Precision-based sampling with missing observations

Philipp Hauber[1,2]

[1]JMU Würzburg
[2]Kiel Institute for the World Economy (IfW)

Kiel Institute Research Seminar

March 22nd, 2022

This presentation

Mash-up of two papers in my dissertation!

Method:

Hauber, P and C. Schumacher (2021). *Precision-based sampling with missing observations: A factor model application*, **Bundesbank Discussion Paper 11/2021**.

Application:

Hauber, P. (2021) *How useful is external information from professional forecasters? Conditional forecasts in large factor models*

Motivation

Essential task in the Bayesian estimation of state space models: drawing from $p(\boldsymbol{\eta}|\mathbf{y}, \Theta)$ where $\boldsymbol{\eta}$ is an unobserved component, $\mathbf{y}$ is data and Θ parameters

Precision-based samplers (Chan and Jeliazkov 2009, **IJMMNO**; McCausland 2012, **JEcmtrics**) exploit the fact the precision matrix of $\boldsymbol{\eta}$ is banded in many macroeconomic application $\rightarrow$ alternative to simulation smoothers that rely on the Kalman filter

Applications in macroeconomics (with complete data) include models of trend inflation (Chan et al. 2013, **JBES**), time-varying Bayesian vector autoregressions (Chan 2020, **JBES**) and factor models (Kaufmann and Schumacher 2017, **JAE**)

Missing observations arise frequently in macroeconomic applications/datasets: different starting dates, different release patterns ("ragged edge"), outliers or mixed frequencies

In our paper, we propose a precision-sampler that can handle (most of these) applications!

Precision-based sampling

Simple example

AR(2) process: $\eta_t = \phi_1 \eta_{t-1} + \phi_2 \eta_{t-2} + u_t$; $u_t \sim \mathcal{N}(0, \sigma^2)$

Stacking the observations over $t = 1, \ldots, T$ yields

$$\mathbf{H}\boldsymbol{\eta} = \mathbf{u}, \text{ where } \mathbf{u} \sim \mathcal{N}(0, \mathbf{I}_T \sigma^2) \text{ and } \mathbf{H} = \begin{bmatrix} 1 & & & & & \\ -\phi_1 & 1 & & & & \\ -\phi_2 & -\phi_1 & 1 & & & \\ & -\phi_2 & -\phi_1 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\phi_2 & -\phi_1 & 1 \end{bmatrix}$$

$\boldsymbol{\eta}$ is Normal with mean $\mathbf{0}_T$ and covariance matrix $\Sigma = \mathbf{H}^{-1}\, \mathbf{I}_T \sigma^2\, {\mathbf{H}^{-1}}^{\mathsf{T}}$

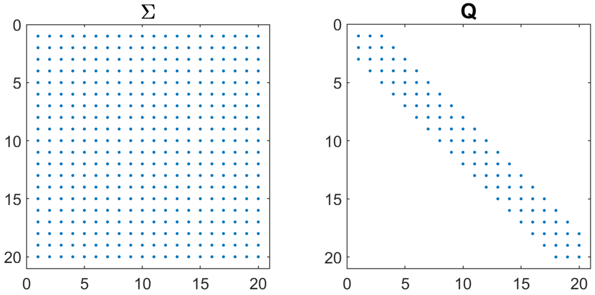
corresponding *precision matrix* is given by $\mathbf{Q} = \Sigma^{-1} = \mathbf{H}^{\mathsf{T}}\, \mathbf{I}_T \sigma^{-2}\, \mathbf{H}$

Precision-based sampling with missing observations

Covariance and precision matrix of η

Properties of the multivariate $\mathcal{N}$:

- $\Sigma_{ij} = 0 \implies$ independence of η_i and η_j
- $\mathbf{Q}_{ij} = 0 \implies$ **conditional** independence of η_i and η_j



Notes: The blue dots indicate the non-zero entries in the covariance matrix Σ and precision matrix $\mathbf{Q}$ of an AR(2) process for $T = 20$ observations. The former is a dense matrix while the latter is sparse and banded with lower and upper bandwidth equal to 2.

Precision-based sampling

Computational advantages of banded precision matrices

Solving linear systems of the form $Ux = b$ where U is an $n \times n$ upper-triangular matrix takes n^2 flops (left); when U has bandwidth p the solution can be obtained in $2np$ flops (right):

```
% solution to Ux = b
% U has maximal bandwidth
for i = n:-1:1
    x(i) = b(i)/U(i,i)
    for j = 1:i-1
        b(j) = b(j) - U(j,i)x(i)
    end
end
```

```
% solution to Ux = b
% U has bandwidth p
for i = n:-1:1
    x(i) = b(i)/U(i,i)
    for j = max{1,i-p}:i-1
        b(j) = b(j) - U(j,i)x(i)
    end
end
```

Even larger gains for matrix factorisations, e.g. Cholesky ($Q = LL^{\mathsf{T}}$) $\implies$ linear instead of cubic costs!

L "inherits" the bandwidth of Q (Golub and Van Loan 2013, Theorem 4.3.1)

Precision-based sampling

Factor model

To fix ideas, consider the following factor model:

$$
\begin{aligned}
\mathbf{y}_t &= \lambda \boldsymbol{\eta}_t + \mathbf{e}_t \\
\mathbf{e}_t &= \phi^e \mathbf{e}_{t-1} + \boldsymbol{\epsilon}_t;\ \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, diag([\sigma_1^2, \cdots, \sigma_N^2])) \\
\boldsymbol{\eta}_t &= \phi^\eta \boldsymbol{\eta}_{t-1} + \mathbf{u}_t;\ \mathbf{u}_t \sim \mathcal{N}(0, \Sigma_u)
\end{aligned}
$$

where $\mathbf{y}_t$ is an $N \times 1$ vector of data and η_t is an $R \times 1$ vector of unobserved factors

Bayesian estimation of the model is done via a Gibbs Sampler which sequentially draws from

- the conditional distribution of factors given data and parameters: $p(\boldsymbol{\eta}|\mathbf{y}, \Theta)$
- the conditional distribution of parameters given data and factors: $p(\Theta|\boldsymbol{\eta}, \mathbf{y})$

Precision-based sampling

Drawing from $p(\boldsymbol{\eta}|\mathbf{y}, \Theta)$

Joint distribution of states $\boldsymbol{\eta} = [\boldsymbol{\eta}_1{}^\mathsf{T}, \cdots, \boldsymbol{\eta}_T{}^\mathsf{T}]^\mathsf{T}$ and data $\mathbf{y} = [\mathbf{y}_1^\mathsf{T}, \cdots, \mathbf{y}_T^\mathsf{T}]^\mathsf{T}$ given parameters:

$$\mathbf{z} = \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}); \; \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_\eta & \mathbf{Q}_{\eta,y} \\ \mathbf{Q}_{\eta,y}^\mathsf{T} & \mathbf{Q}_y \end{bmatrix}$$

Mapping from Θ to $\mathbf{Q}$

Standard result for the multivariate $\mathcal{N}$: $p(\boldsymbol{\eta}|\mathbf{y}, \Theta) = \mathcal{N}(-\mathbf{Q}_\eta^{-1}\mathbf{Q}_{\eta y}\mathbf{y}, \mathbf{Q}_\eta^{-1})$

Sampling from this distribution does **not** require the inversion of (the potentially very large matrix) $\mathbf{Q}_\eta$ and because it is banded

- the mean $-\mathbf{Q}_\eta^{-1}\mathbf{Q}_{\eta y}\mathbf{y}$
- and a random draw given mean and precision matrix

can be obtained efficiently! Rue and Held (2005, Algorithms 2.1, 2.4)

Precision-based sampling with missing observations

Drawing from $p(\eta, \mathbf{y}^m | \mathbf{y}^o, \Theta)$: 3 block ordering

What to do when some observations in $\mathbf{y}$ are missing?

E.g. a model with one factor and data $\mathbf{y} = [y_{11}, y_{12}, \text{NaN}, y_{22}, y_{31}, \text{NaN}, \text{NaN}, \text{NaN}, y_{51}, \text{NaN}]^{\mathsf{T}}$

Reorder the model in 3 blocks as:

$$\mathbf{z}_{3b} = \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{y^m} \\ \mathbf{y^o} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{z}^{3b}_{\eta y^m} \\ \mathbf{z}^{3b}_{y^o} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{3b}^{-1});\ \mathbf{Q}_{3b} = \begin{bmatrix} \mathbf{Q}_{\eta} & \mathbf{Q}_{\eta, y^m} & \mathbf{Q}_{\eta y^o} \\ \mathbf{Q}^{\mathsf{T}}_{\eta, y^m} & \mathbf{Q}_{y^m} & \mathbf{Q}_{y^m, y^o} \\ \mathbf{Q}^{\mathsf{T}}_{\eta, y^o} & \mathbf{Q}^{\mathsf{T}}_{y^m, y^o} & \mathbf{Q}_{y^o} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{Q}^{3b}_{\eta y^m} & \mathbf{Q}^{3b}_{\eta y^m, y^o} \\ \mathbf{Q}^{3b,\mathsf{T}}_{\eta y^m, y^o} & \mathbf{Q}^{3b}_{y^o} \end{bmatrix}$$

Example:

$$\mathbf{z}_{3b} = [\boldsymbol{\eta}, \mathbf{y^m}, \mathbf{y^o}]^{\mathsf{T}} = [\eta_1, \eta_2, \eta_3, \eta_5, \eta_5, y_{21}, y_{32}, y_{41}, y_{42}, y_{52}, y_{11}, y_{12}, y_{22}, y_{31}, y_{51}]^{\mathsf{T}}$$

Precision-based sampling with missing observations

Drawing from $p(\eta, \mathbf{y}^m | \mathbf{y}^o, \Theta)$: time-t ordering

However, $Var(\mathbf{z}^{3b}_{\eta y^m | y^o})^{-1} = \mathbf{Q}^{3b}_{\eta y^m}$ will in general **not** be a banded matrix, so sampling as in the case of complete data not feasible!

$\Longrightarrow$ different ordering of $\mathbf{z}$ that groups conditionally dependent components - $\boldsymbol{\eta}_t$ and $\mathbf{y}^m_t$ (*time-t ordering* - together: $\mathbf{z}_\tau = \mathcal{P}_\tau \mathbf{z}$ and corresponding precision matrix $Q_\tau = \mathcal{P}_\tau \mathbf{Q} \mathcal{P}_\tau{}^\intercal$

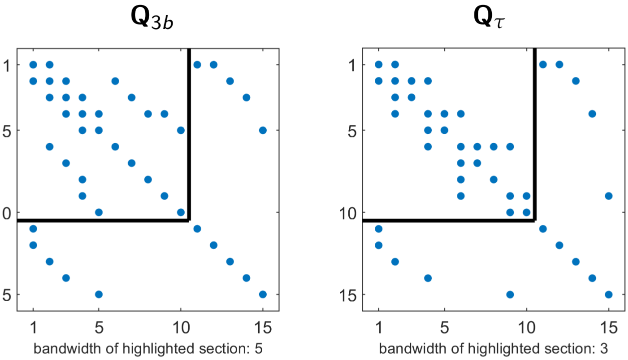
Example: $\mathbf{z}_{\mathcal{P}_\tau} = [\eta_1, \eta_2, y_{21}, \eta_3, y_{32}, \eta_4, y_{41}, y_{42}, \eta_5, y_{52}, y_{11}, y_{12}, y_{22}, y_{31}, y_{51}]^\intercal$

This ensures that the precision matrix of the desired conditional distribution $Var(\mathbf{z}^{\tau}_{\eta y^m | y^o})^{-1} = Q^{\tau}_{\eta y^m}$ is banded!

After sampling from $p(\mathbf{z}^{\tau}_{\eta y^m | y^o})$, reverse the permutation to back out the draw of $\boldsymbol{\eta}$ and $\mathbf{y}^m$

Precision-based sampling with missing observations

Example: Graphical comparison of $\mathbf{Q}_{3b}$ and $\mathbf{Q}_{\tau}$



Notes: The blue dots indicate the non-zero entries in the precision matrix of the 3-block ordering (left) and the time-t permutation (right) of the discussed example with 2 variables and one factor. The highlighted upper left submatrices correspond to the conditional precision matrix of factors and missing values given observations,$\mathbf{Q}^{3b}_{\eta y^m}$ and $Q^{\tau}_{\eta y^m}$.

References I

Chan, Joshua C. C., "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business & Economic Statistics*, 2020, *38* (1), 68–79.

—, **Gary Koop, and Simon M. Potter**, "A New Model of Trend Inflation," *Journal of Business & Economic Statistics*, 2013, *31* (1), 94–106.

Chan, Joshua CC and Ivan Jeliazkov, "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 2009, *1* (1-2), 101–120.

Golub, Gene H. and Charles F. Van Loan, *Matrix Computations* Johns Hopkins Studies in the Mathematical Sciences, 4^{th} ed., Johns Hopkins University Press, 2013.

Kaufmann, Sylvia and Christian Schumacher, "Identifying relevant and irrelevant variables in sparse factor models," *Journal of Applied Econometrics*, 2017, *32* (6), 1123–1144.

McCausland, William J., "The HESSIAN method: Highly efficient simulation smoothing, in a nutshell," *Journal of Econometrics*, 2012, *168* (2), 189–206.

Rue, Havard and Leonhard Held, *Gaussian Markov random fields: theory and applications*, CRC press, 2005.

Appendix

Mapping from Θ to Q

ladida ladida

Appendix

Rue and Held (2005, Algorithm 2.1, 2.4)

Algorithm 2.1 Solving $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A} > 0$

1: Compute the Cholesky factorization $\mathbf{A} = \mathbf{L}\mathbf{L}^\mathsf{T}$
2: Solve $\mathbf{Lv} = \mathbf{b}$ via forward substitution
3: Solve $\mathbf{L}^\mathsf{T}\mathbf{x} = \mathbf{v}$ via backward substitution
4: **return** $\mathbf{x}$

Algorithm 2.4 Sampling $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$

1: Compute the Cholesky factorization $\mathbf{Q} = \mathbf{L}\mathbf{L}^\mathsf{T}$
2: Sample $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I})$
3: Solve $\mathbf{L}^\mathsf{T}\mathbf{v} = \mathbf{z}$
4: Compute $\mathbf{x} = \boldsymbol{\mu} + \mathbf{v}$
5: **return** $\mathbf{x}$

Back