Efficient Estimation of Structural VARMAs with Stochastic Volatility

Eric Eisenstat [1]

[1] The University of Queensland

Vienna University of Economics and Business Seminar
18 April 2018

Motivation

Macroeconomists and policy makers use time series models for structural inference as an alternative to DSGEs.

- robust to explicit theoretical modeling assumptions
- guide development of better DSGEs

Main instrument for inference is impulse response functions and forecast error variance decompositions.

- closely linked to the infinite MA representation (i.e. Wold representation) for multivariate time series data

Key challenge is identifying impulse response functions / structural shocks.

Motivation

Identification is not only about static orthogonal rotations.

If errors are Gaussian white noise, then the characteristic roots of the infinite MA representation are not identified.

The two identification problems can be treated separately.

DSGEs and VARMAs

The equilibrium of a general class of DSGE models can be represented by a finite order VARMA (Fernandez-Villaverde, *et al.*, 2007, AER).

If the characteristic MA roots are all outside the unit circle, the representation is fundamental.

- then, we can obtain a VAR approximation by inverting the MA part and truncating lags (SVAR approach)

Many DSGEs (e.g. permanent income model of Hansen et al., 1991) lead to clearly non-fundamental representations, where at least one root is inside the unit circle.

But theory is rarely precise on which non-fundamental representation (i.e. combination of roots inside the unit circle) is appropriate.

Summarizing the Issue

Fundamental and non-fundamental representations are observationally equivalent under Gaussian errors.

Standard SVAR approach only identifies fundamental structural shocks, which is often inconsistent with theory.

Lippi and Reichlin (1994, JoE) argue IRFs and FEVDs should be based on set identification.

- difficult with SVARs because information about characteristic MA roots is lost when VAR lags are truncated
- even with VARMAs, can be computationally infeasible

Non-Gaussian Errors

Observational equivalence can be eliminated with non-Gaussian errors.

If errors are i.i.d. non-Gaussian, then infinite MA representations are unique up to scaling, order of shocks, and time shifts (Chan *et al.*, 2006, Biometrika, Gouriéroux *et al.*, 2017, WP).

Therefore, specifying i.i.d. non-Gaussian errors simultaneously identifies both:

1. the fundamental or one of the non-fundamental representations,
2. the structural representation (i.e. orthogonal rotations of errors no longer observationally equivalent).

Could this be a viable solution?

Problems With Non-Gaussian Errors

Specifying i.i.d. non-Gaussian errors entails a number of practical problems:

1. Different distributions lead to different representations being identified, so the question becomes which non-Gaussian distribution should be employed?
2. VARMAs with non-Gaussian distribution are highly non-linear in parameters, and no feasible computation methodology is currently available.
3. Statistically identified structural shocks are difficult to interpret.

Proposed Alternative

We argue that an alternative approach is to model errors as conditionally Gaussian by introducing heteroskedastic errors.

A large literature (Clark and Ravazzolo, 2014, JAE; Carriero *et al.*, 2016, JBES, Chan and Eisenstat, 2018, JAE) has consistently demonstrated that allowing for stochastic volatility in VARs is crucial in modeling macroeconomic time series.

Therefore, it is a natural extension to model VARMAs with Gaussian errors and stochastic volatility.

We prove that under mild regularity conditions, we thus can obtain infinite MA representations that are unique up to static orthogonal rotations.

Benefits of VARMAs With Stochastic Volatility

The main advantages the proposed approach are:

1. conditionally Gaussian errors means computation is non-trivial but feasible: we develop an efficient MCMC algorithm in a Bayesian state-space setting,
2. characteristic MA roots are identified, but not the structural representation,
3. structural shocks can be identified using standard theory-driven restrictions as in the SVAR approach,
4. alternatives stochastic volatility specifications can be evaluated using Bayesian model comparison methods.

Structural VARMAs

Consider an $n \times 1$ vector of observations $\mathbf{y}_t$, modeled as a VARMA(p, q) process:

$$\mathbf{B}(L)\mathbf{y}_t = \mathbf{A}(L)\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \mathbf{I}_n).$$

where $\mathbf{A}(L)$ and $\mathbf{B}(L)$ are polynomial matrices in the back shift operator L and $\mathbf{B}_0 = \mathbf{I}_n$.

Assuming $\mathbf{B}(z) \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$, the infinite MA representation is

$$\mathbf{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\Psi}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L). \tag{1}$$

The identification issue we focus on arises from the fact that there exist many $\mathbf{D}(z)$ such that $\mathbf{D}(z)\mathbf{D}(z^{-1})' = \mathbf{I}_n$ and

$$\mathbf{y}_t = \widetilde{\boldsymbol{\Psi}}(L)\widetilde{\boldsymbol{\varepsilon}}_t, \qquad \widetilde{\boldsymbol{\Psi}}(L) = \boldsymbol{\Psi}(L)\mathbf{D}(L), \qquad \widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{D}(L^{-1})'\boldsymbol{\varepsilon}_t, \tag{2}$$

where L^{-1} is the forward shift operator and $\widetilde{\boldsymbol{\varepsilon}}_t \sim \mathcal{N}(0, \mathbf{I}_n)$.

Observational Equivalence

If $\mathbf{D}(z) = \mathbf{D}$, $\mathbf{D}\mathbf{D}' = \mathbf{I}_n$, we obtain the standard case where (1) is observationally equivalent to (2) due static orthogonal rotations.

Even if $\boldsymbol{\Psi}_0 = \widetilde{\boldsymbol{\Psi}}_0$, there exist $\mathbf{D}(z)$ such that $\det \boldsymbol{\Psi}(z) = 0$ implies either $\det \widetilde{\boldsymbol{\Psi}}(z) = 0$ or $\widetilde{\boldsymbol{\Psi}}(z^{-1}) = 0$ and (2) is an observationally equivalent VARMA(p, q) to the one in (1).

- $\mathbf{D}(z)$ is a Blaschke matrix that "flips" some combination of roots of $\det \boldsymbol{\Psi}(z)$ inside /outside the unit circle
- there are up to 2^{nq} countable, observationally equivalent representations generated this way

SVAR approach only considers the fundamental representation and ignores the (up to $2^{nq} - 1$) non-fundamental ones.

- this is inconsistent with theory in many settings

Proposed Solutions to Observational Equivalence

Lippi and Reichlin (1994, JoE) suggest set identification of IRFs and FEVDs based on all fundamental and non-fundamental representations.

- becomes computationally infeasible as n, q increase
- inference typically imprecise

If structural shocks are i.i.d. not Gaussian then all observational equivalence vanishes: (1) can be made unique with trivial restrictions on $\mathbf{\Psi}_0$ (i.e., nonzero, non-decreasing diagonal elements).

If reduced form innovations are conditionally Gaussian with heteroskedastic errors, then the reduced form representation is identified up to static orthogonal rotations, and the structural representation is identified by theory-driven restrictions on $\mathbf{\Psi}_0$.

Multivariate Volatility Process

Let $\{\mathbf{\Sigma}_t : t \in \mathbb{Z}\}$ be a stochastic process where:

- each $\mathbf{\Sigma}_t$ a positive definite symmetric matrix,
- $\|\mathbf{\Sigma}_t\| \leq \varsigma < \infty$ almost surely for all $t \in \mathbb{Z}$,
- $\{\mathbf{\Sigma}_t\}$ is a weakly stationary process with an absolutely integrable inverse spectral density.

Upper bound and stationarity are not necessary and can be replaced other regularity conditions.

- Upper bound allows for clearer proof and is not restrictive in practice since it can be arbitrarily large (we only need it to be finite).

Gaussian Scale Mixture Process

The Gaussian scale mixture process (GSMP) is given by $\{\mathbf{u}_t : t \in \mathbb{Z}\}$ such that

$$(\mathbf{u}_t \,|\, \{\mathbf{\Sigma}_t\}) \overset{iid}{\sim} \mathcal{N}(0, \mathbf{\Sigma}_t). \tag{3}$$

Marginally of $\{\mathbf{\Sigma}_t\}$, $\mathbf{u}_t$ and $\mathbf{u}_{t-j}$ are mean independent for all j.

Marginally of $\{\mathbf{\Sigma}_t\}$, $\mathbf{u}_t$ and $\mathbf{u}_{t-j}$ are stochastically independent if and only if $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_{t-j}$ are stochastically independent.

For any non-singular $\mathbf{D}$, $\widetilde{u}_t = \mathbf{D}\mathbf{u}_t$ is also a GSMP.

VARMAs With GSMP Errors

Consider the reduced form VARMA(p, q):

$$\mathbf{y}_t = \mathbf{B}(L)^{-1}\mathbf{\Theta}(L)\mathbf{u}_t = \mathbf{\Phi}(L)\mathbf{u}_t, \qquad (\mathbf{u}_t \,|\, \mathbf{\Sigma}_t) \sim \mathcal{N}(0, \mathbf{\Sigma}_t), \tag{4}$$

where $\{\mathbf{\Sigma}_t\}$ satisfies the previous assumptions and $\mathbf{\Theta}_0 = \mathbf{I}_n$ and $\mathbf{\Phi}_0 = \mathbf{I}_n$.

A structural form is obtained by:

$$\mathbf{y}_t = \mathbf{B}(L)^{-1}\mathbf{\Theta}(L)\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{Q}_t\boldsymbol{\varepsilon}_t = \mathbf{\Psi}_t(L)\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \mathbf{I}_n),$$

where $\mathbf{Q}_t\mathbf{Q}_t' = \mathbf{I}_n$.

Does there exist an alternative VARMA(p, q) process

$$\widetilde{\mathbf{y}}_t = \mathbf{B}(L)^{-1}\widetilde{\mathbf{\Theta}}(L)\widetilde{\mathbf{u}}_t = \widetilde{\mathbf{\Phi}}(L)\widetilde{\mathbf{u}}_t, \qquad (\widetilde{\mathbf{u}}_t \,|\, \widetilde{\mathbf{\Sigma}}_t) \sim \mathcal{N}(0, \widetilde{\mathbf{\Sigma}}_t), \tag{5}$$

such that $\{\mathbf{y}_t\}$ and $\{\widetilde{\mathbf{y}}_t\}$ are observationally equivalent?

Uniqueness Result

Theorem

Assume $\{\boldsymbol{\Sigma}_t\}$ *satisfies regularity conditions. The process* $\{\widetilde{\mathbf{y}}_t\}$ *given by* (5) *is observationally equivalent to the process* $\{\mathbf{y}_t\}$ *given by* (4) *if and only if* $\{\widetilde{\mathbf{u}}_t\} \stackrel{d}{=} \{\mathbf{u}_t\}$ *and* $\widetilde{\boldsymbol{\Phi}}(L) = \boldsymbol{\Phi}(L)$*, where* $\stackrel{d}{=}$ *denotes equivalence in distribution.*

Two comments regarding the proof:

1. Main challenge: $\mathbf{u}_t$ not stochastically independent.
2. Key insight: if $\widetilde{\boldsymbol{\Phi}}(z) \neq \boldsymbol{\Phi}(z)$, then $\mathbf{D}(z) = \widetilde{\boldsymbol{\Phi}}(z)^{-1}\boldsymbol{\Phi}(z)$ yields $\widetilde{\mathbf{u}}_t = \mathbf{D}(L)\mathbf{u}_t$ such that $\widetilde{\mathbf{u}}_t$ is not mean-independent of $\widetilde{\mathbf{u}}_{t-j}$ for some j; therefore, $\widetilde{\mathbf{u}}_t$ is not a GSMP.

Implications

The data uniquely identifies either a fundamental or one of the non-fundamental representations.

VARs with time-varying volatilities (i.e. Sims and Zha, 2006, AER; Chan and Eisenstat, 2018, JAE) are over-identified because they enforce fundamentalness.

- this holds even for models with log-volatilities modeled as random walks (i.e. non-stationary volatility process)
- can be shown to hold for TVP-VARs with time-varying coefficients (though not a trivial extension)

We can test if the fundamentalness restriction is supported by the data.

Empirical Specification

We consider a Bayesian state-space model given by:

$$\begin{aligned}\mathbf{B}_0\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} \\ + \boldsymbol{\Theta}_1\mathbf{u}_{t-1} + \cdots + \boldsymbol{\Theta}_q\mathbf{u}_{t-q}, \quad (\mathbf{u}_t \,|\, \boldsymbol{\Sigma}_t) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_t), \\ (\boldsymbol{\Sigma}_t^{-1} \,|\, \boldsymbol{\Sigma}_{t-1}) \sim \mathcal{IW}\left(\nu, \frac{1}{\nu}\boldsymbol{\Sigma}_{t-1}^{-\delta}\right),\end{aligned}$$

where $\nu > n$ and $|\delta| < 1$.

$\mathcal{W}(\cdot)$ denotes the Wishart distribution; the state equation defines the Wishart process of Philipov and Glickman (2006, JBES).

Adding the constraint $\|\boldsymbol{\Sigma}_t\| < \varsigma$ ensures that this volatility process satisfies the regularity conditions and has nice properties.

AR and MA Redundancies

We set $\mathbf{B}_0$ to be a lower triangular matrix, with ones on the diagonals.

The free elements of $\mathbf{B}_l$ $(l = 0, \ldots, p)$ and $\boldsymbol{\Theta}_l$ $(l = 1, \ldots, q)$ are modeled with Stochastic Search Variable Selection priors (George *et al.*, 2008, JoE):

$$\left(B_{l,ij} \,|\, \gamma^B_{l,ij}\right) \sim \left(1 - \gamma^B_{l,ij}\right) \mathcal{N}(0, \tau^2_{0,l,ij}) + \gamma^B_{l,ij} \mathcal{N}(0, \tau^2_{1,l,ij}),$$
$$\left(\Theta_{l,ij} \,|\, \gamma^\Theta_{l,ij}\right) \sim \left(1 - \gamma^\Theta_{l,ij}\right) \mathcal{N}(0, \tau^2_{0,l,ij}) + \gamma^\Theta_{l,ij} \mathcal{N}(0, \tau^2_{1,l,ij}),$$

where $\gamma^B_{l,ij} \in \{0, 1\}$, $\gamma^\Theta_{l,ij} \in \{0, 1\}$ and $\tau^2_{0,l,ij} \ll \tau^2_{1,l,ij}$.

This approximates the *Echelon form* for unique VARMA specifications (see Chan *et al.*, 2016, JoE).

- It is needed because $\mathbf{B}(L)$ and $\boldsymbol{\Theta}(L)$ have redundant coefficients when, e.g., $[\mathbf{B}_p; \boldsymbol{\Theta}_q]$ has row rank less than n.

Log Likelihood

Let $\widetilde{\mathbf{\Theta}}_{1:t}$ be the $nt \times n(t+q)$ matrix given by

$$\widetilde{\mathbf{\Theta}}_{1:t} = \begin{pmatrix} \mathbf{\Theta}_q & \mathbf{\Theta}_{q-1} & \cdots & \mathbf{\Theta}_1 & \mathbf{I}_n & & & & \\ & \mathbf{\Theta}_q & \mathbf{\Theta}_{q-1} & \cdots & \mathbf{\Theta}_1 & \mathbf{I}_n & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & \mathbf{\Theta}_q & \mathbf{\Theta}_{q-1} & \cdots & \mathbf{\Theta}_1 & \mathbf{I}_n \end{pmatrix}.$$

Let $\widetilde{\mathbf{\Sigma}}_{1-q:t}$ be the $n(t+q) \times n(t+q)$ matrix given by

$$\widetilde{\mathbf{\Sigma}}_{1-q:t} = \begin{pmatrix} \mathbf{\Sigma}_{1-q} & & \\ & \ddots & \\ & & \mathbf{\Sigma}_t \end{pmatrix}.$$

Let $\mathbf{v}_t = \mathbf{B}_0 \mathbf{y}_t - \boldsymbol{\mu} + \mathbf{B}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{B}_p \mathbf{y}_{t-p}$ and $\widetilde{\mathbf{v}}^t$ the $nt \times 1$ vector

$$\widetilde{\mathbf{v}}_{1:t} = \left(\mathbf{v}_1', \ldots, \mathbf{v}_t'\right)'.$$

Log Likelihood

The log likelihood for the first t observations is given by:

$$\ln p(\mathbf{y}_1, \ldots, \mathbf{y}_t \mid \cdot\ ) = -\frac{1}{2}\ln(2\pi) \\ -\frac{1}{2}|\widetilde{\boldsymbol{\Theta}}_{1:t}\widetilde{\boldsymbol{\Sigma}}_{1:t}\widetilde{\boldsymbol{\Theta}}'_{1:t}| - \frac{1}{2}\widetilde{\mathbf{v}}'_{1:t}\left(\widetilde{\boldsymbol{\Theta}}_{1:t}\widetilde{\boldsymbol{\Sigma}}_{1:t}\widetilde{\boldsymbol{\Theta}}'_{1:t}\right)^{-1}\widetilde{\mathbf{v}}_{1:t}. \quad (6)$$

The full log likelihood is obtained by replacing t with T.

The term $\left(\widetilde{\boldsymbol{\Theta}}_{1:t}\widetilde{\boldsymbol{\Sigma}}_{1:t}\widetilde{\boldsymbol{\Theta}}'_{1:t}\right)^{-1}$ complicates computation because it is nonlinear in $\boldsymbol{\Theta}_l$ and $\boldsymbol{\Sigma}_t$, but the quadratic term is not expensive to compute.

If the $\boldsymbol{\Theta}(L)$ was fundamental, we could set $\boldsymbol{\Sigma}_{1-q} = \boldsymbol{\Sigma}_0 = 0$ and simplify computation (e.g. Chan, 2013, JoE).

- We cannot take this approach here because $\boldsymbol{\Theta}(L)$ is possibly non-fundamental; we instead treat $\boldsymbol{\Sigma}_{1-q}, \ldots, \boldsymbol{\Sigma}_0$ as parameters.

Parameters and Priors

The parameters in this specification are the free elements in $\mathbf{B}_0$, along with $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\boldsymbol{\mu}$, $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q$, $\boldsymbol{\Sigma}_{1-q}, \ldots, \boldsymbol{\Sigma}_0$, $\{\gamma^B_{l,ij}\}$, $\{\gamma^\Theta_{l,ij}\}$, ν and δ.

Along with the priors already discussed, we assign the following:

$$\Pr(\gamma^B_{l,ij} = 0) = 0.5,$$
$$\Pr(\gamma^\Theta_{l,ij} = 0) = 0.5,$$

$$\begin{aligned} \boldsymbol{\mu} &\sim \mathcal{N}(0, 10), \\ \boldsymbol{\Sigma}_{q-1}^{-1} &\sim \mathcal{W}\left(n+1, \mathbf{I}_n\right), \\ \nu - n &\sim \mathcal{G}(1, 1), \\ \delta &\sim \mathcal{U}(-1, 1), \end{aligned}$$

where $\mathcal{G}$ denotes the Gamma distribution and $\mathcal{U}$ the Uniform distribution.

MCMC Outline

The MCMC algorithm involves iterating seven sampling steps:

1. $\left(\boldsymbol{\mu}, \mathbf{B}_0, \ldots, \mathbf{B}_p \,|\, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q, \{\boldsymbol{\Sigma}_t\}, \{\gamma^B_{l,ij}\}\right)$,
2. $\left(\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q \,|\, \boldsymbol{\mu}, \mathbf{B}_0, \ldots, \mathbf{B}_p, \{\boldsymbol{\Sigma}_t\}, \{\gamma^\Theta_{l,ij}\}\right)$,
3. $(\boldsymbol{\Sigma}_{1-q}, \ldots, \boldsymbol{\Sigma}_T \,|\, \boldsymbol{\mu}, \mathbf{B}_0, \ldots, \mathbf{B}_p, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q, \mathbf{S}, \nu, \delta)$,
4. $\left(\gamma^B_{l,ij} \,|\, B_{l,ij}\right)$ for $l = 0, \ldots, p$, $i, j = 1, \ldots, n$ (and $i > j$ for $l = 0$),
5. $\left(\gamma^\Theta_{l,ij} \,|\, \Theta_{l,ij}\right)$ for $l = 1, \ldots, q$, $i, j = 1, \ldots, n$,
6. $(\nu \,|\, \{\boldsymbol{\Sigma}_t\}, \delta)$,
7. $(\delta \,|\, \{\boldsymbol{\Sigma}_t\}, \nu)$.

Sampling Implementation

Step 1 entails standard sampling from the multivariate Gaussian distribution.

Steps 4-5 are also standard in the SSVS literature.

Steps 6-7 are provided in Philipov and Glickman (2006, JBES).

Step 3 is efficiently implemented using the particle Gibbs with ancestry sampling as in Lindsten *et al*. (2014, JMLR).

Step 2 is the most challenging: we implement an independence MH sampler based on the Whittle likelihood approach of Dahlhaus (2002, AoS).

Whittle Likelihood

Let $\check{\boldsymbol{\Theta}}_k = \sum_{l=0}^{q} \boldsymbol{\Theta}_l \mathrm{e}^{il\lambda_k}$, where $i = \sqrt{-1}$ and $\lambda_k = 2\pi k/T$.

Let $\widetilde{\mathbf{V}}_k = (\mathbf{v}_1 \mathrm{e}^{-i\lambda_k}, \dots, \mathbf{v}_T \mathrm{e}^{-iT\lambda_k})$ and $\mathrm{vec}(\widetilde{\mathbf{W}}) = \boldsymbol{\Omega}\,\mathrm{vec}(\check{\boldsymbol{\Theta}}_k^{-1} \widetilde{\mathbf{V}}_k)$, where

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma}_1^{-1} & \boldsymbol{\Sigma}_1^{-1} & \cdots & \cdots & \boldsymbol{\Sigma}_{\lfloor \frac{1+T}{2} \rfloor}^{-1} \\ \boldsymbol{\Sigma}_1^{-1} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \boldsymbol{\Sigma}_{\lfloor \frac{r+c}{2} \rfloor}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{\lfloor \frac{1+T}{2} \rfloor}^{-1} & \cdots & \cdots & \cdots & \boldsymbol{\Sigma}_T^{-1} \end{pmatrix}.$$

$\boldsymbol{\Omega}$ is $nT \times nT$, with r denoting the row, c the column and $\lfloor \cdot \rfloor$ the floor operator.

Whittle Likelihood

The Whittle likelihood is given by:

$$\ln p_W(\mathbf{v};\boldsymbol{\theta}) = -\frac{1}{2T^2}\sum_{k=0}^{T-1}\Bigg((2n)\ln(2\pi)\sum_{t=1}^{T}|\check{\boldsymbol{\Theta}}_k\boldsymbol{\Sigma}_t\check{\boldsymbol{\Theta}}_k^*| + \frac{1}{2\pi}\operatorname{vec}(\widetilde{\mathbf{W}}_k)^*\operatorname{vec}(\widetilde{\mathbf{V}}_k)\Bigg) \quad (7)$$

The score is given by:

$$\frac{d\ln p_W(\mathbf{v};\boldsymbol{\theta})}{d\boldsymbol{\Theta}_l'} = \frac{1}{T}\sum_{k=0}^{T-1}\operatorname{Re}\left[\check{\boldsymbol{\Theta}}_k^{-1} + \frac{1}{2\pi T}\check{\boldsymbol{\Theta}}_k^{-1}\widetilde{\mathbf{V}}_k\widetilde{\mathbf{W}}_k^*\check{\boldsymbol{\Theta}}_k^{-1}\right]. \quad (8)$$

Both are relatively easy to evaluate, which means the Whittle likelihood is easy to maximize.

MH Proposal

Dahlhaus (2002, AoS) proved that if $\widehat{\boldsymbol{\theta}}$ maximizes $p_W(\mathbf{v};\boldsymbol{\theta})$ then it converges to the MLE of the actual likelihood $p(\mathbf{v};\boldsymbol{\theta})$, and it is asymptotically efficient.

Consequently, we set the proposal density to

$$\boldsymbol{\theta}^c \sim \mathcal{N}\left(\widehat{\widehat{\boldsymbol{\theta}}}, -\widehat{\widehat{\mathbf{H}}}_\theta^{-1}\right),$$

where $\widehat{\widehat{\boldsymbol{\theta}}}$ maximizes $\ln p(\boldsymbol{\theta}) + \ln p_W(\mathbf{v};\boldsymbol{\theta})$ and

$$\widehat{\widehat{\mathbf{H}}}_\theta = \frac{d^2 \ln p(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}'} + \frac{d^2 \ln p_W(\mathbf{v};\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}'}.$$

In Progress...

Initial Monte Carlo results that the algorithm is reasonably efficient with VARMA$(2, 2)$ and up to $n = 4$ variables.

Extensive Monte Carlo exercise is under way.

Real data application will re-examine SVARs for effects of monetary policy shocks and news shocks.