governmnent Bond Yields)

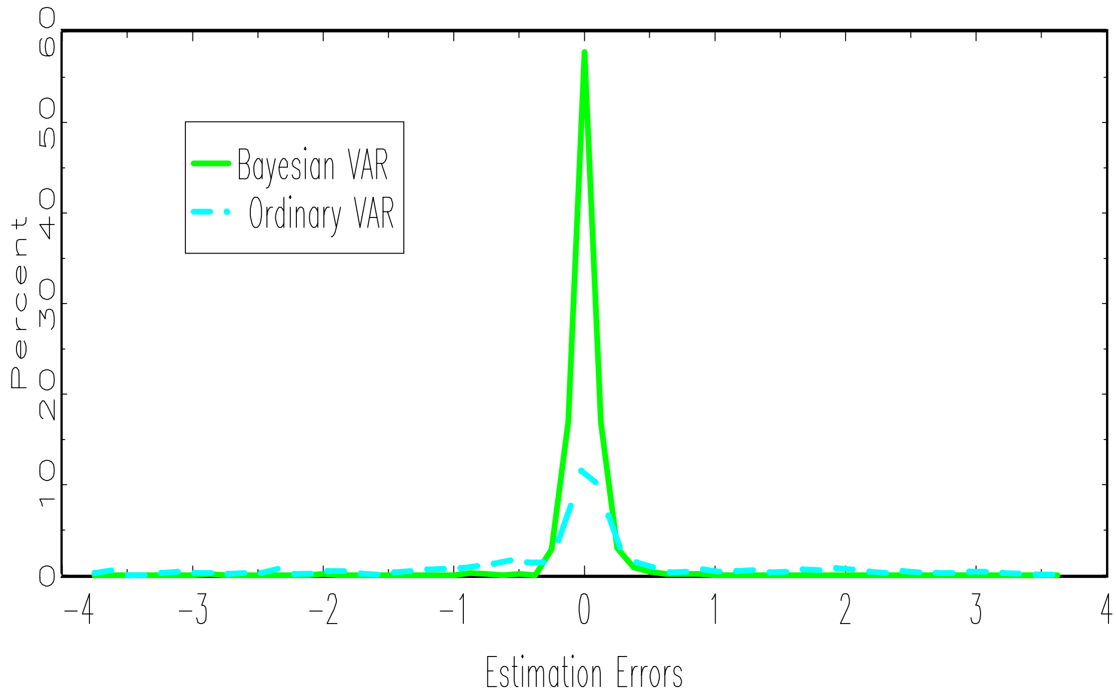

Figure 14. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(Real Fixed Investment)

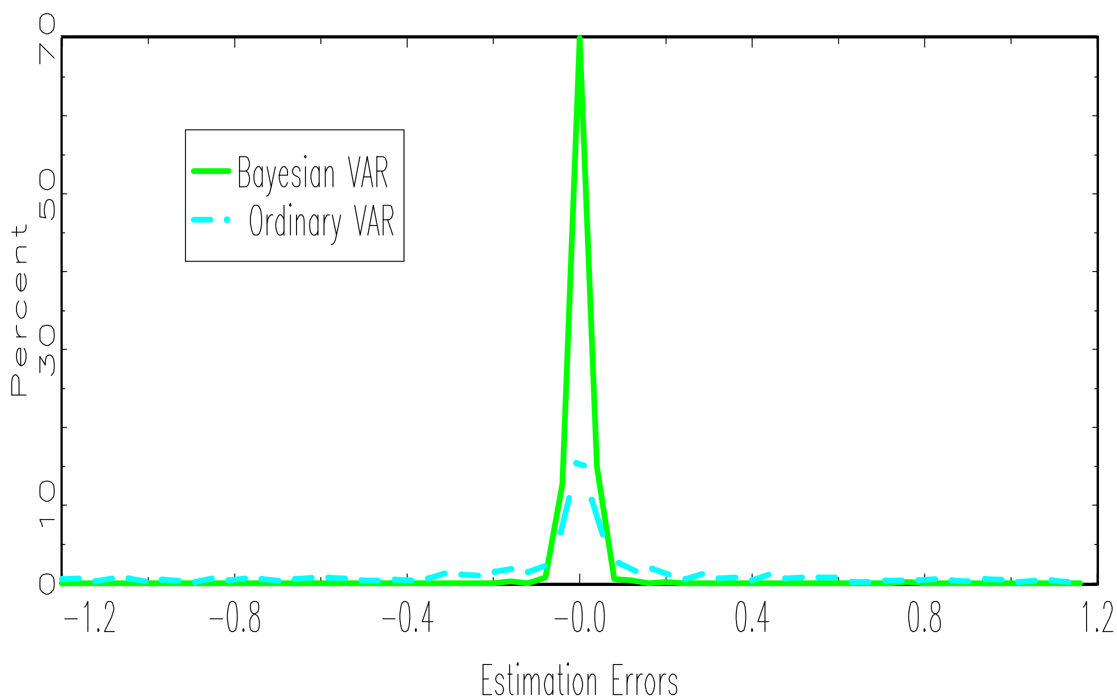

Figure 15. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(Nominal Exchange Rates)

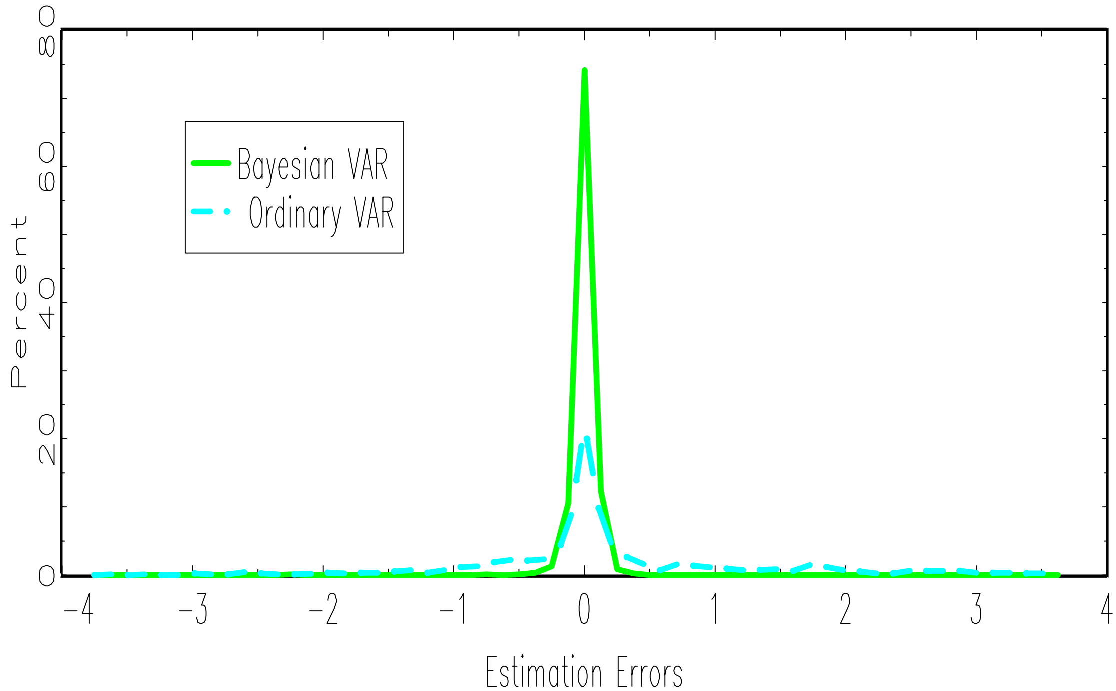

Figure 16. Estimation Errors by Bayesian VAR and Ordinary VAR Models
(Unemployment Rates)