Structural Parameter Expansion for Data Augmentation.

Sébastien Blais

Département des sciences administratives, UQO

May 27, 2017

State-space model and normalization

A structural state-space model often involves parameter elements that are not identified.

Example:

$$
\begin{aligned}
y &\sim \mathcal{N}\left(b + h\xi, r\right) \\
\xi &\sim \mathcal{N}\left(e, q\right) \\
\theta &= (b, h, r, e, q) \in \Theta
\end{aligned}
$$

A normalization is a parameter subspace $\Theta^N \subset \Theta$ that is not observationally restrictive.

Example:

$$
\begin{aligned}
\Theta^N &= \{(b, r, h, e, q) | \, (b, h, r, e, q) \in \Theta; (e, q) = (0, 1)\} \subset \Theta \\
&= \left\{(\theta_1, \theta_2) \middle| \, (\theta_1, \theta_2) \in \Theta; \theta_2 = \bar{\theta}_2\right\} \subset \Theta_1^N \times \Theta_2^N = \Theta
\end{aligned}
$$

Why care?

Normalization

1. defines the shape of the parameter posterior and its interpretation;
2. influences the efficiency of posterior sampling;
   - Efficiency under a particular normalization depends on the data-generating process. (Pitt and Shephard, 1999; Papaspiliopoulos, Roberts and Sköld, 2003)
3. “interacts with” prior specification.
   - parameter estimation (Hamilton, Waggoner and Zha, 2007)
   - model choice (Frühwirth-Schnatter and Wagner, 2010)

Why care?

Normalization

1. defines the shape of the parameter posterior and its interpretation;
2. influences the efficiency of posterior sampling;
   - Efficiency under a particular normalization depends on the data-generating process. (Pitt and Shephard, 1999; Papaspiliopoulos, Roberts and Sköld, 2003)
3. "interacts with" prior specification.
   - parameter estimation (Hamilton, Waggoner and Zha, 2007)
   - model choice (Frühwirth-Schnatter and Wagner, 2010)

My contributions:

1. A posterior sampler that is simple and efficient, independently of normalization choice and the data-generating process.
2. Parameter priors that do not interact with normalization.

Why normalize?

If $p_\Theta(\theta)$ is proper, so is $p(\theta|\mathbf{y})$. Sampling from $p(\theta|\mathbf{y})$ could prove beneficial.

1. Drawing from a higher-dimensional parameter space improves convergence.
2. Drawing from an unnormalized parameter space is analytically convenient.

Outline

1. Background
2. Parameter Expansion (Liu and Wu, 1999)
3. Analytical convenience
4. Structural Parameter Expansion
5. Empirical results
6. Prior specification

Group invariance

The likelihood function of a state-space model is invariant under a group of parameter transformations induced by a group of transformations of the state vector.

Let Ξ denote the sample space of ξ. For a group Γ, let $\mathbf{f}_\xi : \Gamma \times \Xi \to \Xi$ specify Γ acting on the left of Ξ.

There exists a function $\mathbf{f}_\theta : \Gamma \times \Theta \to \Theta$ such that

$$p\left(\mathbf{y} \middle| \mathbf{f}_\xi\left(\gamma, \xi\right), \mathbf{f}_\theta\left(\gamma, \theta\right)\right) = p\left(\mathbf{y} \middle| \xi, \theta\right)$$

for all $\gamma \in \Gamma$.

For each $\gamma \in \Gamma$, define $\mathbf{M}_\gamma\left(\theta\right) = \mathbf{f}_\theta\left(\gamma, \theta\right)$ ($\mathbf{M}_\gamma : \Theta \to \Theta$).

Define also $\mathbf{M}_{\Theta^N}\left(\gamma, \theta_1\right) = \mathbf{f}_\theta\left(\gamma, \left(\theta_1, \bar{\theta}_2\right)\right)$ ($\mathbf{M}_{\Theta^N} : \Gamma \times \Theta_1^N \to \Theta$).

Group invariance

There exists a function $\mathbf{f}_\theta : \Gamma \times \Theta \to \Theta$ such that

$$p\left(\mathbf{y} \middle| \mathbf{f}_\xi\left(\gamma, \xi\right), \mathbf{f}_\theta\left(\gamma, \theta\right)\right) = p\left(\mathbf{y} \middle| \xi, \theta\right)$$

for all $\gamma \in \Gamma$.

For each $\gamma \in \Gamma$, define $\mathbf{M}_\gamma\left(\theta\right) = \mathbf{f}_\theta\left(\gamma, \theta\right)$ ($\mathbf{M}_\gamma : \Theta \to \Theta$).

Define also $\mathbf{M}_{\Theta^N}\left(\gamma, \theta_1\right) = \mathbf{f}_\theta\left(\gamma, \left(\theta_1, \bar{\theta}_2\right)\right)$ ($\mathbf{M}_{\Theta^N} : \Gamma \times \Theta_1^N \to \Theta$).

Remarks:

1. $\mathcal{M}_\Gamma = \{\mathbf{M}_\gamma | \gamma \in \Gamma\}$ is a group under function composition.
2. The unnormalized parameter space is induced by Γ: $\Theta = \mathbf{f}_\theta\left(\Gamma, \Theta^N\right)$ for any normalization $\Theta^N \subset \Theta$
3. $\mathbf{M}_{\Theta^N}$ is invertible: For each normalization Θ^N, there exists a function $\gamma : \Theta \to \Gamma$ such that $\mathbf{f}_\theta\left(\gamma\left(\theta\right), \theta\right) \in \Theta^N$.

Example

Consider the following normalized model:

$$y \sim \mathcal{N}\left(b + h\xi, r\right)$$
$$\xi \sim \mathcal{N}\left(0, 1\right)$$
$$\theta_1 = (b, h, r) \in \Theta_1^N$$

Define $\xi' = \mathbf{f}_\xi\left(\left(l, g\right), \xi\right) = g\xi + l$ with $(l, g) \in \mathcal{L} \times \mathcal{G}$.

$$\mathbf{y} \sim \mathcal{N}\left(b - hg^{-1}l + hg^{-1}\xi', r\right)$$
$$\xi' \sim \mathcal{N}\left(l, g^2\right)$$
$$\theta' = \mathbf{f}_\theta\left(l, g, b, h, r, 0, 1\right) = \left(b - hg^{-1}l, hg^{-1}, r, l, g^2\right)$$
$$= \left(b', h', r, e', q'\right) \in \Theta$$
$$\mathbf{M}_{\Theta^N}^{-1}\left(\theta'\right) = \left(b' + \sqrt{q'}h', \sqrt{q'}h', r, e', \sqrt{q'}\right) \in \Theta_1^N \times \mathcal{L} \times \mathcal{G}$$

Data augmentation

For a state-space model defined by

$$p\left(\mathbf{y}\middle|\,\theta_1,\bar{\theta}_2,\xi\right)$$
$$p\left(\xi\middle|\,\theta_1,\bar{\theta}_2\right)$$
$$\theta_1 \in \Theta_1^N$$

a standard data-augmentation algorithm would proceed as follows:

1. Draw ξ' from $p\left(\xi\middle|\,\theta_1,\bar{\theta}_2,\mathbf{y}\right)$
2. Draw θ_1' from $p_{\Theta_1^N}\left(\theta_1\middle|\,\bar{\theta}_2,\xi',\mathbf{y}\right) \propto p\left(\mathbf{y},\xi'\middle|\,\theta_1,\bar{\theta}_2\right) p_{\Theta_1^N}\left(\theta_1\right)$

Outline

1. Background
2. Parameter Expansion (Liu and Wu, 1999)
3. Analytical convenience
4. Structural Parameter Expansion
5. Empirical results
6. Prior specification

PX-DA

If $\mathcal{A}$ is a group and $\alpha \in \mathcal{A}$ indexes a differentiable mapping $\mathbf{M}_\alpha(\xi)$ such that

$$p\left(\mathbf{y}\middle|\,\mathbf{M}_\alpha(\xi), \theta_1, \bar{\theta}_2, \alpha\right) = p\left(\mathbf{y}\middle|\,\mathbf{M}_\alpha(\xi), \theta_1, \bar{\theta}_2\right),$$

PX-DA (Liu and Wu, 1999) proceeds a follows:

1. Draw ξ' from $p\left(\xi\middle|\,\theta_1, \bar{\theta}_2, \mathbf{y}\right)$;
2. Draw (θ_1', α^*) from

$$p^{PX-DA}_{\Theta_1^N \times \mathcal{A}}\left(\theta_1, \alpha\middle|\,\bar{\theta}_2, \xi', \mathbf{y}\right) \propto p\left(\mathbf{y}, \mathbf{M}_\alpha(\xi')\middle|\,\theta_1, \bar{\theta}_2\right) \left|\mathbf{J}_{\mathbf{M}_\alpha}\right| p_{\Theta_1^N}(\theta_1)\, p_{\mathcal{A}}\left(\alpha\middle|\,\theta_1\right),$$

where $\mathbf{J}_{\mathbf{M}_\alpha}$ denotes the Jacobian of $\mathbf{M}_\alpha(\xi)$ evaluated at ξ.

$p^{PX-DA}_{\Theta_1^N \times \mathcal{A}}\left(\theta_1, \alpha\middle|\,\bar{\theta}_2, \xi', \mathbf{y}\right)$ is not always a standard distribution, even if $p_{\Theta_1^N}\left(\theta_1\middle|\,\bar{\theta}_2, \xi', \mathbf{y}\right)$ is so.

Properties

1. DA and PX-DA define the same parameter posterior if $p_{\mathcal{A}}\left(\alpha|\,\theta_1\right)$ is
   - a proper density function
   - the improper limit of a sequence of proper priors
   - proportional to the left Haar measure on $\mathcal{A}$
2. Converge rate:
   - If $p_{\mathcal{A}}\left(\alpha|\,\theta_1\right) = p_{\mathcal{A}}\left(\alpha\right)$, PX-DA converges at least as fast as DA on Θ_1^N.
   - If $p_{\mathcal{A}}\left(\alpha\right)$ is proportional to the left Haar measure on $\mathcal{A}$, PX-DA converges at least as fast as DA on any normalization of $\Theta_1^N \times \mathcal{A}$.

Outline

1. Background
2. Parameter Expansion (Liu and Wu, 1999)
3. Analytical convenience
4. Structural Parameter Expansion
5. Empirical results
6. Prior specification

Analytical convenience

Example: Multinomial probit model

$$\mathbf{y}_i = \begin{cases} 1 & \text{if } \xi_i \geq \max(\xi) \\ 0 & \text{otherwise} \end{cases}$$

$$\xi \sim \mathcal{N}(\beta \mathbf{x}, \mathbf{Q})$$

$$\theta = (\beta, \mathbf{Q}) \in \Theta$$

θ is not identified because $p(\mathbf{y} | \beta, \mathbf{Q}) = p(\mathbf{y} | g\beta, g^2\mathbf{Q})$ for all $g \neq 0$.

Implementing DA on $\Theta^N = \{\theta \in \Theta | \mathbf{Q}_{11} = 1\}$ is challenging.

Operationalizing normalization as a mapping

McCulloch and Rossi (1994) proceed in two steps:

1. Generate a posterior sample on Θ
   - Sampling is simple because $p\left(\beta \mid \xi, \mathbf{Q}, \mathbf{y}\right)$ and $p\left(\mathbf{Q} \mid \xi, \beta, \mathbf{y}\right)$ are standard distributions.
2. Operationalize normalization by postprocessing the sample

$$\mathbf{M}_{\Theta^N}^{-1}\left(\beta, \mathbf{Q}\right) = \left(\beta / \sqrt{\mathbf{Q}_{11}}, \mathbf{Q}/\mathbf{Q}_{11}, g^* = \sqrt{\mathbf{Q}_{11}}\right) \in \Theta_1^N \times \mathcal{G}$$

Prior specification

1. Given p_Θ, what is $p_{\Theta_1^N}(\theta_1)$?

$$p_{\Gamma\times\Theta_1^N}(\gamma,\theta_1) = p_\Theta\left(\mathbf{M}_{\Theta^N}(\gamma,\theta_1)\right)\left|\mathbf{J}_{\mathbf{M}_{\Theta^N}}\right|$$

$$p_{\Theta_1^N}(\theta_1) = \int_\Gamma p_\Theta\left(\mathbf{M}_{\Theta^N}(\gamma,\theta_1)\right)\left|\mathbf{J}_{\mathbf{M}_{\Theta^N}}\right| \nu(\mathrm{d}\gamma),$$

where $\mathbf{J}_{\mathbf{M}_{\Theta^N}}$ denotes the Jacobian of $\mathbf{M}_{\Theta^N}$.

2. Given a target $p_{\Theta_1^N}(\theta_1)$ and a prior $p_\Gamma(\gamma|\theta_1)$,

$$p_\Theta(\theta) = p_{\Theta_1^N\times\Gamma}\left(\mathbf{M}_{\Theta^N}^{-1}(\theta)\right)\left|\mathbf{J}_{\mathbf{M}_{\Theta^N}^{-1}}\right|$$

Outline

1. Background
2. Parameter Expansion (Liu and Wu, 1999)
3. Analytical convenience
4. Structural Parameter Expansion
5. Empirical results
6. Prior specification

SPX-DA

Recipe: Exploit the invariance property of the likelihood function for generating a parameter space Θ that

1. simplifies implementation of posterior sampling
2. includes every possible normalization (so that convergence is at least as fast as DA under any normalization)

Sampling proceeds as follows:

1. Draw ξ' from $p\left(\xi \middle| \theta_1, \bar{\theta}_2, \mathbf{y}\right)$
2. Draw θ^* from $p\left(\theta \middle| \xi', \mathbf{y}\right)$
3. Compute $\left(\theta_1', \gamma^*\right) = \mathbf{M}_{\Theta^N}^{-1}\left(\theta^*\right)$

Comparisons

- McCulloch and Rossi (1994) map the posterior sample; I map after each sweep.
  - $p(\theta | \xi', \mathbf{y})$ is improper if $p(\theta)$ is improper and Θ is not compact: $p(\xi | \theta, \mathbf{y})$ is not well defined.
  - I want to use an improper prior...
- PX-DA and DA define the same parameter posterior if $p(\alpha)$ is proper or the left Haar measure on $\mathcal{A}$; SPX-DA and DA define the same posterior if

$$p_{\Theta}(\theta) = p_{\Theta_1^N \times \Gamma}\left(\mathbf{M}_{\Theta^N}^{-1}(\theta)\right) \left|\mathbf{J}_{\mathbf{M}_{\Theta^N}^{-1}}\right|$$

$$p_{\Theta_1^N}(\theta_1) = \int_{\Gamma} p_{\Theta}\left(\mathbf{M}_{\Theta^N}(\gamma, \theta_1)\right) \left|\mathbf{J}_{\mathbf{M}_{\Theta^N}}\right| \nu(\mathrm{d}\gamma).$$

Remark: The DA and SPX-DA algorithms I consider do not define the same parameter posterior.

Outline

1. Background
2. Parameter Expansion (Liu and Wu, 1999)
3. Analytical convenience
4. Structural Parameter Expansion
5. Empirical results
6. Prior specification

Linear state-space model

$$\mathbf{y}_t \sim \mathcal{N}\left(\tilde{\mathbf{B}} + \tilde{\mathbf{H}}\xi_t, \mathbf{R}\right)$$
$$\xi_t \sim \mathcal{N}\left(\tilde{\mathbf{F}}\xi_{t-1}, \mathcal{I}\right)$$
$$\left(\tilde{\mathbf{B}}, \tilde{\mathbf{H}}, \mathbf{R}, \tilde{\mathbf{F}}\right) \in \Theta_1^N$$

where $\tilde{\mathbf{H}}_{n,k} = 0$ for $k > n$ and $\tilde{\mathbf{H}}_{n,k} > 0$. The likelihood function is invariant under a group of parameter transformations induced by the left-action of the affine group on the state vector defined by $(\mathbf{L}, \mathbf{G})\,\xi = \mathbf{G}\xi + \mathbf{L}$.

$$\mathbf{y}_t \sim \mathcal{N}\left(\mathbf{B} + \mathbf{H}\xi_t, \mathbf{R}\right)$$
$$\xi_t \sim \mathcal{N}\left(\mathbf{E}, \mathbf{F}\xi_{t-1}, \mathbf{Q}\right)$$
$$\left(\mathbf{B}, \mathbf{H}, \mathbf{R}, \mathbf{E}, \mathbf{F}, \mathbf{Q}\right) \in \Theta$$

with $\left(\mathbf{B}, \mathbf{H}, \mathbf{R}, \mathbf{E}, \mathbf{F}, \mathbf{Q}\right) = \mathbf{M}_{\Theta^N}\left(\tilde{\mathbf{B}}, \tilde{\mathbf{H}}, \mathbf{R}, \tilde{\mathbf{F}}, \mathbf{L}, \mathbf{G}\right)$

$$= \left(\tilde{\mathbf{B}} - \tilde{\mathbf{H}}\mathbf{G}^{-1}\mathbf{L}, \tilde{\mathbf{H}}\mathbf{G}^{-1}, \mathbf{R}, \left(\mathcal{I} - \mathbf{G}^{-1}\tilde{\mathbf{F}}\mathbf{G}\right)\mathbf{L}, \mathbf{G}^{-1}\tilde{\mathbf{F}}\mathbf{G}, \mathbf{G}\mathbf{G}^{\top}\right).$$

SPX-DA for LSSM

1. Draw ξ^* from $p\left(\xi \middle| \tilde{\mathbf{B}}, \tilde{\mathbf{H}}, \mathbf{R}, \mathbf{0}, \tilde{\mathbf{F}}, \mathcal{I}, \mathbf{y}\right)$;
2. Draw $(\mathbf{B}^*, \mathbf{H}^*, \mathbf{R}^*, \mathbf{E}^*, \mathbf{F}^*, \mathbf{Q}^*)$ from
$$p_{\Theta}^{SPX-DA}(\mathbf{B}, \mathbf{H}, \mathbf{R}, \mathbf{E}, \mathbf{F}, \mathbf{Q} | \xi^*, \mathbf{y}) \propto p(\mathbf{y}, \xi^* | \mathbf{B}, \mathbf{H}, \mathbf{R}, \mathbf{E}, \mathbf{F}, \mathbf{Q}) \times p_{\Theta}^{SPX-DA}(\mathbf{B}, \mathbf{H}, \mathbf{R}, \mathbf{E}, \mathbf{F}, \mathbf{Q});$$
3. Compute
$$\left(\tilde{\mathbf{B}}', \tilde{\mathbf{H}}', \mathbf{R}', \tilde{\mathbf{F}}', \mathbf{L}^*, \mathbf{G}^*\right) = \mathbf{M}_{\Theta^N}^{-1}(\mathbf{B}^*, \mathbf{H}^*, \mathbf{R}^*, \mathbf{E}^*, \mathbf{F}^*, \mathbf{Q}^*),$$

Implementation requires only minor modification of a standard DA algorithm.

Simulation setup

Data-generating process:

$$\xi_{t+1} = \mathbf{0} + \alpha_F \begin{bmatrix} 1 & 0 \\ & 0.75 \end{bmatrix} \xi_t + \mathbf{v}_t$$

$$\mathbf{Y}_t = \mathbf{0} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \xi_t + \sqrt{\alpha_R}\mathbf{w}_t,$$

where $\mathbf{v}_t$ and $\mathbf{w}_t$ are vectors of independent standard normal variables, $(\alpha_F, \alpha_R) \in \{0.7, 0.9, 0.95\} \times \{0.01, 0.1, 0.5\}$.

$$\mathbf{R} = r\mathcal{I}$$

$$p_{\Theta_1^N}(\theta_1) \propto p(r)$$

$$p_\Theta(\theta) \propto p(r) \det(\mathbf{Q})^{-\frac{1-N+(K+1)}{2}}$$

$$r^{-1} \sim Gamma\left(2, \alpha_R^{-1}\right)$$

Inefficiency factors

For a posterior sample of M iterations,

$$\text{Var}\left[h\left(\theta\right)\right] = \frac{\text{Var}\left(h\left(\theta\right)\right)}{M} \times \tau,$$

where τ is the autocorrelation time of $h\left(\theta\right)$: an MCMC sampler requires a simulation size τ times larger than an i.i.d. sampler for estimating $h\left(\theta\right)$ with the same precision.

The inefficiency factor of a quantity is an estimator of its autocorrelation time. I computed it as

$$\hat{\tau} = 1 + 2\sum_{q=1}^{500}\left(1 - \frac{q}{500}\right)\hat{\rho}\left(q\right).$$

where $\hat{\rho}\left(q\right)$ is the sample autocorrelation of the quantity at lag q.

$\alpha_F = 0.9$ and $\alpha_R = 0.1$

	DA				SPX-DA			
	min	median	max	mean	min	median	max	mean
Element	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\tilde{\mathbf{B}}_1$	290.4	443.7	488.5	436.2	0.6	1.0	1.4	1.0
$\tilde{\mathbf{B}}_2$	106.3	253.1	391.8	256.7	0.6	1.0	1.3	1.0
$\tilde{\mathbf{B}}_3$	300.9	453.4	489.4	447.7	0.6	1.0	1.4	1.0
$\tilde{\mathbf{B}}_4$	303.1	453.5	489.4	447.8	0.6	1.0	1.4	1.0
$\tilde{\mathbf{H}}_{1,1}$	61.4	106.0	218.2	108.1	0.9	1.4	2.1	1.4
$\tilde{\mathbf{H}}_{2,1}$	122.0	195.3	303.4	201.7	0.8	1.2	2.0	1.3
$\tilde{\mathbf{H}}_{3,1}$	155.2	241.3	370.6	247.2	0.7	1.1	2.1	1.2
$\tilde{\mathbf{H}}_{4,1}$	154.9	242.0	370.4	247.2	0.8	1.1	2.1	1.2
$\tilde{\mathbf{H}}_{2,2}$	26.9	54.9	93.9	56.0	0.8	1.2	1.9	1.2
$\tilde{\mathbf{H}}_{3,2}$	31.6	58.7	101.4	60.1	0.9	1.2	1.9	1.2
$\tilde{\mathbf{H}}_{4,2}$	31.5	58.9	102.2	60.1	0.7	1.2	1.6	1.2
$\mathbf{R}_{1,1}$	2.4	3.6	5.4	3.7	1.9	2.9	4.2	2.9
$\tilde{\mathbf{F}}_{1,1}$	7.8	40.7	351.7	63.7	0.6	1.1	1.5	1.1
$\tilde{\mathbf{F}}_{2,1}$	10.2	62.4	206.2	78.9	0.8	1.2	1.7	1.2
$\tilde{\mathbf{F}}_{1,2}$	2.9	8.2	72.1	11.8	0.7	1.1	1.6	1.1
$\tilde{\mathbf{F}}_{2,2}$	4.1	11.1	51.6	13.2	0.7	1.1	1.7	1.2

Inefficiency factors for 101 artificial data samples of 200 observations. Based on $50,000$ iterations.

DA algorithm

α_R \ α_F	0.70	0.90	0.99
0.01	440.2	462.2	467.2
0.10	290.0	400.9	430.4
0.50	118.3	322.2	390.2

Panel a - $\tilde{\mathbf{B}}$

α_R \ α_F	0.70	0.90	0.99
0.01	305.3	355.2	393.9
0.10	79.8	136.7	231.3
0.50	27.5	59.1	150.0

Panel b - $\tilde{\mathbf{H}}$

α_R \ α_F	0.70	0.90	0.99
0.01	2.8	2.9	3.1
0.10	3.1	3.6	4.7
0.50	4.1	4.8	8.6

Panel c - $\mathbf{R}$

α_R \ α_F	0.70	0.90	0.99
0.01	16.7	30.3	46.5
0.10	9.8	30.6	70.0
0.50	7.3	28.8	76.8

Panel d - $\tilde{\mathbf{F}}$

SPX-DA algorithm

α_R \ α_F	0.70	0.90	0.99
0.01	1.0	1.0	1.1
0.10	1.0	1.0	1.1
0.50	1.0	1.0	1.1

Panel a - $\tilde{\mathbf{B}}$

α_R \ α_F	0.70	0.90	0.99
0.01	1.0	1.0	1.1
0.10	1.1	1.2	1.4
0.50	1.9	2.1	2.5

Panel b - $\tilde{\mathbf{H}}$

α_R \ α_F	0.70	0.90	0.99
0.01	2.9	2.9	3.3
0.10	2.9	2.9	3.2
0.50	3.1	3.0	3.3

Panel c - $\mathbf{R}$

α_R \ α_F	0.70	0.90	0.99
0.01	1.0	1.0	1.1
0.10	1.1	1.1	1.3
0.50	2.1	1.8	2.0

Panel d - $\tilde{\mathbf{F}}$

Panel of interest rates

	DA	SPX-DA
Element	(1)	(2)
$\tilde{\mathbf{B}}$	481.7	1.1
$\tilde{\mathbf{H}}$	344.8	1.9
$\mathbf{R}$	8.6	2.7
$\tilde{\mathbf{F}}$	99.3	1.4

$\mathbf{R} = r\mathcal{I}$

	DA	SPX-DA
Element	(1)	(2)
$\tilde{\mathbf{B}}$	483.8	1.3
$\tilde{\mathbf{H}}$	399.5	2.0
$\mathbf{R}$	8.4	5.8
$\tilde{\mathbf{F}}$	113.8	1.4

$\mathbf{R}$ is diagonal

$K = 3$, $N = 7$, 216 monthly observations.

Outline

1 Background
2 Parameter Expansion (Liu and Wu, 1999)
3 Analytical convenience
4 Structural Parameter Expansion
5 Empirical results
6 Prior specification

Prior specification

Normalization should not be observationally restrictive. Yet it sometimes interacts with prior specification and distorts statistical inference...

1. $p\left(\mathbf{y}\middle|\,\theta\right)$:
   For each $\theta \in \Theta$, there exists a $\theta' \in \Theta^N$ such that $p\left(\mathbf{y}\middle|\,\theta'\right) = p\left(\mathbf{y}\middle|\,\theta\right)$
2. $p\left(\mathbf{y}\right)$:
   For all normalizations $\tilde{\Theta}, \Theta^N \subset \Theta$

$$\int_{\Theta} p\left(\mathbf{y}\middle|\,\theta\right) p_{\Theta}\left(\theta\right) \mathbf{1}_{\tilde{\Theta}}\left(\theta\right) \nu\left(\mathrm{d}\theta\right) = \int_{\Theta} p\left(\mathbf{y}\middle|\,\theta\right) p_{\Theta}\left(\theta\right) \mathbf{1}_{\Theta^N}\left(\theta\right) \nu\left(\mathrm{d}\theta\right),$$

Requiring that normalization contains no information about the observables thus severely constrains prior specification.

Invariant priors

Definition: A density function $p(\theta)$ (which might not be proper) is invariant under $\mathcal{M}_\Gamma(\Theta)$ if

$$p(\mathbf{M}_\gamma(\theta)) \left|\mathbf{J}_{\mathbf{M}_\gamma}\right| = p(\theta)$$

for all $\mathbf{M}_\gamma \in \mathcal{M}_\Gamma(\Theta)$, where $\mathbf{J}_{\mathbf{M}_\gamma}$ is the Jacobian of $\mathbf{M}_\gamma$ evaluated at θ.

Remark: Invariant priors are noninformative about dimensions that have no substantive interpretation, but they can (usefully) be informative about other dimensions.

If $p_\Theta(\theta)$ is invariant under $\mathcal{M}_\Gamma(\Theta)$,

$$\int_\Theta p(\mathbf{y}|\,\theta)\, p_\Theta(\theta)\, \mathbf{1}_{\Theta^N}(\theta)\ \nu(\mathrm{d}\theta) = \int_\Theta p(\mathbf{y}|\,\theta)\, p_\Theta(\theta)\, \mathbf{1}_{\Theta^N_\gamma}(\theta)\ \nu(\mathrm{d}\theta)$$

for all $\gamma \in \Gamma$, where $\Theta^N_\gamma = \left\{\theta \middle| \mathbf{M}_\gamma(\theta) \in \Theta^N\right\}$.

Example: LSSM

$\left|\mathbf{J}_{\mathbf{M}_{\mathbf{L},\mathbf{G}}}\right| = |\det \mathbf{G}|^{K+2-N}$ and $p_\Theta(\mathbf{B},\mathbf{H},\mathbf{R},\mathbf{E},\mathbf{F},\mathbf{Q})$ is invariant under $\mathcal{M}_{\mathcal{L}\times\mathcal{G}}(\Theta)$ if it is proportional to

- $p(\mathbf{R}) \det(\mathbf{Q})^{-\frac{1-N+(K+1)}{2}}$;
- $p(\mathbf{R}) \det(\Sigma)^{-\frac{1-N+(K+1)}{2}}$, where $\Sigma = \text{Cov}[\xi]$;
- $p(\mathbf{R}) \left|\det\left(\mathbf{H}^\top\mathbf{H}\right)\right|^{1-N+(K+1)}$;
- $p(\lambda)\, p(\mathbf{R}) \det(\mathbf{Q})^{-\frac{1-N+(K+1)}{2}}$, where λ is the vector of $\mathbf{F}$'s eigenvalues;
- $p(\delta|\,\mathbf{H},\mathbf{R},\mathbf{F},\mathbf{Q})\, p(\mathbf{R}) \det(\mathbf{Q})^{-\frac{1-N+(K+1)}{2}}$, where $\delta_n = \frac{\left[\mathbf{H}\Sigma\mathbf{H}^\top\right]_{n,n}}{\left[\mathbf{H}\Sigma\mathbf{H}^\top\right]_{n,n}+R_{n,n}} = \frac{\text{Cov}[\mathbf{H}\xi]_{n,n}}{\text{Cov}[\mathbf{y}]_{n,n}}$;
- ...

Remark: $\mathbf{R}$, λ and δ are invariant quantities. For instance, $p(\mathbf{R}|\,\xi,\mathbf{Y}) = p(\mathbf{R}|\,\mathbf{G}\xi+\mathbf{L},\mathbf{Y})$.

Conclusion

- SPX-DA is simple to implement and numerically efficient.
- SPX-DA's simplicity and efficiency are independent of
  1. normalization choice
  2. data-generating process
- Invariant priors do not interact with normalization.

To do:

- Analyze the empirical performance of invariant priors in model selection and forecasting applications.
- Find useful invariant prior for structural VARs, discrete choice models, cointegration models, stochastic volatility models, etc.