Efficient Estimation in Non-linear Non-Gaussian State Space Models

Joshua Chan Rodney Strachan

Research School of Economics
Australian National University

14 May 2013

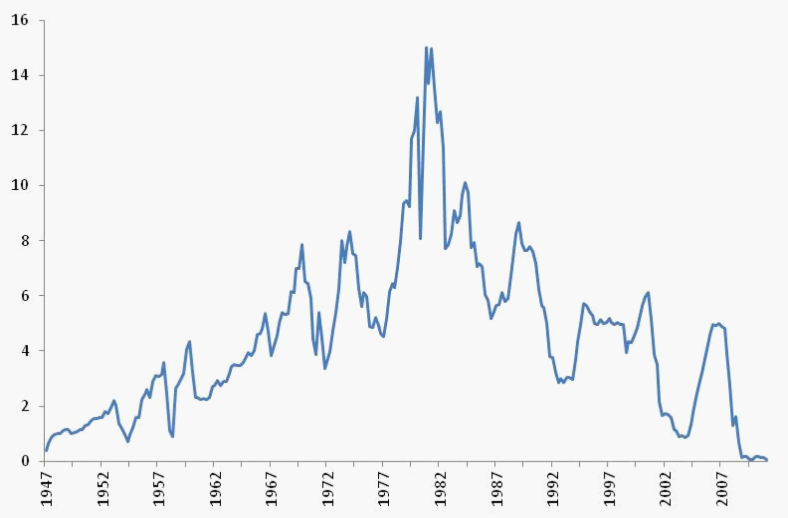
Motivation and Application

- Time Varying Parameter VARs have proven very insightful for macro-policy analysis

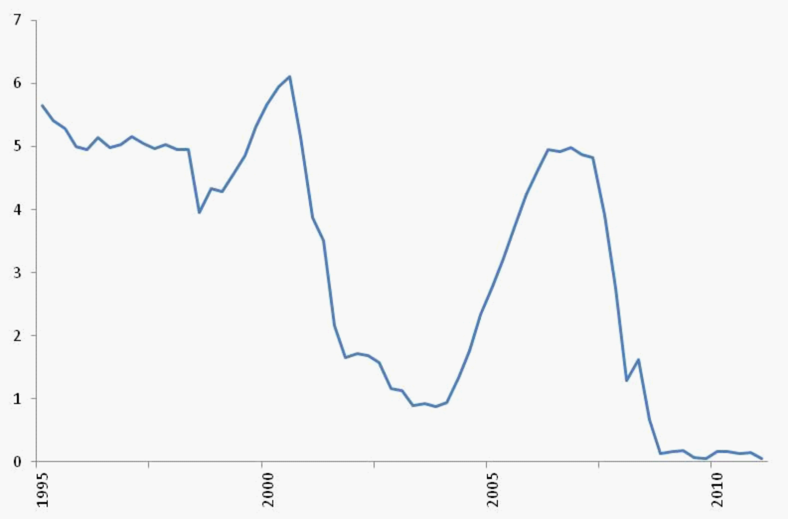
$$y_t = X_t \eta_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \Sigma^{-1})$$
$$\eta_t = \eta_{t-1} + \zeta_t, \quad \zeta_t \sim \mathrm{N}(0, \Omega^{-1})$$

- Since the (global) financial crisis (GFC), things have changed
- some important variables are now at, or near, their bounds
  - e.g., short-term interest rates; $y_{1,t} > 0$

US 3 month T-Bill Interest Rate



US 3 month T-Bill Interest Rate



Motivation (cont...)

- These bounds affect parameter estimation and imply non-linear models
- Another example we are working on
  - bounds on exchange rates (Swiss franc); $y_{2,t} \leq \overline{e}$

The Features of The Models We Can Consider

- State space representations
- Non-linearity becomes relevant only in the last few years
- Large dimensions: e.g., VARs
  - univariate non-linear methods not much use
- non-Gaussian, but 'Gaussian-like', errors

The Framework

- Measurement equation: $p(y_t \mid \eta_t, \theta)$, where
  - y_t is an $n \times 1$ vector of observations
  - η_t is an $m \times 1$ vector of latent states
  - θ denotes the set of model parameters
- State equation: $p(\eta_t \mid \eta_{t-1}, \theta)$
- Note 1: $p(y_t \mid \eta_t, \theta)$ may depend on previous observations y_{t-1}, y_{t-2}, etc. and other covariates
- Note 2: it can be generalized to: $p(y_t \mid \eta_t, \eta_{t-1}, \ldots, \eta_{t-l}, \theta)$ or $p(\eta_t \mid \eta_{t-1}, \ldots, \eta_{t-l}, \theta)$

Estimation methods

Substantive progress for the linear Gaussian case:

- Kalman filter-based algorithms: Carter and Kohn (1994), Fruwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002)
- Precision-based algorithms: Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011)

Non-linear Non-Gaussian case: a very active research area

Non-linearity in many states is tricky and we present an approach for one important application

Non-linear Non-Gaussian case: Three Broad Approaches

Auxiliary mixture sampling:

- Use data augmentation and finite Gaussian mixtures to approximate non-Gaussian errors
- Applicable to various stochastic volatility models and state space models for Poisson counts
- Efficient and easy to implement when applicable
- Typically model-specific

Three Broad Approaches (cont.)

Particle filter:

- A Broad class of techniques that involves sequential importance sampling and bootstrap resampling
- In the state space setting, it is used to evaluate the integrated likelihood via sequential importance sampling and resampling
- Popular for estimating (non-linear) DSGE models (Rubio-Ramirez and Fernandez-Villaverde, 2005; Fernandez-Villaverde and Rubio-Ramirez, 2007)
- Very general approach, but computationally demanding (computation time in days)

Three Broad Approaches (cont.)

Direct sampling via MH:

- Construct an approximation for the conditional density of the states, which is used to generate candidate draws for the MH step
- Common choice: Gaussian. e.g., Durbin and Koopman (1997), Shephard and Pitt (1997), Strickland, Forbes, and Martin (2006), etc.
- Difficulties:
  - Obtaining the approximation and generating draws from it at every iteration of the MCMC cycle is not trivial;
  - MH acceptance rate can be quite low: Gaussian approximation not sufficiently good
- Better approximation: HESSIAN method (McCausland, 2008).
- Highly efficient, but currently only applicable to univariate state models (i.e., $m = 1$)

Main Goals

1. Describe a fast routine to construct a Gaussian approximation based on the precision-based method (as a by-product, also get a t approximation)
2. Discuss two more efficient sampling schemes for simulation of the states: ARMH and collapsed sampler
3. Application: TVP-VAR with stochastic volatility and a non-negativity restriction

Linear Gaussian Case

- For now, consider

$$y_t = X_t \eta_t + \varepsilon_t,$$
$$\eta_t = \Gamma_t \eta_{t-1} + \zeta_t,$$

for $t = 1, \ldots, T$, with

$$\begin{pmatrix} \varepsilon_t \\ \zeta_t \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} \Sigma_t^{-1} & 0 \\ 0 & \Omega_t^{-1} \end{pmatrix}\right)$$

- Σ_t and Ω_t are respectively the precision of ε_t and ζ_t
- Let $y = (y_1', \ldots, y_T')'$, $\eta = (\eta_1', \ldots, \eta_T')'$, and $\theta = \{\eta_0, \{\Gamma_t\}, \{\Sigma_t\}, \{\Omega_t\}\}$

The Measurement Equation

Stacking the measurement equation over the T time periods:

$$y = X\eta + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \Sigma^{-1}),$$

where $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_T')'$,

$$X = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_T \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} \Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_T^{-1} \end{bmatrix}$$

- $\log p(y \,|\, \theta, \eta) \propto -\frac{1}{2}\log|\Sigma^{-1}| - \frac{1}{2}(y - X\eta)'\Sigma(y - X\eta)$
- Note: Σ is a banded matrix

The State Equation

Stacking the state equation over the T time periods:

$$\begin{pmatrix} I_m & & & & \\ -\Gamma_2 & I_m & & & \\ & -\Gamma_3 & I_m & & \\ & & \ddots & \ddots & \\ & & & -\Gamma_T & I_m \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_T \end{pmatrix} = \begin{pmatrix} \Gamma_1 \eta_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \vdots \\ \zeta_T \end{pmatrix},$$

i.e., $K\eta = \gamma + \zeta, \quad \zeta \sim \mathrm{N}(0, \Omega^{-1})$

- Let $\eta^0 = K^{-1}\gamma$. Since $|K| = 1$, we have

$$\log p(\eta \,|\, \theta) \propto -\frac{1}{2} \log |\Omega^{-1}| - \frac{1}{2} \left(\eta - \eta^0\right)' K'\Omega K \left(\eta - \eta^0\right))$$

- Note $K'\Omega K$ is a also a banded matrix

The Conditional Density for the States

- Therefore, the log conditional density $\ln p(\eta \mid y, \theta)$ is

$$\begin{aligned}&\propto \ln p(y \mid \theta, \eta) + \ln p(\eta \mid \theta)\\&\propto -\frac{1}{2}\left[\eta'(X'\Sigma X + K'\Omega K)\eta - 2\eta'(X'\Sigma y + K'\Omega K\eta^0)\right]\end{aligned}$$

- In other words, $(\eta \mid y, \theta) \sim \mathrm{N}(\hat{\eta}, H^{-1})$, where

$$\begin{aligned}H &= K'\Omega K + X'\Sigma X,\\\hat{\eta} &= H^{-1}(K'\Omega K\eta^0 + X'\Sigma y)\end{aligned}$$

- Since $X'\Sigma X$ is banded, it follows that H is also banded

What this process gives us ...

- At this point we have the mean, $\hat{\eta}$, and precision, H
- Note that the precision, H, is a banded and sparse matrix

Efficient State Simulation for the Linear Gaussian Case

1. Compute H and obtain its Cholesky decomposition C_H such that $H = C'_H C_H$
2. Sample $u \sim \mathrm{N}(0, I_{Tm})$, and solve $C_H x = u$ for x by back-substitution Then $x \sim \mathrm{N}(0, H^{-1})$
3. Solve
$$C'_H C_H \hat{\eta} = K'\Omega K \eta^0 + X'\Sigma y$$
for $\hat{\eta}$ by forward- and back-substitution.
4. Finally return $\eta = \hat{\eta} + x$, so that $\eta \sim \mathrm{N}(\hat{\eta}, H^{-1})$

Key features:

- Can compute $\hat{\eta}$ and C_H fast
- Marginal cost of sampling from $\mathrm{N}(\hat{\eta}, H^{-1})$ is low
- Built-in routines for sparse matrices in Matlab and Gauss
- Can also generate from $\mathrm{t}(\nu, \hat{\eta}, H^{-1})$

General State Space: Measurement Equation

Idea: approximate the log-likelihood $\ln p(y \mid \eta, \theta)$ via a second-order Taylor expansion around $\tilde{\eta} = (\tilde{\eta}_1', \ldots, \tilde{\eta}_T')'$:

$$\begin{aligned}
\ln p(y \mid \eta, \theta) &\approx \ln p(y \mid \tilde{\eta}, \theta) + (\eta - \tilde{\eta})' f - \frac{1}{2}(\eta - \tilde{\eta})' G (\eta - \tilde{\eta}) \\
&\propto -\frac{1}{2}\left[\eta' G \eta - 2\eta'(f + G\tilde{\eta})\right],
\end{aligned}$$

$$f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_T \end{bmatrix},$$

$$f_t \equiv \frac{\partial}{\partial \eta_t} \ln p(y_t \mid \eta_t, \theta)\Big|_{\eta_t = \tilde{\eta}_t}, \quad G_t \equiv -\frac{\partial^2}{\partial \eta_t \eta_t'} \ln p(y_t \mid \eta_t, \theta)\Big|_{\eta_t = \tilde{\eta}_t}$$

The Gaussian Approximation

- State equation: linear Gaussian as before (for simplicity):

$$\ln p(\eta \mid \theta) \propto \frac{1}{2} \ln |\Omega| - \frac{1}{2}(\eta - \eta^0)' K' \Omega K (\eta - \eta^0)$$

- Combining this and the approximation for the measurement equation:

$$\begin{aligned} \ln p(\eta \mid y, \theta) &\propto \ln p(y \mid \eta, \theta) + \ln p(\eta \mid \theta) \\ &\approx -\frac{1}{2} \left[\eta'(G + K'\Omega K)\eta - 2\eta'(f + G\tilde{\eta} + K'\Omega K \eta^0) \right] \end{aligned}$$

- That is, the approximating distribution is Gaussian with precision $H \equiv G + K'\Omega K$

What we need for the Gaussian approximation

- Expand the Taylor approximation at the mode $\hat{\eta} = \tilde{\eta}$
- This then gives us the precision matrix, H
- Note that, again, the precision, H, is a banded and sparse matrix
- This structure will give us the necessary computational speed

We Investigate Three Sampling Schemes

- Sampling Scheme 1 (S1): MH with a Gaussian Proposal
  - Expand the Taylor approximation at the mode $\hat{\eta}$
  - Generate candidates from $q(\eta \mid y, \theta) = \mathrm{N}(\hat{\eta}, H^{-1})$ or $q(\eta \mid y, \theta) = t\left(\nu, \hat{\eta}, H^{-1}\right)$ for the MH step
- Sampling Scheme 2 (S2): ARMH with a Gaussian or t Proposal
- Sampling Scheme 3 (S3): Collapsed Sampling with Cross Entropy
  - Used to draw $\theta \sim p(\theta \mid y)$ then draw $\eta \sim p(\eta \mid y, \theta)$

Sampling Scheme 1: MH with a Gaussian Proposal

- Expand the Taylor approximation at the mode $\hat{\eta}$
- The mode can be found by Newton-Raphson method: given the current location $\eta^{(s)}$, compute

$$\begin{aligned}\eta^{(s+1)} &= \eta^{(s)} + H(\eta^{(s)})^{-1} \frac{\partial}{\partial \eta} \log p(\eta \mid y, \theta) \Big|_{\eta=\eta^{(s)}} \\ &= H(\eta^{(s)})^{-1} \left(f(\eta^{(s)}) + G(\eta^{(s)})\eta^{(s)} + K'0 \right)\end{aligned}$$

- Continue until $||\eta^{(s+1)} - \eta^{(s)}|| < \epsilon$, set $\hat{\eta} = \eta^{(s+1)}$
- Generate candidates from $\mathrm{N}(\hat{\eta}, H^{-1})$ for the MH step

Accept-reject Sampling

- Target density: $p(\eta \mid y, \theta) \propto p(y \mid \eta, \theta) p(\eta \mid \theta)$; proposal density $q(\eta \mid y, \theta)$
- In the classic AR sampling, we need a constant c such that

  $$p(y \mid \eta, \theta) p(\eta \mid \theta) \leq c q(\eta \mid y, \theta),$$

  for all η in the support of $p(\eta \mid y, \theta)$
- Difficult to obtain c efficiently (especially when θ is revised at every iteration)

Accept-reject Metropolis-Hastings

- Combination of the classic accept-reject sampling with the MH algorithm
- The ARMH relaxes the domination condition. When it is not satisfied, use MH
- Let
$$\mathcal{D} = \{\eta : p(y \mid \eta, \theta) p(\eta \mid \theta) \leq c q(\eta \mid y, \theta)\},$$
and let $\mathcal{D}^c$ denote its complement

Sampling Scheme 2: ARMH with a Gaussian Proposal

1. AR step: Generate a draw $\eta^* \sim q(\eta \,|\, y, \theta)$. Accept η^* with probability

$$\alpha_{\mathrm{AR}}(\eta^* \,|\, y, \theta) = \min\left\{1, \frac{p(y \,|\, \eta^*, \theta)p(\eta^* \,|\, \theta)}{cq(\eta^* \,|\, y, \theta)}\right\}.$$

Continue the process until a draw η^* is accepted

2. MH-step: Given the current draw η and the proposal η^*
   - if $\eta \in \mathcal{D}$, set $\alpha_{\mathrm{MH}}(\eta, \eta^* \,|\, y, \theta) = 1$;
   - if $\eta \in \mathcal{D}^c$ and $\eta^* \in \mathcal{D}$, set

$$\alpha_{\mathrm{MH}}(\eta, \eta^* \,|\, y, \theta) = \frac{cq(\eta \,|\, y, \theta)}{p(y \,|\, \eta, \theta)p(\eta \,|\, \theta)};$$

   - if $\eta \in \mathcal{D}^c$ and $\eta^* \in \mathcal{D}^c$, set

$$\alpha_{\mathrm{MH}}(\eta, \eta^* \,|\, y, \theta) = \min\left\{1, \frac{p(y \,|\, \eta^*, \theta)p(\eta^* \,|\, \theta)q(\eta \,|\, y, \theta)}{p(y \,|\, \eta, \theta)p(\eta \,|\, \theta)q(\eta^* \,|\, y, \theta)}\right\}$$

Return η^* with prob. $\alpha_{\mathrm{MH}}(\eta, \eta^* \,|\, y, \theta)$; otherwise return η

Another Way to Look at ARMH

- The AR step: a way to sample from

$$q_{\mathrm{AR}}(\eta \,|\, y, \theta) = d^{-1} \alpha_{\mathrm{AR}}(\eta \,|\, y, \theta) q(\eta \,|\, y, \theta)$$

- By adjusting the original proposal density $q(\eta \,|\, y, \theta)$ by the function $\alpha_{\mathrm{AR}}(\eta \,|\, y, \theta)$, a better approximation is achieved
- In fact, we have

$$q_{\mathrm{AR}}(\eta \,|\, y, \theta) = \begin{cases} p(y \,|\, \eta, \theta) p(\eta \,|\, \theta)/cd, & \eta \in \mathcal{D}, \\ q(\eta \,|\, y, \theta)/d, & \eta \in \mathcal{D}^c, \end{cases}$$

- Better approximation, but requires multiple draws in the AR step

Joint Sampling of (θ, η)

- Typically sample from $p(\eta \mid y, \theta)$ and $p(\theta \mid y, \eta)$ sequentially
- In some settings, η and θ might be highly correlated
- Hence, sample (θ, η) jointly by first drawing from $p(\theta \mid y)$ marginally of the states η followed by a draw from $p(\eta \mid y, \theta)$
- Need a mechanism to generate candidates for θ. Often use random walk

Sampling Scheme 3: Collapsed Sampling with CE

- We propose an independence chain MH sampler instead
- The proposal density for θ, denoted as $q(\theta \mid y)$, is obtained optimally: given a parametric family of densities $\mathcal{P}$, use the member in $\mathcal{P}$ that is the closest to the marginal density $p(\theta \mid y)$ in the Kullback-Leibler divergence or the cross-entropy distance
- Generate $\theta^* \sim q(\theta \mid y)$, then evaluate the acceptance probability (that involves estimating the integrated likelihood via importance sampling)

Illustration: TVP-VAR with SV

- Write the VAR(l) in SUR form:

$$y_t = x_t\beta_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \Sigma_t^{-1}),$$

  where $x_t = I_n \otimes [1, y'_{t-1}, \ldots, y'_{t-l}]$ and $\beta_t = \mathrm{vec}([\mu_t : A_{t1} : \cdots : A_{tl}]')$ is a $k \times 1$ vector of VAR coefficients'with $k = n^2 l + n$

- Following Primiceri (2005), the time-varying precision matrix Σ_t is modeled as $\Sigma_t = L'_t D_t^{-1} L_t$, where $D_t = \mathrm{diag}(\mathrm{e}^{h_{t1}}, \ldots, \mathrm{e}^{h_{tn}})$ and

$$L_t = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{t21} & 1 & 0 & \cdots & 0 \\ a_{t31} & a_{t32} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{tn1} & a_{tn2} & \cdots & a_{tn,n-1} & 1 \end{pmatrix}$$

State Equations

- Let $h_t = (h_{t1}, \dots, h_{tn})'$ and a_t be the free elements of L_t, i.e., $a_t = (a_{t21}, a_{t31}, a_{t32}, \dots, a_{tn,n-1})'$
- Random walks for all the states:

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \Omega_\beta^{-1}), \\ h_t &= h_{t-1} + \xi_t, \quad \xi_t \sim \mathrm{N}(0, \Omega_h^{-1}), \\ a_t &= a_{t-1} + \zeta_t, \quad \zeta_t \sim \mathrm{N}(0, \Omega_a^{-1}), \end{aligned}$$

where Ω_β, Ω_h, and Ω_a are all diagonal matrices

Inequality Restriction

- For the application, we have $n = 3$ variables: nominal interest rate (3-month Tbill), inflation rate (CPI) and GDP growth
- U.S. quarterly data from 1947 Q1 to 2011 Q2
- Impose the restriction that the nominal interest rate is always non-negative (a model for computing liquidity trap)
- Assume $y_{t1} \geq 0$ is the nominal interest rate, and let x_{t1} be the first row of x_t
- The marginal distribution of y_{t1} is

$$(y_{t1} \mid \beta_t, \Sigma_t) \sim \mathrm{N}(x_{t1}\beta_t, \mathrm{e}^{h_{t1}})\mathbf{1}(y_{t1} \geq 0)$$

Inequality Restriction (cont.)

- Hence,

$$\mathbb{P}(y_{t1} \geq 0 \mid \beta_t, \Sigma_t) = 1 - \Phi\left(-x_{t1}\beta_t / e^{\frac{1}{2}h_{t1}}\right) = \Phi\left(x_{t1}\beta_t e^{-\frac{1}{2}h_{t1}}\right),$$

- The log-likelihood is $\ln p(y \mid \beta, \Sigma) = \sum_{t=1}^{T} \ln p(y_t \mid \beta_t, \Sigma_t)$, where

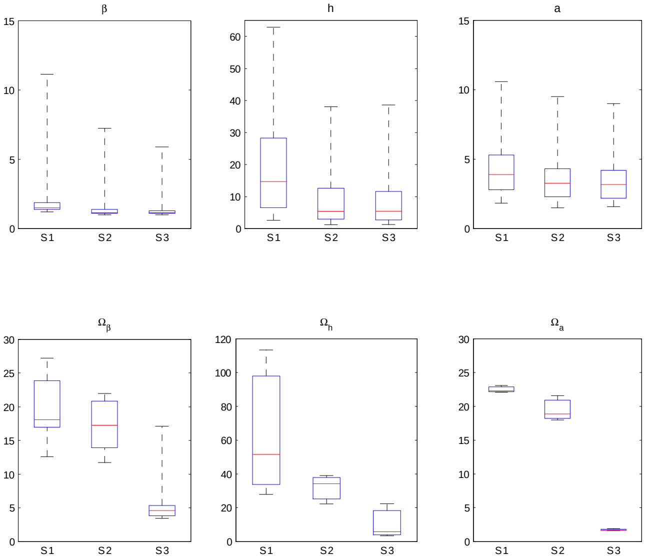
$$\ln p(y_t \mid \beta_t, a_t, h_t) \propto -\frac{1}{2}\sum_{i=1}^{n} h_{ti} - \frac{1}{2}(y_t - x_t\beta_t)' L_t' D_t^{-1} L_t (y_t - x_t\beta_t) - \ln \Phi\left(x_{t1}\beta_t e^{-\frac{1}{2}h_{t1}}\right)$$

Acceptance Rate and Running Time

Table: Acceptance rate (in %) and running time (in minutes; 50000 draws) of the three sampling schemes: MH (S1), ARMH (S2) and the collapsed sampler with CE (S3).

scheme	β	$h_{\cdot 1}$	$h_{\cdot 2}$	$\cdot 3$	Ω_β	Ω_h	Ω_a	time
S1	68	28	35	59	–	–	–	23
S2	95	71	79	97	–	–	–	27
S3	98	69	79	97	62	58	76	182

Inefficiency Factors



Estimation Results: volatilities and correlations

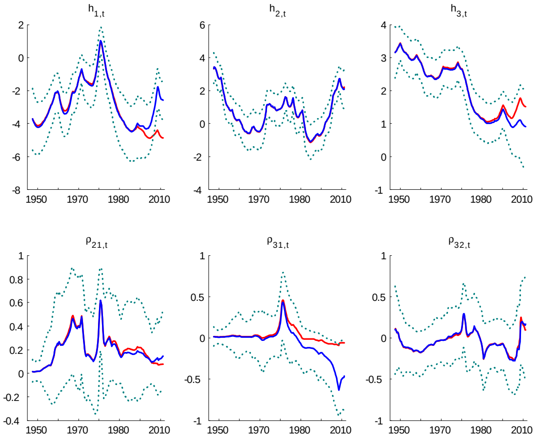


Figure: Evolution of the log-volatilities and correlations. Solid *red line* is the estimated posterior mean under the *unrestricted* model. The solid *blue line* is the estimated posterior mean under the *restricted* model with 5%-tile and 95%-tile, respectively.

Estimation Results: Impulse responses

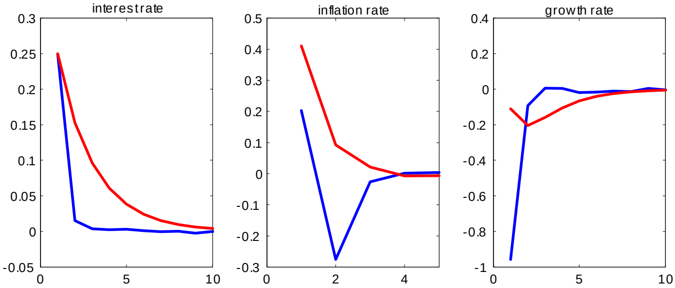


Figure: Impulse response to a credit shock under the unrestricted model (red solid line) and the model with the inequality restrictions imposed (blue solid line).

Concluding Remarks and Future Research

- Building on recent developments in precision-based algorithms, we propose a practical approach to simulating the states in a more general state space model
- A general approach that is much less computationally demanding than PF

Future research:

- non-linear DSGE models - limitations to invertible states
- a state space model for bounded inflation rate (already done)
- time-varying-parameter MA models (already done)
- non-linear factor models (wrestling with this and about to give up)