**Computer Tutorial 3: TVP-VARs and Multivariate Stochastic Volatility in VARs**

In this exercise sheet, I provide a series of questions and answers for a question about the TVP-VAR and one about the VAR with stochastic volatility. Given that I am providing the answers, what do I expect you to do in the computer sessions? First I am not expecting you to go through the theoretical derivations and proofs (part a of both questions). However, since I do not have time to do proofs in the lectures, I felt I should make them available to you in case you want to see them or have them as a resource for your future Bayesian research. So I suggest you do only a quick reading of this exercise sheet (without worrying about the details of the proofs) before the computer tutorial to get an overview of the model each exercise relates to. Instead focus on the computational part of the each exercise (part b of the two exercises). Code which does these exercises is provided. In the computer lab, I suggest you experiment with these codes (e.g. try different priors, lag lengths, etc.) to familiarize yourself with Bayesian programming in VARs.

Some of the notation I use in this exercise sheet is different than that used in the lecture.

**Exercise 1 (The Time Varying Parameter VAR)**

In many macroeconomic applications, the inter-relationships between variables changes over time implying that the VAR coefficients should change. The Time-Varying Parameter VAR (TVP-VAR) is a popular model in such cases.

Consider again the VAR($p$) but now with time-varying parameters:

$$y_t = a_t + A_{1t} y_{t-1} + \cdots + A_{pt} y_{t-p} + \epsilon_t, \tag{1}$$

where $\epsilon_t \sim N(0, \Sigma)$. Define $X_t = I_n \otimes [1, y'_{t-1}, \ldots, y'_{t-p}]$ and $\beta_t = \text{vec}([a_t,\ A_{1t},\ \ldots,\ A_{pt}]')$ and rewrite the above system as

$$y_t = X_t \beta_t + \epsilon_t.$$

The time-varying parameters $\beta_t$ are assumed to evolve as a random walk

$$\beta_t = \beta_{t-1} + u_t, \tag{2}$$

where $u_t \sim N(0, Q)$ and the initial conditions $\beta_0$ are treated as parameters. In this question, make the simplifying assumption that the covariance matrix $Q$ is diagonal, i.e., $Q = \text{diag}(q_1, \ldots, q_{kn})$ where $k = np + 1$ is the number of explanatory variables in each equation of the TVP-VAR. One can also consider the possibility of a block-diagonal matrix or even a full matrix, but the form of the Gibbs sampler will change slightly. Note that the TVP-VAR is a state space model with measurement equation (1) and state equation (2).

To complete the model specification, consider independent priors for $\Sigma$, $\beta_0$ and the diagonal elements of $Q$:

$$\Sigma \sim IW(\nu_0, S_0), \quad \beta_0 \sim N(a_0, B_0), \quad q_i \sim IG(\nu_{0,q_i}, S_{0,q_i}).$$

(a) Derive a Gibbs sampler which allows for posterior inference in the TVP-VAR.

(b) Use the data set provided. Estimate a TVP-VAR using this data set. Carry out an impulse response analysis and discuss whether impulse responses have changed over time.

**Solution 1**

(a) The parameters in the TVP-VAR are $\beta_0$, $\Sigma$ and $Q$, and the states are $\beta = (\beta'_1, \ldots, \beta'_T)'$ and we derive a Gibbs sampler involving these four blocks.

First, we derive the posterior for $\beta$, conditional on the other parameters. Re-write (1) as

$$y = X\beta + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, I_T \otimes \Sigma)$ and

$$X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_T \end{pmatrix}.$$

Hence, we have

$$(y \mid \beta, \Sigma) \sim N(X\beta, I_T \otimes \Sigma).$$

So we have re-framed the TVP-VAR as a Normal linear regression model. Next, we derive the prior of $\beta$. To that end, rewrite (2) in matrix notation:

$$H\beta = \tilde{\alpha}_{\boldsymbol{\beta}} + u,$$

where $u \sim N(0, I_T \otimes Q)$, $\tilde{\alpha}_\beta = (\beta_0', 0, \ldots, 0)'$ and

$$H = \begin{pmatrix} I_{nk} & 0 & 0 & \cdots & 0 \\ -I_{nk} & I_{nk} & 0 & \cdots & 0 \\ 0 & -I_{nk} & I_{nk} & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -I_{nk} & I_{nk} \end{pmatrix}.$$

Note that $H$ is of dimension $Tnk \times Tnk$, and is a multivariate generalization of the usual first difference matrix and that $|H| = 1$, and is therefore invertible. It can be shown that $H^{-1}\tilde{\alpha}_\beta = 1_T \otimes \beta_0$. Therefore, the prior of $\beta$ is given by

$$(\beta \mid \beta_0, Q) \sim N(1_T \otimes \beta_0, (H'(I_T \otimes Q^{-1})H)^{-1}).$$

Thus, the TVP-VAR can also be seen as a multivariate regression with a Normal prior and standard multivariate regression can be used to derive:

$$(\beta \mid y, \Sigma, \beta_0, Q) \sim \mathcal{N}(\hat{\beta}, K_{\boldsymbol{\beta}}^{-1}),$$

where

$$\begin{aligned} K_\beta &= H'(I_T \otimes Q^{-1})H + X'(I_T \otimes \Sigma^{-1})X, \\ \hat{\beta} &= K_\beta^{-1}\left(H'(I_T \otimes Q^{-1})H(1_T \otimes \beta_0) + X'(I_T \otimes \Sigma^{-1})y\right). \end{aligned}$$

The conditional posterior for $\Sigma$ can be derived using a similar derivation as for the VAR questions of Computer Session 1. That is, conditional on $\beta$, the TVP-VAR has the same structure as a VAR with known coefficients and it can be shown that

$$(\Sigma \mid y, \beta, Q) \sim \mathcal{IW}\left(\nu_0 + T, S_0 + \sum_{t=1}^{T}(y_t - X_t\beta_t)(y_t - X_t\beta_t)'\right).$$

In a similar vein, conditional on $\beta$, the state equations have the same structure as regressions with the diagonal elements of $Q$=diag$(q_1, \ldots, q_{nk})$ having the same role as error variances. Thus, it can be seen that:

$$(q_i \mid y, \beta, \beta_0) \sim IG\left(\nu_{0,q_i} + \frac{T}{2}, S_{0,q_i} + \frac{1}{2}\sum_{t=1}^{T}(\beta_{it} - \beta_{i(t-1)})^2\right).$$

Finally, since $\beta_0$ only appears in the first state equation

$$\beta_1 = \beta_0 + u_1,$$

where $u_1 \sim N(0, Q)$. Given the Normal prior $\beta_0 \sim N(a_0, B_0)$, we can use standard linear regression results to get

$$(\beta_0 \,|\, y, \beta, Q) \sim \mathcal{N}(\hat{\beta}_0, K_{\boldsymbol{\beta}_0}^{-1}),$$

where

$$K_{\boldsymbol{\beta}_0} = B_0^{-1} + Q^{-1}, \quad \hat{\beta}_0 = K_{\boldsymbol{\beta}_0}^{-1}\left(B_0^{-1}a_0 + Q^{-1}\beta_1\right).$$

We summarize the Gibbs sampler as follows:

**Algorithm 1** ***(Gibbs Sampler for the TVP-VAR($p$)).***
*Pick some initial values for $\beta^{(0)}$, $\Sigma^{(0)}$, $Q^{(0)}$ and $\beta_0^{(0)}$. Then, repeat the following steps from $r = 1$ to $R$:*

1. *Draw $\beta^{(r)} \sim (\beta \,|\, y, \Sigma^{(r-1)}, Q^{(r-1)}, \beta_0^{(r-1)})$ (multivariate Normal).*
2. *Draw $\Sigma^{(r)} \sim (\Sigma \,|\, y, \beta^{(r)}, Q^{(r-1)}, \beta_0^{(r-1)})$ (inverse-Wishart).*
3. *Draw $Q^{(r)} \sim (Q \,|\, y, \beta^{(r)}, \Sigma^{(r)}, \beta_0^{(r-1)})$ (independent inverse-gammas).*
4. *Draw $\beta_0^{(r)} \sim (\beta_0 \,|\, y, \beta^{(r)}, \Sigma^{(r)}, Q^{(r)})$ (multivariate Normal).*

(b) In the empirical example in Computer Session 1, we estimated a 3-variable VAR(2) of inflation, unemployment and the interest rate and used it to compute impulse response functions to a 100-basis-point monetary policy shock. Given the sample covers a long period, one might wonder if the responses to the monetary policy shocks were different early in the sample from later in the sample. To address this question, we revisit this example using a TVP-VAR. In particular, we compare the impulse responses associated with the VAR coefficients in 1975 to those in 2005. We use the same data set, lag length and scheme to identify the monetary policy shock with the TVP-VAR as we did in Computer Session 1.

MATLAB script VAR_TVP.m implements the Gibbs sampler derived in the solution to part a). The MATLAB script construct_IR.m calculates the impulse responses.

Figure 1 reports two sets of impulse response functions to a 100-basis-point monetary policy shock. The first set is computed using the VAR coefficients at 1975Q1; the second set uses those at 2005Q1. As the figure shows, the impulse-response functions of both inflation and the interest rate are very similar across the two time periods, whereas those of the unemployment seem to be more different.

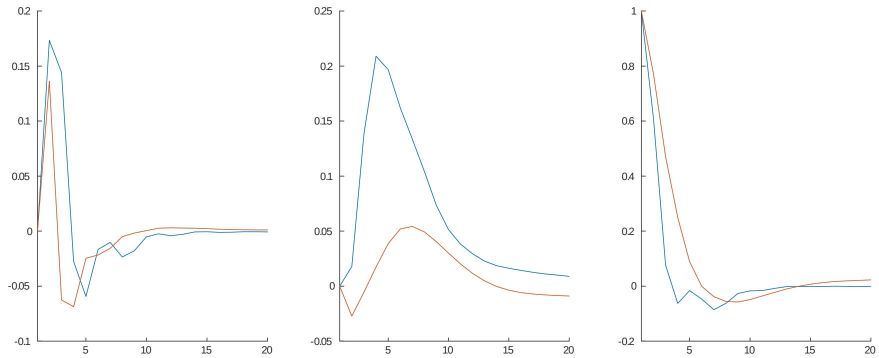


Figure 1: Impulse-response functions of inflation (left panel), unemployment (middle panel) and interest rate (right panel) to a 100-basis-point monetary policy shock for 1975Q1 and 2005Q1

We report in Figure 2 the mean differences between the two sets of impulse-response functions of the three variables, as well as the associated 90% credible intervals. Despite some large absolute differences, parameter uncertainty is high and most of the credible intervals contain zero. There seems to be some evidence that the responses of unemployment are different across the two periods, but the evidence is not conclusive.

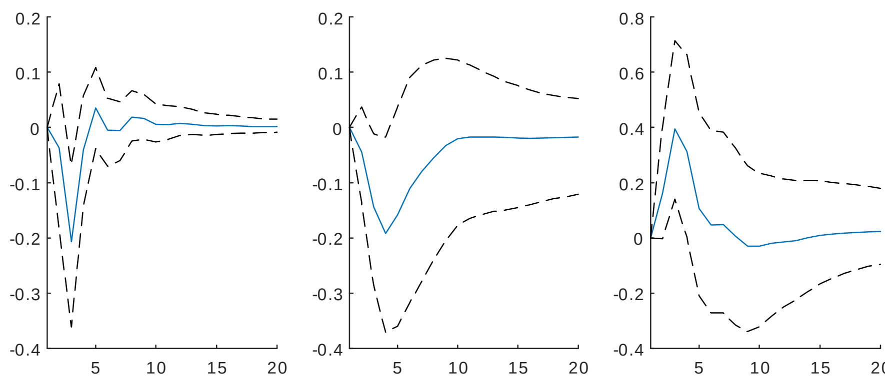


Figure 2: Differences between the impulse-response functions of inflation (left panel), unemployment (middle panel) and interest rate (right panel) at 2005Q1 and 1975Q1. The dotted lines are the 5% and 95% quantiles

**Exercise 2 (The VAR with Stochastic Volatility)**

This exercise shows how one can extend the standard VAR with homoscedastic errors to produce a VAR with stochastic volatility (VAR-SV). We start with the constant-coefficient VAR of Computer Session 1. However, we extend it by allowing the errors to have a time-varying covariance matrix $\Sigma_t$:

$$y_t = X_t\beta + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_t). \tag{3}$$

We want $\Sigma_t$ to evolve smoothly, while at each time period $\Sigma_t$ must be a valid covariance matrix—i.e., it is symmetric and positive definite. There are a few approaches to model such an evolution, and here we follow the approach in the seminal paper by Cogley and Sargent (2005). They model $\Sigma_t$ as

$$\Sigma_t^{-1} = L' D_t^{-1} L,$$

where $D_t$ is a diagonal matrix and $L$ is a lower triangular matrix with ones on the main diagonal, i.e.,

$$D_t = \begin{pmatrix} e^{h_{1t}} & 0 & \cdots & 0 \\ 0 & e^{h_{2t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{h_{nt}} \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)} & 1 \end{pmatrix}.$$

By construction $\Sigma_t$ is symmetric and positive definite for any values of $h_t = (h_{1t}, \dots, h_{nt})'$ and $a = (a_{21}, a_{31}, a_{32}, \dots, a_{n1}, \dots, a_{n(n-1)})'$. Note that the dimension of $a$ is $m = n(n-1)/2$. Then, each $h_{it}$ is specified independently using a univariate stochastic volatility model. More precisely, each $h_{it}$ evolves according to the following random walk

$$h_{it} = h_{i(t-1)} + u_{it}^h,$$

where $u_{it}^h \sim \mathcal{N}(0, \sigma_{h,i}^2)$ and $h_{i0}$ is treated as an unknown parameter.

It is worth noting that the parameters in $a$ are restricted to be constant here. Primiceri (2005) considers an extension where these parameters are time-varying and modeled as random walks (i.e. both $a$ and $\beta$ are time varying). The Gibbs sampler for Primiceri's model requires extra blocks relative to the Gibbs sampler derived in this exercise. These can be obtained using standard results for the Normal linear state space model.

By construction, $\Sigma_t = L^{-1} D_t (L^{-1})'$, and therefore we can express each element of $\Sigma_t$ in terms of the elements of $D_t$ and $L^{-1} = (a^{ij})$. More precisely, we have

$$\sigma_{ii,t} = e^{h_{it}} + \sum_{k=1}^{i-1} e^{h_{kt}} (a^{ik})^2, \;\; i = 1, \dots, n,$$

$$\sigma_{ij,t} = a^{ij} e^{h_{jt}} + \sum_{k=1}^{j-1} a^{ik} a^{jk} e^{h_{kt}}, \;\; 1 \le j < i \le n,$$

where $\sigma_{ij,t}$ is the $(i,j)$ element of $\Sigma_t$. Note that the log-volatility $h_{1t}$ affects the variances of all the variables, whereas $h_{nt}$ impacts only the last variable.

In addition, despite the assumption of a constant matrix $L$, this setup also allows for some form of time-varying correlations among the errors. This can be seen via a simple example. Using the formulas above, we have

$$\sigma_{11,t} = \mathrm{e}^{h_{1t}}, \quad \sigma_{22,t} = \mathrm{e}^{h_{2t}} + \mathrm{e}^{h_{1t}}(a^{21})^2, \quad \sigma_{12,t} = a^{21}\mathrm{e}^{h_{1t}}.$$

We have used the fact that $L^{-1}$ is a lower triangular matrix with ones on the main diagonal, and therefore $a^{11} = 1$ and $a^{12} = 0$. Now, the (1,2) correlation coefficient is given by

$$\frac{\sigma_{12,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}} = \frac{a^{21}}{\sqrt{\mathrm{e}^{h_{2t}-h_{1t}} + (a^{21})^2}}.$$

Hence, as long as $h_{1t}$ and $h_{2t}$ are not identical for all $t$, this correlation coefficient is time-varying.

To complete the model specification, use independent priors for $\beta$, $a$, $\sigma^2_h = (\sigma^2_{h,1}, \ldots, \sigma^2_{h,n})'$ and $h_0 = (h_{10}, \ldots, h_{n0})'$:

$$\beta \sim N(\beta_0, V_\beta), \quad a \sim N(a_0, V_a), \quad \sigma^2_{h,i} \sim IG(\nu_{0,h_i}, S_{0,h_i}), \quad h_0 \sim N(b_0, B_0).$$

(a) Derive an MCMC algorithm which allows for posterior inference in this VAR-SV.

(b) Estimate the VAR-SV using the data set of Computer Session 1 but extended to 2013Q4 so as to investigate whether the Great Recession involved changes in volatilities (this is US_macrodata1.csv on the book's website). Plot the volatilities and discuss your results.

**Solution 2**

(a) To estimate the VAR-SV model, we derive a Gibbs sampler that combines standard Bayesian VAR methods (see Exercise 21.1) which methods for drawing the stochastic volatilities. Remember that stochastic volatility models are state space models and the auxiliary mixture sampler discussed in the lecture allows for Bayesian estimation of univariate stochastic volatility processes.

The model parameters are $\beta$, $a$, $\sigma^2_{h,i}$ and $h_{0,i}$, and the states are the log-volatilities $h_{i,1:T} = (h_{i1}, \ldots, h_{iT})'$ for $i = 1, .., n$. We derive a 5-block Gibbs sampler involving each of these. The two non-standard steps are sampling of $a$ and $h = (h'_{1,1:T}, \ldots, h'_{n,1:T})'$, and we will describe them in detail below.

We begin with the sampling of $a$, the lower triangular elements of $L$. First observe that given $y$ and $\beta$, $\epsilon = y - X\beta$ is known. Thus, we can rewrite the model as a system of regressions in which $\epsilon_{it}$ is regressed on $\epsilon_{1t}, \ldots, \epsilon_{(i-1)t}$ for $i = 2, \ldots, n$, and $a_{i1}, \ldots, a_{i(i-1)}$ are the corresponding regression coefficients. If we can rewrite the model this way, then we can apply standard linear regression results to sample $a$.

To be precise, note that

$$L\epsilon_t = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \vdots \\ \epsilon_{nt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} + a_{21}\epsilon_{1t} \\ \epsilon_{3t} + a_{31}\epsilon_{1t} + a_{32}\epsilon_{2t} \\ \vdots \\ \epsilon_{nt} + \sum_{j=1}^{n-1} a_{nj}\epsilon_{jt} \end{pmatrix}$$

$$= \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \vdots \\ \epsilon_{nt} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ -\epsilon_{1t} & 0 & 0 & 0 & 0 & \cdots & & \vdots \\ 0 & -\epsilon_{1t} & -\epsilon_{2t} & 0 & 0 & \cdots & & 0 \\ \vdots & & & \ddots & \ddots & & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & -\epsilon_{1t} & \cdots & -\epsilon_{t(n-1)} \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{31} \\ a_{32} \\ \vdots \\ a_{n(n-1)} \end{pmatrix}$$

Or more succinctly,

$$L\epsilon_t = \epsilon_t - E_t a.$$

Noting that $|\Sigma_t| = |D_t|$, we can rewrite the likelihood implied by (3) as

$$\begin{aligned} p(y \,|\, \beta, a, h) &\propto \left(\prod_{t=1}^{T} |D_t|^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \epsilon_t'(L'D_t^{-1}L)\epsilon_t\right) \\ &= \left(\prod_{t=1}^{T} |D_t|^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2}\sum_{t=1}^{T} (L\epsilon_t)'D_t^{-1}(L\epsilon_t)\right) \\ &= \left(\prod_{t=1}^{T} |D_t|^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2}\sum_{t=1}^{T} (\epsilon_t - E_t a)'D_t^{-1}(\epsilon_t - E_t a)\right). \end{aligned}$$

In other words, the likelihood is the same as that implied by the regression

$$\epsilon_t = E_t a + \eta_t, \tag{4}$$

where $\eta_t \sim N(0, D_t)$. Therefore, stacking (4) over $t = 1, \ldots, T$, we have

$$\epsilon = Ea + \eta,$$

where $\eta \sim N(0, D)$ with $D = \mathrm{diag}(D_1, \ldots, D_T)$. Given the prior $a \sim N(a_0, V_a)$, it then follows that

$$(a \,|\, y, \beta, h) \sim N(\hat{a}, K_a^{-1}),$$

where

$$K_a = V_a^{-1} + E'D^{-1}E, \quad \hat{a} = K_a^{-1}\left(V_a^{-1}a_0 + E'D^{-1}\epsilon\right).$$

To sample the log-volatility $h$, we first compute the orthogonalized innovations: $\tilde{\epsilon}_t = L(y_t - X_t\beta)$ for $t = 1, \ldots, T$. It can be seen that $E(\tilde{\epsilon}_t \,|\, a, h, \beta) = 0$ and

$$cov(\tilde{\epsilon}_t \,|\, a, h, \beta) = L(LD_t^{-1}L)^{-1}L' = D_t.$$

Hence, $(\tilde{\epsilon}_{it} \,|\, a, h, \beta) \sim N(0, \mathrm{e}^{h_{it}})$. Therefore, we can apply the auxiliary mixture sampler described in the lecture to each of the series $\tilde{\epsilon}_{i1}, \ldots, \tilde{\epsilon}_{iT}$ for $i = 1, \ldots, n$. The Gibbs sampler steps for $\sigma^2_{h,i}$ and $h_{0,i}$ for $i = 1, .., n$ are standard and will not be repeated here.

To sample $\beta$, we rewrite (3) as

$$y = X\beta + \epsilon,$$

where $\epsilon \sim N(0, \tilde{\Sigma})$ and $\tilde{\Sigma} = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_T)$ is a block-diagonal matrix. Together with the prior $\beta \sim N(\beta_0, V_{\boldsymbol{\beta}})$, we have

$$(\beta \,|\, y, a, h) \sim \mathcal{N}(\hat{\beta}, K_{\boldsymbol{\beta}}^{-1}),$$

where

$$K_{\boldsymbol{\beta}} = V_{\boldsymbol{\beta}}^{-1} + X'\tilde{\Sigma}^{-1}X, \quad \hat{\beta} = K_{\boldsymbol{\beta}}^{-1}\left(V_{\boldsymbol{\beta}}^{-1}\beta_0 + X'\tilde{\Sigma}^{-1}y\right).$$

Note that $\tilde{\Sigma}^{-1} = \mathrm{diag}(\Sigma_1^{-1}, \ldots, \Sigma_T^{-1})$ with $\Sigma_t^{-1} = L'D_t^{-1}L$.

(b) MATLAB script VAR_SV.m contains the Gibbs sampler of part a) for estimating the VAR-SV model using US quarterly observations on CPI inflation, unemployment and the interest rate from 1959Q2 to 2013Q4.

We report the time-varying volatility of the three equations expressed as standard deviations in Figure 3. The volatility of all three equations decreases substantially in the early 1980s, the timing of which matches the onset of the Great Moderation. The volatility remains low until the Great Recession. These results show that the error variances of all three equations have substantial time variation. Extending the standard VAR with constant variance to one with stochastic volatility is therefore empirically relevant.

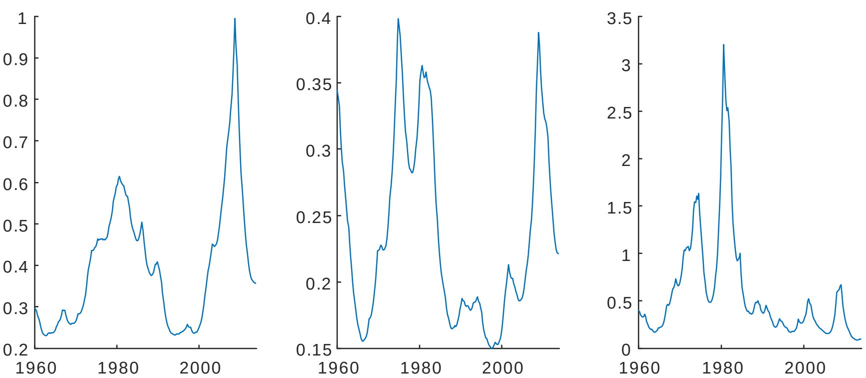


Figure 3: Time-varying volatility of the inflation equation (left panel), unemployment equation (middle panel) and interest rate equation (right panel) expressed as standard deviations.