Forecasting with Weakly Identified Linear State-Space Models.

Sébastien Blais

Bank of Canada

4 November 2009

Outline

Outline

Likelihood-based inference

Consider a model with $\mathbb{E}\left[y_{t+1}|y_{1:t},\theta\right] = f(y_{1:t},\theta)$ and a prior distribution $p(\theta)$.

The predictive density is

$$p(y_{T+1}|y_{1:T}) = \int p(y_{T+1}|y_{1:T},\theta)p(\theta|y_{1:T})\,\mathrm{d}\theta.$$

Likelihood-based inference

Consider a model with $\mathbb{E}\left[y_{t+1}|y_{1:t},\theta\right] = f(y_{1:t},\theta)$ and a prior distribution $p(\theta)$.

The predictive density is

$$p(y_{T+1}|y_{1:T}) = \int p(y_{T+1}|y_{1:T},\theta)p(\theta|y_{1:T})\,\mathrm{d}\theta.$$

Minimizing expected square error yields a point forecast

$$\begin{aligned} \mathbb{E}\left[y_{T+1}|y_{1:T}\right] &= \arg\min_{\delta} \int (y_{T+1}-\delta)^2 p(y_{T+1}|y_{1:T})\,\mathrm{d}y_{T+1} \\ &\neq f(y_{1:T},\hat{\theta}_{MLE}) \text{ in finite sample} \end{aligned}$$

Likelihood-based inference

Consider a model with $\mathbb{E}\left[y_{t+1}|y_{1:t},\theta\right] = f(y_{1:t},\theta)$ and a prior distribution $p(\theta)$.

The predictive density is

$$p(y_{T+1}|y_{1:T}) \quad = \quad \int p(y_{T+1}|y_{1:T},\theta)p(\theta|y_{1:T})\,\mathrm{d}\theta.$$

Minimizing expected square error yields a point forecast

$$\begin{aligned}\mathbb{E}\left[y_{T+1}|y_{1:T}\right] \quad &= \quad \arg\min_{\delta} \int (y_{T+1}-\delta)^2 p(y_{T+1}|y_{1:T})\,\mathrm{d}y_{T+1} \\ &\neq \quad f(y_{1:T},\hat{\theta}_{MLE}) \text{ in finite sample}\end{aligned}$$

When does it matter?

- Posterior averaging is beneficial when a linear state-space model is weakly identified.

When is a LSSM weakly identified?

When the “true” parameter values are “close” to the region where the model is locally unidentified.

When is a LSSM weakly identified?

When the “true” parameter values are “close” to the region where the model is locally unidentified.

- Weak identification (empirical underidentification, local almost nonidentification) is a finite-sample problem.

When is a LSSM weakly identified?

When the “true” parameter values are “close” to the region where the model is locally unidentified.

Examples:

- Multicollinearity - correlation is “close” to being equal to 1
- Weak instruments - correlation is “close” to being equal to 0
- ARMA processes - MA and AR roots are “close” to canceling out

When is a LSSM weakly identified?

When the “true” parameter values are “close” to the region where the model is locally unidentified.

Consequences: irregular finite-sample distributions (far from a symmetric normal, e.g. multimodal)

- Parameter point estimators are bad summaries of uncertainty (e.g. strong bias)
- Unreliable asymptotics
- Confidence regions can be disjoint (challenges for communication)

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and “irregular” distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions (“Label switching”)
     - Dick and Bowden (1973), Redner and Walker (1984), Stephen (1997,2000), Celeux, Hurn and Robert (2000), Frühwirth-Schnatter (2001), Geweke (2007)
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections (“Sign switching”)
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
     - Jennrich (1978)
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
     - Blais (this paper), Box and Jenkins (1976), Stoffer and Wall (1991), Kleibergen and Hoek (2000), Frühwirth-Schnatter and Wagner (2008)
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators

Related literature

1. Weak identification and "irregular" distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions ("Label switching")
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections ("Sign switching")
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007).

Related literature

1. Weak identification and “irregular” distributions.
   - The likelihood function of a finite mixture distribution is invariant with respect to permutations of its component distributions (“Label switching”)
   - The likelihood function of latent factor models is invariant with respect to factor permutations
   - The likelihood function of latent factor models is invariant with respect to factor reflections (“Sign switching”)
   - No valid bounded confidence interval for a parameter exists if this parameter is not identifiable on a subset of the parameter space.
     - Gleser and Hwang (1987), Dufour (1997)
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification

Related literature

1. Weak identification and "irregular" distributions.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and "irregular" distributions.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)

Related literature

1. Weak identification and "irregular" distributions.
2. How best to normalize?
   - Normalization in structural equations models affects the finite-sample distribution of OLS and 2SLS estimators (Hillier, 1990).
   - Normalization becomes critical when weak identification issues arise (Hamilton, Waggoner and Zha, 2007)
     - "A poor normalization can lead to multimodal distributions, disjoint confidence intervals, and very misleading characterizations of the true statistical uncertainty."
     - They propose an "identification principle".

Outline

The likelihood function of Gaussian Linear State-Space Models (**LSSMs**)

$$\begin{aligned} \mathbf{y}_{t\,(N\times 1)} &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1\,(K\times 1)} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

The likelihood function of Gaussian Linear State-Space Models (**LSSMs**)

$$
\begin{aligned}
\mathbf{y}_{t\,(N\times 1)} &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\
\xi_{t+1\,(K\times 1)} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q})
\end{aligned}
$$

is invariant with respect to linear transformations: for any invertible matrix **M**,

$$
l\left(\mathbf{B},\mathbf{H},\mathbf{R},\mathbf{F},\mathbf{Q}\,\middle|\, y\right) = l\left(\mathbf{B},\mathbf{M}^{-1'}\mathbf{H},\mathbf{R},\mathbf{M}\mathbf{F}\mathbf{M}^{-1},\mathbf{M}\mathbf{Q}\mathbf{M}'\,\middle|\, y\right) \quad \forall y \in \mathcal{Y}.
$$

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{B} + \mathbf{H}'\mathbf{M}^{-1}\mathbf{M}\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\
\mathbf{M}\xi_{t+1} &= \mathbf{M}\mathbf{F}\mathbf{M}^{-1}\mathbf{M}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{M}\mathbf{Q}\mathbf{M}')
\end{aligned}
$$

The likelihood function of Gaussian Linear State-Space Models (**LSSMs**)

$$\begin{aligned} \mathbf{y}_{t\,(N\times 1)} &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1\,(K\times 1)} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

is invariant with respect to linear transformations: for any invertible matrix **M**,

$$l\left(\mathbf{B},\mathbf{H},\mathbf{R},\mathbf{F},\mathbf{Q} \,\middle|\, y\right) = l\left(\mathbf{B},{\mathbf{M}^{-1}}'\mathbf{H},\mathbf{R},\mathbf{M}\mathbf{F}\mathbf{M}^{-1},\mathbf{M}\mathbf{Q}\mathbf{M}' \,\middle|\, y\right) \quad \forall y \in \mathcal{Y}.$$

Elementary linear transformations:

$$\begin{aligned} \mathbf{M} &= \mathbf{D}: && \text{diagonal scale matrix} \\ \mathbf{M} &= \mathbf{O}: && \text{rotation matrix} \\ \mathbf{M} &= \mathbf{P}: && \text{permutation matrix} \\ \mathbf{M} &= \mathbf{S}: && \text{diagonal reflection matrix} \end{aligned}$$

The likelihood function of Gaussian Linear State-Space Models (**LSSMs**)

$$\begin{aligned} \mathbf{y}_{t\,(N \times 1)} &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1\,(K \times 1)} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

Example

LSSMs are invariant with respect to a (finite) set of 2^K reflections. With $K = 2$, these transformations are

$$\mathbf{S} \in \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

Definition

A normalization is a parameter subspace

$$\Theta^N \subseteq \Theta$$

Example

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

The normalization

$$\Theta^{\mathbf{Q}_1} = \left\{\theta \in \Theta \middle| \mathbf{Q}_{kk} = 1, k = 1, \ldots, K\right\}$$

breaks scale invariance.

What can be done?

- Taking parameter uncertainty into account will improve forecasts → Simulation results.
  - Taking parameter uncertainty into account becomes more beneficial the weaker the identification of some parameters.
- Normalization ensuring global identification may help communication → An identification principle.
  - Normalizations satifying the identification principle are more likely to yield unimodal distributions.
  - Many normalizations satisfy this principle: it may be useful to try several of them.

What can be done?

- Taking parameter uncertainty into account will improve forecasts $\longrightarrow$ Simulation results.
  - Taking parameter uncertainty into account becomes more beneficial the weaker the identification of some parameters.
- Normalization ensuring global identification may help communication $\longrightarrow$ An identification principle.
  - Normalizations satifying the identification principle are more likely to yield unimodal distributions.
  - Many normalizations satisfy this principle: it may be useful to try several of them.

Outline

Gibbs sampler

- Every parameter except γ admits a conditionally conjugate prior.
- I use a random-walk Metropolis-Hastings step to draw γ with latent factors as a single block.

$$q\left(\gamma',\xi'\middle|\mathbf{y},\gamma,\Phi,\Sigma_\gamma\right) = p\left(\xi'\middle|\mathbf{y},\gamma',\Phi\right)\phi\left(\gamma'\middle|\gamma,\Sigma_\gamma\right),$$

where $p\left(\xi'\middle|\mathbf{y},\gamma',\Phi\right)$ is available inclosed form and

$$\Phi \equiv \{\mathbf{B},\mathbf{Q},\mathbf{R},\mathbf{F},\xi_1\}.$$

Note: the joint proposal does not depend on ξ.

Posterior averaging is beneficial for weakly identified LSSMs.

	H	All	ARMA(1,1)	AR(1)
Weaker reflection identification	0.005	**0,795** (0,020) M=1000	0,817 (0,022) M=898	0,655 (0,048) M=102
↓	0.010	**0,852** (0,024) M=1000	0,880 (0,027) M=884	0,737 (0,050) M=116
	0.050	**0,883** (0,024) M=1000	0,919 (0,028) M=871	0,748 (0,051) M=129
Stronger reflection identification	0.100	**0,961** (0,023) M=1000	0,968 (0,026) M=876	0,908 (0,059) M=124

The data-generating process (an ARMA(1,1)) is

$$\begin{aligned} \xi_t &= F\xi_{t-1} + v_t, \\ y_t &= B + H'\xi_t + w_t, \end{aligned}$$

with $B = 0$, $R = 1$, $Q = 1$, $F = 0.95$.

Outline

Definition

Let Θ^I denote the nonidentification subset. A normalization $\Theta^N \subseteq \Theta$ satisfies the *identification principe* *if it*

a) $\text{int}\left(\Theta^N\right) \cap \Theta^I = \emptyset$ *and* $\Theta^I \subseteq \text{fr}\left(\Theta^N\right)$*;*

b) *is connected;*

c) *provides global identification.*

Note: Intersections of hyperplanes and half-spaces are connected.

- Hamilton, Waggoner and Zha (2007)
  - "Our proposal is that the boundaries of [a normalization set] A should correspond to the loci along which the structure is locally unidentified or the log likelihood is $-\infty$."
  - "One easy way to check whether a proposed normalization set A conforms to this identification principle is to make sure that the model is locally identified at all interior points of A."

Example (Harvey, 1989, $K = 2$)

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\
\xi_{t+1} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q})
\end{aligned}
$$

Normalization $\{\theta \in \Theta | H_{12} = 0\}$ *does not provide global identification because it does not break permutation invariance if* $H_{22}^0 = 0$ *under the data-generating process.*

$$
\mathbf{H} = \begin{bmatrix} H_{11} & 0 \\ H_{21} & H_{22} \end{bmatrix}
$$

The sampling distribution of $\hat{H}_{11}$ *will be bimodal for sufficiently large samples if* H_22 *is close enough to being equal to 0.*

Example (Harvey, 1989, $K = 2$)

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

Normalization $\{\theta \in \Theta | H_{12} = 0\}$ does not provide global identification because it does not break permutation invariance if $H_{22}^0 = 0$ under the data-generating process.

$$\mathbf{H} = \begin{bmatrix} H_{11} & 0 \\ H_{21} & 0 \end{bmatrix}$$

The sampling distribution of $\hat{H}_{11}$ will be bimodal for sufficiently large samples if $H_2 2$ is close enough to being equal to 0.

Example (Harvey, 1989, $K = 2$)

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B} + \mathbf{H}'\xi_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(0,\mathbf{R}) \\ \xi_{t+1} &= \mathbf{F}\xi_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0,\mathbf{Q}) \end{aligned}$$

Normalization $\{\theta \in \Theta | H_{12} = 0\}$ *does not provide global identification because it does not break permutation invariance if* $H^0_{22} = 0$ *under the data-generating process.*

$$\mathbf{H} = \begin{bmatrix} H_{21} & 0 \\ H_{11} & 0 \end{bmatrix}$$

The sampling distribution of $\hat{H}_{11}$ *will be bimodal for sufficiently large samples if* $H_2 2$ *is close enough to being equal to 0.*

A normalization of LSSMs that ensures global identification.

I propose

$$\Theta^{\mathbf{H_O}} = \left\{ \theta \in \Theta \middle| \mathbf{HH}' = \mathcal{I} \right\},$$

which breaks scale and rotation invariance, but preserves permutation and reflection invariance.

I parameterize a $K \times N$ row-orthogonal matrix as:
$\mathbf{H}' = \mathbf{B}_1 \mathbf{B}_2 \ldots \mathbf{B}_K \mathbf{U}$, où

$$\mathbf{B}_k = \rho_{k,k+1} \rho_{k,k+2} \cdots \rho_{k,N} \quad , \quad \gamma_{k,n} = \arctan \left(\frac{\mathbf{H}_{k,n+1}}{\sqrt{\sum_{i=1}^{n} \mathbf{H}_{k,i}^2}} \right)$$

$$\rho_{i,j} = \begin{bmatrix} \mathcal{I} & & & & \\ & \cos \gamma_{i,j} & & -\sin \gamma_{i,j} & \\ & & \mathcal{I} & & \\ & \sin \gamma_{i,j} & & \cos \gamma_{i,j} & \\ & & & & \mathcal{I} \end{bmatrix}_{(N \times N)} \quad , \quad \mathbf{U}_{(N \times K)} = \begin{bmatrix} \mathcal{I} \\ \mathbf{0} \end{bmatrix}$$

Outline

Conclusion

This paper

- argues that LSSMs are subject to weak identification issues;
- shows that posterior averaging is beneficial when forecasting with weakly identified LSSMs;
- offers a normalization which can alleviate communication problems caused by weak identification;
- describes a novel Gibbs sampler for Gaussian LSSMs.