Time-varying Parameter Vector Autoregressions

Mike Ellington

Haindorf Seminar 2020; Hejnice, Czech Republic

HUMBOLDT-UNIVERSITÄT ZU BERLIN

CHARLES UNIVERSITY

January 22, 2020

A Brief Motivation

- We know that the world is not static. So why would we impose static relationships on Financial and Economic Data?
- The revolution in computing power facilitates researchers to estimate sophisticated models *quickly*.
- As a result we have seen an influx of papers estimating time-varying parameter vector autoregressions on economic and financial data.
- Primiceri (2005) (# citations ≈1842); Cogley and Sargent (2005) (# citations ≈1500)

A Brief Motivation

- We know that the world is not static. So why would we impose static relationships on Financial and Economic Data?
- The revolution in computing power facilitates researchers to estimate sophisticated models *quickly*.
- As a result we have seen an influx of papers estimating time-varying parameter vector autoregressions on economic and financial data.
- Primiceri (2005) (# citations ≈1842); Cogley and Sargent (2005) (# citations ≈1500)

A Brief Motivation

- We know that the world is not static. So why would we impose static relationships on Financial and Economic Data?
- The revolution in computing power facilitates researchers to estimate sophisticated models *quickly*.
- As a result we have seen an influx of papers estimating time-varying parameter vector autoregressions on economic and financial data.
- Primiceri (2005) (# citations ≈1842); Cogley and Sargent (2005) (# citations ≈1500)

A Brief Motivation

- We know that the world is not static. So why would we impose static relationships on Financial and Economic Data?
- The revolution in computing power facilitates researchers to estimate sophisticated models *quickly*.
- As a result we have seen an influx of papers estimating time-varying parameter vector autoregressions on economic and financial data.
- Primiceri (2005) (# citations ≈1842); Cogley and Sargent (2005) (# citations ≈1500)

These models are becoming an "industry" standard.

Replication and Example code can be found at my GITHUB **page.** `https://github.com/mte00/TVP-VAR-Estimation-Workshop`

Roadmap

1. Bayesian Estimation
   - Linear Regression using Bayesian Methods
     - Conjugate Priors and Analytical Posterior Distributions
   - Example using Financial Data
   - Stochastic Volatility Models
   - Stochastic Volatility and Parameter Evolution

2. Time-varying Parameter VARs with Stochastic Volatility

3. Review of Paper: Ellington (2019) The Empirical Relevance of the Shadow Rate and the Zero Lower Bound

Departing from the Frequentist Approach I

In Finance and Economics, the workhorse of all data analysis is the Normal linear regression model

$$Y_t = BX_t + \epsilon_t, \quad \epsilon_t \backsim \mathrm{N}\left(0, \sigma^2\right)$$

with Y_t is a $T \times 1$ matrix of the dependent variable, X_t is a $T \times N$ matrix of deterministic terms and independent variables.

Departing from the Frequentist Approach II

The frequentist econometrician obtains estimates for B and σ^2 by maximising the likelihood function

$$\begin{aligned} \mathrm{L} &= (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y_t - BX_t)^\top (Y_t - BX_t)\right) \\ \hat{B} &= (X_t^\top X_t)^{-1} X_t^\top Y_t \\ \hat{\sigma}^2 &= \frac{\epsilon_t^\top \epsilon_t}{T} \end{aligned}$$

Departing from the Frequentist Approach III

A Bayesian Approach applies Bayes' theorem to estimation and inference.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

Departing from the Frequentist Approach IV

Suppose we have a model characterised by some likelihood function $p(y|\theta)$ and we want to learn about θ. According to Bayes theorem we have

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \propto p(\theta)p(y|\theta)$$

where $p(\theta)$ is some prior distribution for parameter vector θ.

Gibbs Sampling with Conjugate Priors I

In our Normal linear regression model, the model parameters are B and σ^2.

$$Y_t = BX_t + \epsilon_t, \quad \epsilon_t \backsim \mathrm{N}\left(0, \sigma^2\right)$$

A typical prior that is convenient for a researcher to use is one that assumes prior independence such that

$$p\left(B, \sigma^2\right) = p(B)p(\sigma^2)$$

Gibbs Sampling with Conjugate Priors II

We consider the following priors

$$B \backsim \mathrm{N}\left(B_0, V_B\right), \quad \sigma^2 \backsim \mathrm{iG}\left(v_0, S_0\right)$$

The conditional densities we need to work with are given by

$$\begin{aligned}
\left(\sigma^2 | Y_t, B\right) &\backsim \mathrm{iG}\left(v_0 + \frac{T}{2}, S_0 + \frac{1}{2}\epsilon_t^\top \epsilon_t\right) \\
\left(B | Y_t, \sigma^2\right) &\backsim \mathrm{N}\left(\hat{B}, \mathrm{D}_B\right)
\end{aligned}$$

with posterior parameters

$$\mathrm{D}_B = \left(V_B^{-1} + \sigma^{-2} X_t^\top X_t\right)^{-1}, \quad \hat{B} = \mathrm{D}_B\left(V_B^{-1} B_0 + \sigma^{-2} X_t^\top Y_t\right)$$

Gibbs Sampling with Conjugate Priors III

The Gibbs Sampler in this case allows us to take random draws of model parameters from **full conditional posterior distributions** because they define a posterior for each block conditional on all other blocks.

We know it is simple to draw from Normal and Gamma distributions and this Gibbs sampler will sequentially draw from the Normal and Gamma distributions. **Note: The sequence of draws exhibit correlation as the sequence is a Markov Chain.**

Gibbs Sampling with Conjugate Priors IV

The Algorithm:

1. Pick initial values for B_0; V_B; σ_0^2; v_0; and S_0.
2. Draw $\sigma_k^2 \backsim p(\sigma^2 | Y_t, B_{k-1})$ from inverse Gamma distribution.
3. Draw $B_k \backsim p(B | Y_t, \sigma_k^2)$ from multivariate Normal distribution.
4. Repeat steps 2 and 3 K times allowing for a number of iterations to eliminate the value of the prior.

A Note on Gibbs Sampling Markov Chains I

Gibbs Samplers, like the one we outline, are generally called Markov Chain Monte Carlo Algorithms (MCMC). We can test whether results are reliable using MCMC diagnostics.

A Note on Gibbs Sampling Markov Chains II

Geweke (1992) suggests a diagnostic which we will call CD. If a large enough number of draws are taken, estimate of $g(\theta)$ based on first half of draws should be essentially the same as estimates based on the last half.

Let K_1 denote the number of retained draws. We now split this into three subsets K_A, K_B, K_C respectively. We drop K_B when computing the diagnostics and use the first 10% of K_1 as K_A. We use the last 40% of K_1 as K_C.

A Note on Gibbs Sampling Markov Chains III

Define $\widehat{g}K_A$, $\widehat{g}K_C$ be the estimates of $\mathbb{E}\left(g(\theta)|y\right)$ using the draws in K_A, K_C respectively. Also let $\frac{\widehat{\sigma_A}}{\sqrt{K_A}}$, $\frac{\widehat{\sigma_C}}{\sqrt{K_C}}$ be the respective standard errors.

The convergence diagnostic is

$$CD = \frac{\widehat{g}K_A - \widehat{g}K_C}{\frac{\widehat{\sigma_A}}{\sqrt{K_A}} + \frac{\widehat{\sigma_C}}{\sqrt{K_C}}}, \quad CD \to \mathrm{N}\left(0, 1\right)$$

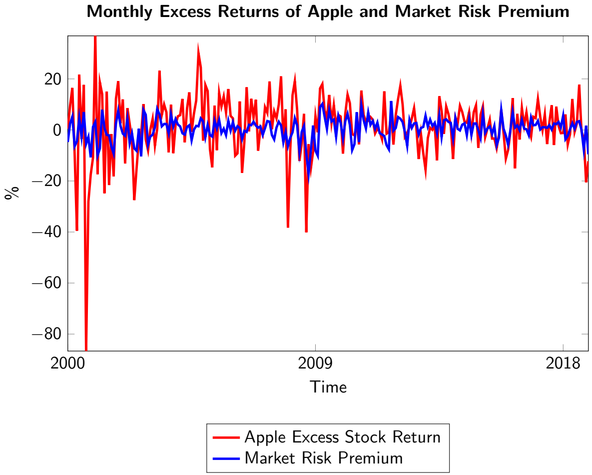
Apple Stock Returns and the Market Risk Premium I

Suppose we wish to fit the independent Normal inverse-Gamma model to Apple's excess stock return and the market risk premium. The model is:

$$
\begin{aligned}
r_t &= \alpha + \beta r_{M,t} + \nu_t, \quad \nu_t \backsim \mathrm{N}\left(0, \sigma^2\right) \\
Y &= XB + \nu
\end{aligned}
$$

here Y is a $T \times 1$ matrix, X is a $T \times N$ matrix, and B is a 2×1 matrix(vector) of coefficients.

Apple Stock Returns and the Market Risk Premium II



Apple Stock Returns and the Market Risk Premium III

We now run EXAMPLE1.m file.

Estimation procedure:

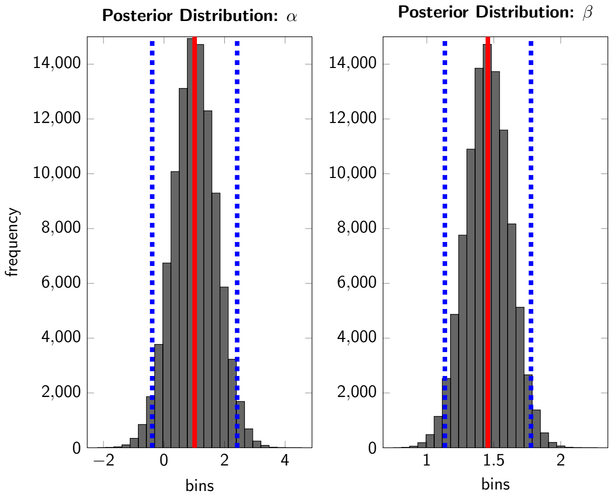
- Set $B_0 = [0, 0]^\top$, $V_B^{-1} = 0.01 \times I_N$, $v_0 = 3$, and $S_0 = 1 \times (v_0 - 1)$.
- Choose number of draws to retain (100,000) and number of draws to discard as burn-in (10,000).
- Initialise Markov Chain using OLS estimates.
- Sample B and σ^2 from their conditional posterior $K = 110,000$ times.

Apple Stock Returns and the Market Risk Premium IV

The Gibbs Sampler estimates are:

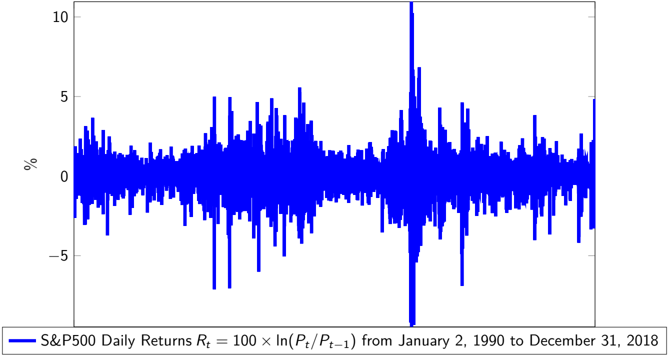
	OLS Estimates	Posterior Mean	95% Confidence Intervals	*CD*
α	1.0113 (0.72)	1.0082	[-0.3905 2.416]	0.25
β	1.456 (0.17)	1.4553	[1.1351 1.7778]	-0.26
σ^2	117.57	116.024	[96.6854 139.3304]	-0.12

Apple Stock Returns and the Market Risk Premium V



Modelling Volatility: Stochastic Volatility Models

We know that financial time-series are subject to the phenomenon of volatility clustering. Models like those we discuss previously cannot accommodate time-varying volatility.



Modelling Volatility: Stochastic Volatility Models

We consider a simple stochastic volatility model for S&P500 daily stock returns as

$$\begin{aligned} y_t &= \epsilon_t \sqrt{\exp(\ln h_t)} \\ \ln h_t &= \ln h_{t-1} + \nu_t, \quad \nu_t \backsim \mathrm{N}(0, g) \end{aligned}$$

where h_t is time-varying variance. It is clear that we have a state space model where the observation equation is nonlinear in the state variable. In order to estimate the model using Bayesian methods, we need to use a **Metropolis Hastings (MH) algorithm**.

MH algorithm of Jacquier et al. (2002) I

We use the algorithm of Jacquier et al. (2002) where we apply an independence MH algorithm at each point in time to sample from the conditional distribution of h_t; given by $f(h_t|h_{-t}, y_t)$. Since volatility evolves as a random walk, knowledge of h_{t+1}, h_t contains all relevant information for h_t.

MH algorithm of Jacquier et al. (2002) II

The conditional distribution has the following density

$$
\begin{aligned}
f(h_t|h_{t-1}, h_{t+1}, y_t) &= h_t^{-0.5} \exp\left(\frac{-y_t^2}{2h_t}\right) \times h_t^{-1} \exp\left(\frac{-(\ln h_t - \mu)^2}{2\sigma_h}\right) \\
\mu &= \frac{1}{2}(\ln h_{t+1} + \ln h_{t-1}) \\
\sigma_h &= \frac{g}{2}
\end{aligned}
$$

MH algorithm of Jacquier et al. (2002) III

Sampling is carried out on a date-by-date application of the MH algorithm with the candidate density and acceptance probability

$$
\begin{aligned}
q\left(\Phi^{k+1}\right) &= h_t^{-1} \exp\left(\frac{-(\ln h_t - \mu)^2}{2\sigma_h}\right) \\
\alpha &= \min\left(\frac{\pi(\Phi^{k+1})/q(\Phi^{k+1}/\Phi^{k})}{\pi(\Phi^{k})/q(\Phi^{k}/\Phi^{k+1})}, 1\right) \\
&= \min\left(\frac{h_{t,new}^{-0.5}\exp\left(\frac{-y_t^2}{2h_{t,new}}\right)}{h_{t,old}^{-0.5}\exp\left(\frac{-y_t^2}{2h_{t,old}}\right)}\right)
\end{aligned}
$$

MH algorithm of Jacquier et al. (2002) IV

Of course, μ requires knowledge of h_{t+1}, h_{t-1}. Therefore Jacquier et al. (2002) suggest the prior $\ln h_0 \backsim \mathrm{N}\left(\bar{\mu}, \bar{\sigma}\right)$ which has a posterior of

$$
\begin{aligned}
f\left(h_0|h_1\right) &= h_0^{-0.5}\exp\left(\frac{-(\ln h_0 - \mu_0)^2}{2\sigma_0}\right) \\
\sigma_0 &= \bar{\sigma}g/(\bar{\sigma}+g) \\
\mu_0 &= \sigma_0\left(\bar{\mu}/\bar{\sigma}+\ln h_1/g\right)
\end{aligned}
$$

MH algorithm of Jacquier et al. (2002) V

For the final value (with $t = T$) the modified candidate generating density is

$$\begin{aligned} q(\Phi^{k+1}) &= h_t^{-1} \exp\left(\frac{-(\ln h_t - \mu)^2}{2\sigma_h}\right) \\ \mu &= \ln h_{t-1} \\ \sigma_h &= g \end{aligned}$$

MH algorithm of Jacquier et al. (2002) VI

1 Obtain starting values for h_t, $t = 0, \dots, T$ as ϵ_t^2 and set priors for $\bar{\mu}$, $\bar{\sigma}$ (Perhaps $\bar{\mu}$ is the OLS estimate of $\ln \epsilon_t$ and $\bar{\sigma}$ set large to reflect uncertainty in initial condition)

2 for $t = 0$, sample initial value of $h_t = h_0$

2b for $t = 1–T - 1$, draw new value for h_t from candidate and compute α. Then draw $u \backsim \mathrm{U}(0,1)$. If $\alpha > u$ set $h_t = h_{t,new}$. Otherwise retain old draw.

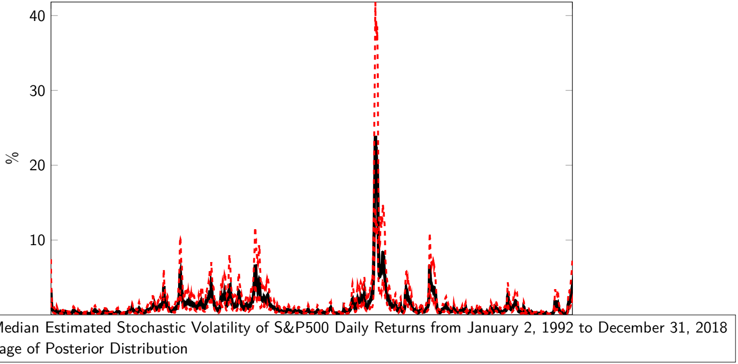
2c for $t = T$ compute from modified candidate, compute alpha, and if $\alpha > u$ set $h_t = h_{t,new}$. Otherwise retain old draw.

3 Given draw for h_t compute residuals $\nu_t = \ln h_t - \ln h_{t-1}$. Draw g from iG distribution with scale $\frac{\nu_t^\top \nu_t + g_0}{2}$ and DoF $\frac{T+\nu_0}{2}$.

4 Repeat steps 2–3 K times and retain K - burn draws of h_t, g as an approximation to the marginal posterior distributions.

Modelling Volatility: Stochastic Volatility Models

We now run EXAMPLE2.m file and obtain the posterior distribution for the stochastic volatility of S&P500 market returns from January 1992–December 2018.



Allowing for Parameter Variation I

The simple stochastic model assumes no mean or autoregressive dynamics. If the true DGP possesses these factors, it is important to account for them in the modelling process.

$$
\begin{aligned}
y_t &= a_t + b_t y_{t-1} + \epsilon_t \sqrt{\exp(\ln h_t)} \\
B_t &= B_{t-1} + \eta_t, \;\; \eta_t \backsim \mathrm{N}\left(0, Q\right) \\
\ln h_t &= \ln h_{t-1} + \nu_t, \;\; \nu_t \backsim \mathrm{N}\left(0, g\right)
\end{aligned}
$$

where $B_t = \{a_t, b_t\}$.

Allowing for Parameter Variation II

We now use the Carter Kohn algorithm in order to sample B_t. This involves using a Kalman filter.

In order to calibrate the initial conditions, we use a training sample over the first T_0 observations by estimating an linear regression to obtain $B_0 = \{a_0, b_0\}$. We also need a prior for Q, and of course the stochastic volatility.

Allowing for Parameter Variation III

The prior for Q is inverse Wishart such that $P(Q) \backsim \mathrm{iW}(Q_0, T_0)$.

Prior for Q is crucial: since this induces how much time-variation we have in the model.

Allowing for Parameter Variation IV

Kalman Filter: Starts with initial $B_{t-1|t-1} = B_{0|0}$ and $P_{t-1|t-1} = P_{0|0}$ and then iterates through the estimation sample through the following steps:

1. Prediction step: $B_{t|t-1} = B_{t|t-1}$ and then computes the variance using $P_{t|t-1} = P_{t-1|t-1} + Q$
2. Compute fitted value of y_t at t and calculates prediction error $\mathrm{e}_{t|t-1} = y_t - X_t B_{t|t-1}$
3. Compute variance of prediction error $\mathrm{f}_{t|t-1} = X_t P_{t|t-1} X_t^\top + h_t$. Then update by calculating Kalman gain $\mathrm{K}_t = P_{t|t-1} X_t^\top \mathrm{f}_{t|t-1}^{-1}$
4. Update estimate using information in the prediction error $B_{t|t} = B_{t|t-1} + \mathrm{K}_t \mathrm{e}_{t|t-1}$, and also update variance $P_{t|t} = P_{t|t-1} - \mathrm{K}_t X_t P_{t|t-1}$.

Allowing for Parameter Variation V

Carter Kohn Algorithm: The conditional distribution of $\mathbf{B}_T = \{B_1, B_2, \ldots, B_T\}$ is

$$p(\mathbf{B}_T) = p(B_T, y_t) \prod_{t=1}^{T-1} p(B_t | B_{t+1}, y_t)$$

The Kalman filter pins down the first element on the RHS which is $\backsim \mathrm{N}\left(\mathbf{B}_T, P_T\right)$

Allowing for Parameter Variation VI

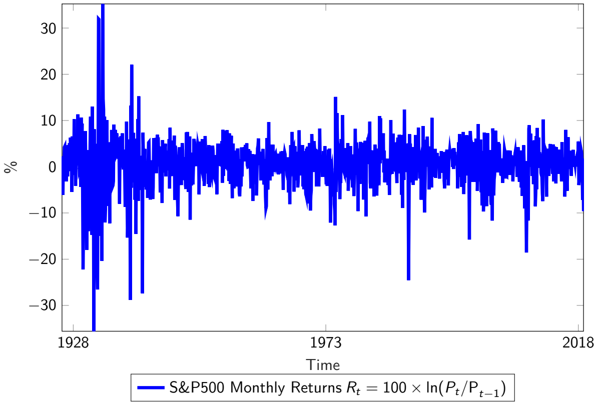
The remaining elements are obtained via backward recursion as B_t is conditionally Normal we have

$$\begin{aligned} B_{t|t+1} &= P_{t|t} P_{t+1|t}^{-1} (B_{t+1} - B_t) \\ P_{t|t+1} &= P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t} \end{aligned}$$

which yields for every $t \in \{T-1, \ldots, 1\}$.

Allowing for Parameter Variation VII

We now combine the Carter Kohn algorithm with our MH algorithm and conduct our analysis on the Monthly S&P500 market return from January 1926–December 2018.



Allowing for Parameter Variation VIII

Running EXAMPLE3.m consists of the following steps:

1. Set iW prior for Q. $Q_0 = m \times Q_{OLS} \times T_0$. Here m is some scaling parameter, Q_{OLS} is the covariance matrix of B_0 and $T_0 = 240$ is the number of observations in the training sample.
2. Obtain starting values for h_t, $t = 0, \ldots, T$ as ϵ_t^2 and set priors for $\bar{\mu}$, $\bar{\sigma}$ (Perhaps $\bar{\mu}$ is the OLS estimate of $\ln \epsilon_t$ and $\bar{\sigma}$ set large to reflect uncertainty in initial condition)
3. for $t = 0$, sample initial value of h_t=h_0

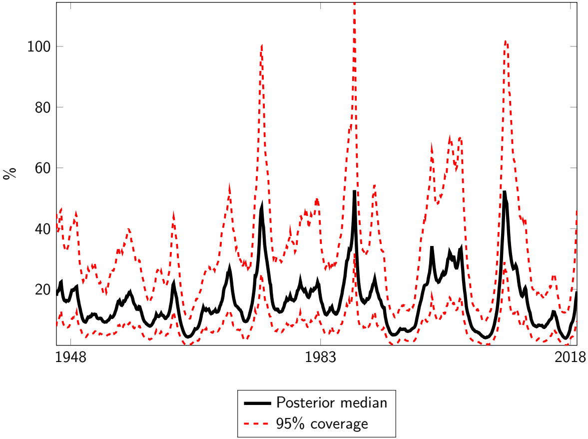
3b for $t = 1$–$T - 1$, draw new value for h_t from candidate and compute α. Then draw $u \backsim \mathrm{U}(0, 1)$. If $\alpha > u$ set $h_t = h_{t,new}$. Otherwise retain old draw.

3c for $t = T$ compute from modified candidate, compute alpha, and if $\alpha > u$ set $h_t = h_{t,new}$. Otherwise retain old draw.

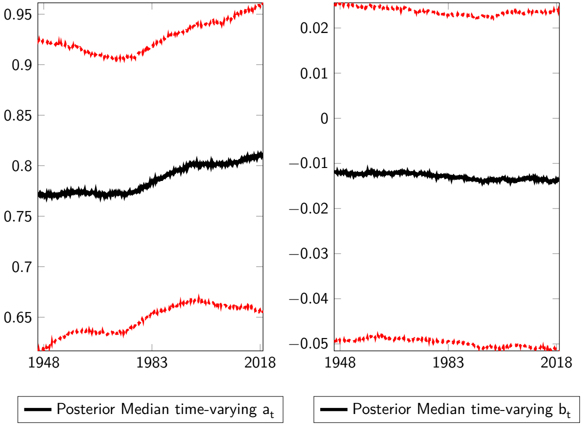
Allowing for Parameter Variation IX

4. Given draw for h_t compute residuals $\nu_t = \ln h_t - \ln h_{t-1}$. Draw g from iG distribution with scale $\frac{\nu_t^\top \nu_t + g_0}{2}$ and DoF $\frac{T+\nu_0}{2}$.
5. Conditional on h_t, Q, sample B_t using Carter Kohn algorithm.
6. Conditional on B_t, sample Q from iW distribution with scale matrix $(B_t - B_{t-1})^\top (B_t - B_{t-1}) + Q_0$ and degrees of freedom $T_0 + T$.
7. Repeat steps 3–6 K times and retain K - burn of the draws of model parameters as an approximation of the marginal posterior distributions.

Allowing for Parameter Variation X



Allowing for Parameter Variation XI



TVP VAR with Stochastic Volatility I

Consider the following TVP VAR model with p lags and N variables:

$$\mathbf{Y}_t = \beta_{0,t} + \beta_{1,t}\mathbf{Y}_{t-1} + \beta_{2,t}\mathbf{Y}_{t-2} + \mathbf{e}_t \equiv \mathbf{X}_t^\top \mathbf{B}_t + \mathbf{e}_t$$

where Y_t is a $T \times N$ matrix of dependent variables; $\mathbf{X}_t^\top$ contains lagged values of $\mathbf{Y}_t$ and a constant.

$$\mathbf{B}_t = \mathbf{B}_{t-1} + u_t, \quad u_t \backsim N(0, \mathbf{Q})$$

$\mathbf{Q}$ is a full matrix allowing parameters across equations to be correlated.

TVP VAR with Stochastic Volatility II

The innovations of the measurement equation, $\mathbf{e}_t$ are Normal with zero mean and time-varying covariance matrix Ω_t which is factored as

$$\begin{aligned}
\Omega_t &= \mathbf{A}_t^{-1}\mathbf{H}_t(\mathbf{A}_t^{-1})^\top \\
\ln \mathbf{h}_{i,t} &= \ln \mathbf{h}_{i,t-1} + \eta_t, \quad \eta_t \backsim N(0, \mathbf{Z_h}) \\
\mathbf{a}_t &= \mathbf{a}_{t-1} + \zeta_t, \quad \zeta_t \backsim N(0, \mathbf{S})
\end{aligned}$$

The innovations in the model, collected in the diagonal matrix $\mathbf{V}$, are jointly Normal and the structural shocks, $\mathbf{v}_t$ are such that, $\mathbf{e}_t \equiv \mathbf{A}_t^{-1}\mathbf{H}_t^{\frac{1}{2}}\mathbf{v}_t$.

TVP VAR with Stochastic Volatility III

Note: $\mathbf{S}$ is a block diagonal matrix, which implies that the non–zero and non–unit elements of $\mathbf{A}_t$ that belong to different rows evolve independently. This is a simplifying assumption that permits estimation of $\mathbf{A}_t$ equation by equation (Primiceri, 2005).

TVP VAR with Stochastic Volatility IV

A typical prior specification: Estimates simple VAR using OLS and sets initial conditions using these estimates.

- $\mathbf{B}_0 \backsim \mathrm{N}\left(\hat{\mathbf{B}}_{OLS}, 4 \times V(\hat{\mathbf{B}}_{OLS})\right)$
- Let $C^\top C = \Sigma_{OLS}$ and $\tilde{C}$ be a matrix that divides all elements in each column of C by the corresponding diagonal element.

$$\begin{aligned} \ln \mathbf{h}_0 &\backsim \mathrm{N}\left(\ln \mu_0, 10 \times I_N\right) \\ \mathbf{a}_0 &\backsim \mathrm{N}\left(\tilde{\mathbf{a}}_0, \tilde{V}(\tilde{\mathbf{a}}_0)\right) \end{aligned}$$

here $\tilde{\mathbf{a}}_0$ collects the elements below the main diagonal of $\tilde{C}$.
$\tilde{V}(\tilde{\mathbf{a}}_0) = \mathrm{diag}\left(10 \times |\, [\tilde{\mathbf{a}}_0]_{i,i} \,|\right)$

TVP VAR with Stochastic Volatility V

- $\mathbf{Q} \backsim \text{iW}\left(\mathbf{Q}_0^{-1}, T_0\right)$, with $\mathbf{Q}_0 = (1 + \dim(\mathbf{B}_t) \times V(\hat{\mathbf{B}}_{OLS}) \times m$
- The blocks of $\mathbf{S}$ are $\mathbf{S}_i \backsim \text{iW}\left(\mathbf{S}_{0,i}, (1 + \dim(\mathbf{S}_i)\right)$. For example if N=3 then we have

$$\begin{aligned} \mathbf{S}_{0,1} &= 10^{-3} \times |\tilde{\mathbf{a}}_{0,11}| \\ \mathbf{S}_{0,2} &= 10^{-3} \times \text{diag}\left([|\tilde{\mathbf{a}}_{0,21}|, \ \tilde{\mathbf{a}}_{0,31}|]\right) \end{aligned}$$

- The variances of the stochastic volatility innovations follow an iG distribution (Cogley and Sargent, 2005) such that: $\mathbf{Z}_{\mathbf{h}_{i,i}} \backsim \text{iG}\left(\frac{10^{-4}}{2}, \frac{1}{2}\right)$

TVP VAR with Stochastic Volatility VI

Posterior Simulation:

- *Draw elements of* $\mathbf{B}_t$*:* using Carter Kohn Algorithm.
- *Draw elements of* $\mathbf{a}_t$*:* We can use Carter Kohn algorithm to get $\mathbf{A}_t$. If N=3, conditional on $\mathbf{Y}^T$, $\mathbf{B}^T$ and $\mathbf{H}^T$ note the TVP VAR can be written as

$$\begin{aligned}\mathbf{A}_t\tilde{\mathbf{Y}}_t \equiv \mathbf{A}_t(\mathbf{Y}_t - \mathbf{X}_t^{'}\mathbf{B}_t) &= \mathbf{A}_t\mathbf{e}_t \equiv \mathbf{v}_t \\ \text{Var}(\mathbf{v}_t) &= \mathbf{H}_t\end{aligned}$$

with $\tilde{\mathbf{Y}}_t \equiv [\tilde{\mathbf{Y}}_{1,t},\ \tilde{\mathbf{Y}}_{2,t},\ \tilde{\mathbf{Y}}_{3,t}]^{'}$ and

$$\begin{aligned}\tilde{\mathbf{Y}}_{1,t} &= \mathbf{v}_{1,t} \\ \tilde{\mathbf{Y}}_{2,t} &= -\mathbf{a}_{21,t}\tilde{\mathbf{Y}}_{1,t} + \mathbf{v}_{2,t} \\ \tilde{\mathbf{Y}}_{3,t} &= -\mathbf{a}_{31,t}\tilde{\mathbf{Y}}_{1,t} - \mathbf{a}_{32,t}\tilde{\mathbf{Y}}_{2,t} + \mathbf{v}_{3,t}\end{aligned}$$

TVP VAR with Stochastic Volatility VII

- *Drawing elements of* $\mathbf{H}_t$ Conditional on $\mathbf{Y}^T$, $\mathbf{B}^T$ and $\mathbf{a}^T$, the orthogonal innovations u_t, $\text{Var}(\mathbf{e}_t) = \mathbf{H}_t$ are observable. Following Jacquier et al. (2002) the stochastic volatilities, $\mathbf{h}_{i,t}$'s, are sampled element by element.

- *Drawing the hyperparameters* Conditional on Y^T, $\mathbf{B}^T$, $\mathbf{H}_t$ and $\mathbf{a}^T$, the innovations in $\mathbf{B}_t$, $\mathbf{a}_t$ and $\mathbf{h}_{i,t}$'s are observable, which allows one to draw the elements of $\mathbf{Q}_t = \mathbf{Q}$, $\mathbf{S}_1$, $\mathbf{S}_2$ and the $\mathbf{Z}_{\mathbf{h},i,i}$ from their respective distributions.

Review of Ellington (2019) I

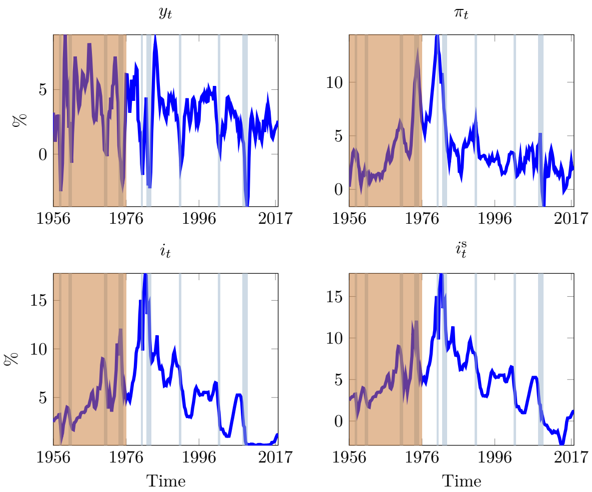
Abstract:

This paper tests the statistical and economic differences in monetary policy implications using the shadow rate proposed in Wu and Xia (2016). Time-varying coefficient VAR models are fitted to US data from 1966–2017 that reveal stark economic and statistical differences in the structural implications of monetary policy that arise when replacing conventional interest rates with their shadow rate counterparts. Results provide strong support for utilising shadow rates within models of monetary policy under a binding zero lower bound constraint.

Review of Ellington (2019) II

- N=3, p=2 TVP VARs for $\mathbf{Y}_t \equiv [y_t, \pi_t, \mathrm{r}_t]$ with $\mathrm{r}_t = \{i_t, i_t^s\}$.
- Modify the traditional OLS prior by estimating a Bayesian VAR on the training sample and using posterior mean and variances to calibrate the initial conditions of the model.
- Map reduced-from VAR to structural model using Arias et al. (2018) and Rubio-Ramirez et al. (2010).
- Conduct structural analysis in a **generalised framework**.

Review of Ellington (2019) III



Review of Ellington (2019) IV

Model Evaluation: Bayesian DIC stats (similar to AIC)

Model:	$r = i_t$ DIC	$r = i_t^s$ DIC
TVP VAR $\mathbf{Q}_t = \mathbf{Q}$, with stochastic volatility	88.32	116.13
TVP VAR $\mathbf{Q}_t = \mathbf{Q}_t$, with stochastic volatility	99.76	148.66
Two-regime MS-VAR	190.19	216.20
TVP VAR constant covariance matrix	1034.41	1091.20
BVAR with time-varying covariance matrix	161.92	171.99
BVAR with stochastic volatility	170.72	181.21
constant coefficient BVAR	1758.59	1762.55

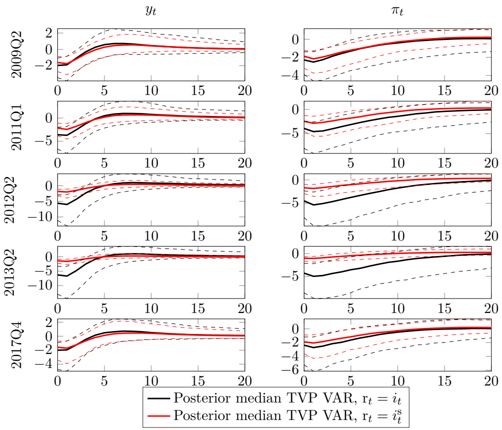
Review of Ellington (2019) V

The structural model is mapped using contemporaneous sign restrictions:

	$\mathbf{v}_t^{\mathbf{s}}$	$\mathbf{v}_t^{\mathbf{d}}$	$\mathbf{v}_t^{\mathbf{mp}}$
y_t	$\geq$	$\geq$	$\leq$
π_t	$\leq$	$\geq$	$\leq$
r_t	x	$\geq$	$\geq$

Review of Ellington (2019) VI

Monetary Policy Shocks: Impulse Response Functions



Review of Ellington (2019) VII

Structural Monetary Policy Rules:

$$\Omega_t = \bar{\mathbf{P}}_t^{-1}\bar{\mathbf{D}}_t\bar{\mathbf{D}}_t'(\bar{\mathbf{P}}_t^{-1})'$$

where $\bar{\mathbf{P}}_t$ and $\bar{\mathbf{D}}_t$ are 3×3 matrices of the form

$$\bar{\mathbf{P}}_t = \begin{bmatrix} 1 & -\mathbf{p}_t^{y,\pi} & -\mathbf{p}_t^{y,\mathrm{r}} \\ -\mathbf{p}_t^{\pi,y} & 1 & -\mathbf{p}_t^{\pi,\mathrm{r}} \\ -\mathbf{p}_t^{\mathrm{r},y} & -\mathbf{p}_t^{\mathrm{r},\pi} & 1 \end{bmatrix}, \quad \bar{\mathbf{D}}_t = \begin{bmatrix} \mathbf{d}_{y,t} & 0 & 0 \\ 0 & \mathbf{d}_{\pi,t} & 0 \\ 0 & 0 & \mathbf{d}_{\mathrm{r},t} \end{bmatrix}$$

Review of Ellington (2019) VIII

Therefore the structural representation of our models may be written as

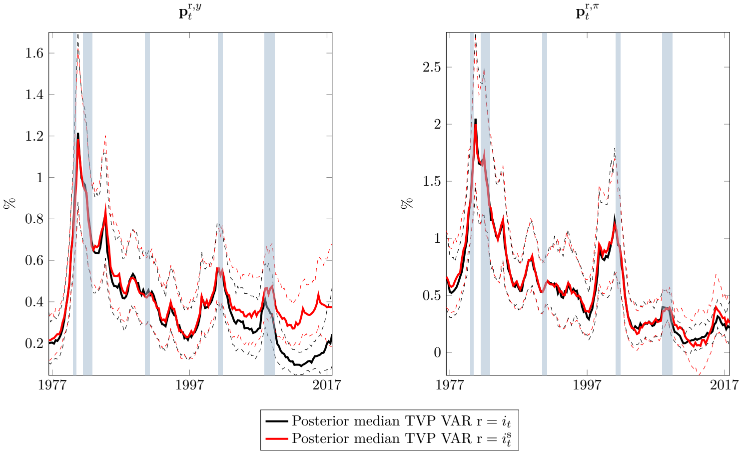
$$\bar{\mathbf{P}}_t \mathbf{Y}_t = \mathbf{G}_{0,t} + \mathbf{G}_{1,t}\mathbf{Y}_{t-1} + \mathbf{G}_{2,t}\mathbf{Y}_{t-2} + \bar{\mathbf{D}}_t \mathbf{v}_t$$

The third row of delivers the structural monetary policy rule of the model.

$$\begin{aligned} \mathrm{r}_t &= \mathbf{g}_0^{\mathrm{r},t} + \mathbf{p}_t^{i,y} y_t + \mathbf{g}_{1,t}^{\mathrm{r},y} y_{t-1} + \mathbf{g}_{2,t}^{\mathrm{r},y} y_{t-2} + \mathbf{p}_t^{\mathrm{r},\pi}\pi_t \\ &+ \mathbf{g}_{1,t}^{\mathrm{r},\pi}\pi_{t-1} + \mathbf{g}_{2,t}^{\mathrm{r},\pi}\pi_{t-2} + \mathbf{g}_{1,t}^{\mathrm{r},\mathrm{r}}\mathrm{r}_{t-1} + \mathbf{g}_{2,t}^{\mathrm{r},\mathrm{r}}\mathrm{r}_{t-2} \end{aligned}$$

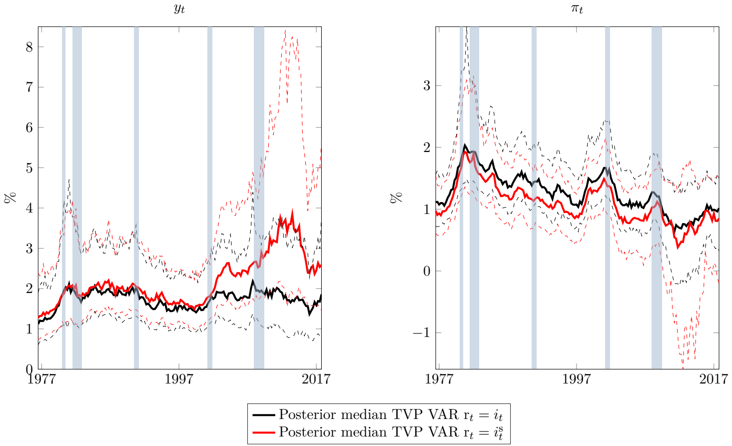
Review of Ellington (2019) IX

Monetary Policy Rules: Structural Impact Coefficients



Review of Ellington (2019) X

Monetary Policy Rules: Long-run Coefficients



Review of Ellington (2019) XI

Conclusions

- Model evaluation provides strong support for TVP VAR.
- Statistically significant and economically meaningful differences in the response of economic activity throughout time.
- Monetary policy rules shows response of variables differs when using shadow rates (Wu and Xia, 2016).
- Policy makers should pay attention to shadow rates of interest when the economy's short-term interest rate has a binding zero lower bound constraint.

References I

Arias, J., Rubio-Ramirez, J., and Waggoner, D. (2018). Inference Based on SVARS Identified with Sign and Zero Restrictions: Theory and Applications. *Econometrica*, 86(2):685–720.

Cogley, T. and Sargent, T. J. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Ellington, M. (2019). The Empirical Relevance of the Shadow Rate and the Zero Lower Bound. *Available from: SSRN*.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-based Approaches to the Calculations of Posterior Moments. *Bayesian statistics*, 4:641–649.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 20(1):69–87.

References II

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

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies*, 77(2):665–696.

Wu, J. C. and Xia, F. D. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.