Slow EM Convergence in Low-Noise Dynamic Factor Models

Daan Opschoor
Econometric Institute, Erasmus University Rotterdam
Tinbergen Institute
opschoor@ese.eur.nl

Dick van Dijk
Econometric Institute, Erasmus University Rotterdam
Tinbergen Institute
djvandijk@ese.eur.nl

Key Takeaways

- Large-scale dynamic factor models (DFM) infeasible to estimate with direct numerical likelihood maximization.
  $\Longrightarrow$ Expectation-Maximization (EM) algorithm provides alternative.
- However, the **EM algorithm fails in a low-noise environment**.
  $\Longrightarrow$ Extremely slow convergence leading to poor estimates.
- We **solve** these issues with the Adaptive EM algorithm and/or with carefully injecting artificial noise.

Low-Noise DFM

- Popular practice in macroeconomic forecasting/nowcasting with DFMs is to allow for serial correlation in idiosyncratic component $\boldsymbol{\varepsilon}_t$.
  $\Longrightarrow$ Possible efficiency/forecasting gains.
- Use framework of Bańbura and Modugno (2014) to achieve this by including $\boldsymbol{\varepsilon}_t$ **in state vector** and introduce **artificial error term $\boldsymbol{e}_t$ with small variance κ** in order to apply EM in its usual form.
- Low-noise DFM with measurement equation

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}, \kappa \boldsymbol{I}), \tag{1}$$

with κ a small pre-fixed value (e.g., 10^{-4}) and (V)AR dynamics for states.

Failure of EM in Low-Noise DFM

- The M-step of the factor loading matrix $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda}_{j+1} = \boldsymbol{\Lambda}_j + \left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\Theta}_j}\left(\boldsymbol{e}_t \boldsymbol{f}_t' | \boldsymbol{Y} \right) \right) \left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\Theta}_j}\left(\boldsymbol{f}_t \boldsymbol{f}_t' | \boldsymbol{Y} \right) \right)^{-1}.$$

- In fact, Petersen et al. (2005) show that

$$\boldsymbol{\Lambda}_{j+1} = \boldsymbol{\Lambda}_j + \kappa \tilde{\boldsymbol{\Lambda}}_j + \mathcal{O}(\kappa^4), \tag{2}$$

  highlighting that the learning rate of M-step for $\boldsymbol{\Lambda}$ is proportional to the artificial noise level κ.
- This implies that if the variance of $\boldsymbol{e}_t$ becomes smaller (i.e., $\kappa \to 0$) that the **EM parameter update stagnates** (i.e., $\boldsymbol{\Lambda}_{j+1} \to \boldsymbol{\Lambda}_j$).

Solutions to EM failure in Low-Noise DFM

Adaptive EM

- The Adapative Overrelaxed EM (AEM) algorithm of Salakhutdinov and Roweis (2003) **boosts the parameter updates** by an **adaptive factor** η_j.
- The M-step of the factor loading matrix $\boldsymbol{\Lambda}$ in the AEM is

$$\boldsymbol{\Lambda}_{j+1}^{AEM} = \boldsymbol{\Lambda}_j^{AEM} + \eta_j \left(\boldsymbol{\Lambda}_{j+1} - \boldsymbol{\Lambda}_j^{AEM} \right).$$

- Combining this with equation (2) gives

$$\boldsymbol{\Lambda}_{j+1}^{AEM} = \boldsymbol{\Lambda}_j^{AEM} + \eta_j \kappa \tilde{\boldsymbol{\Lambda}}_j^{AEM} + \mathcal{O}(\kappa^4),$$

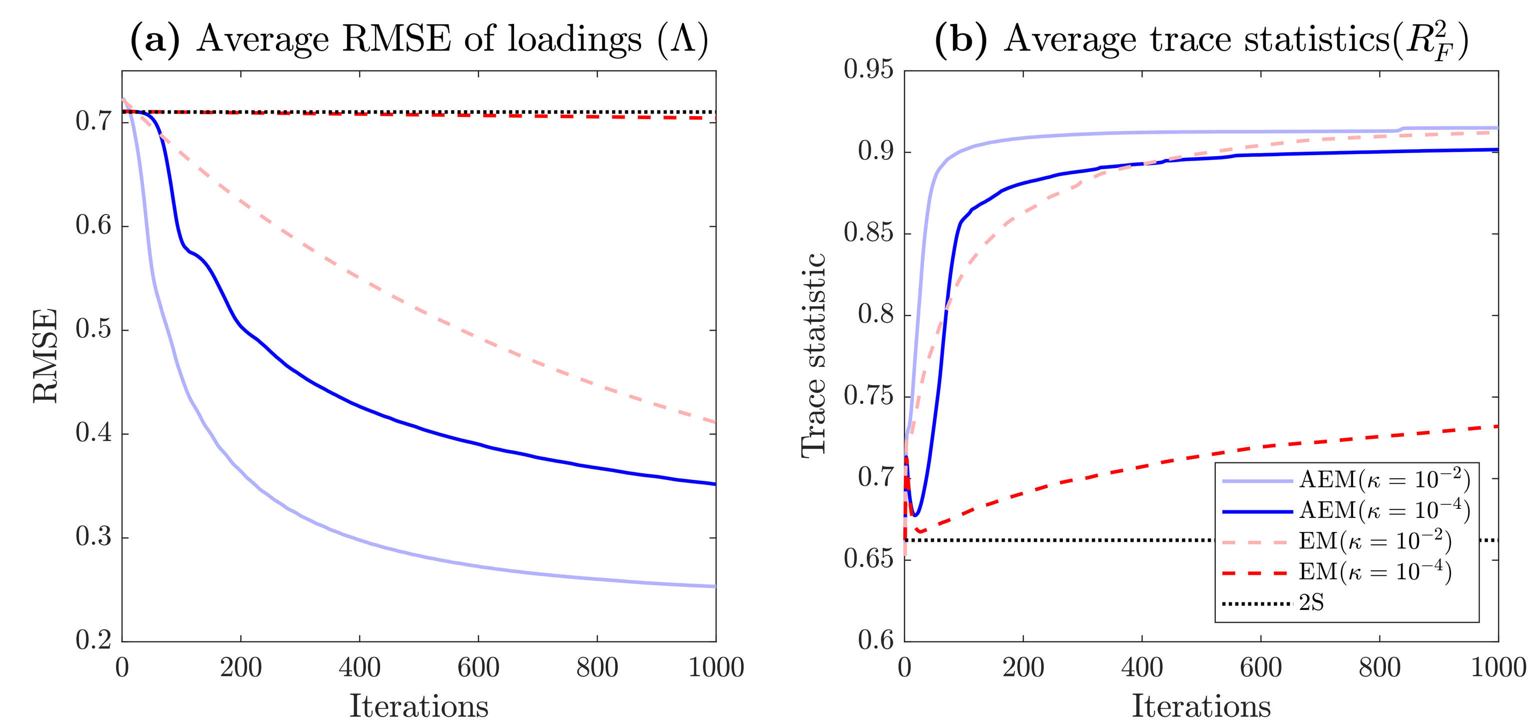
  showing that η_j counters low noise level κ and speeds up convergence.
- Following Salakhutdinov and Roweis (2003), use $\eta_{j+1} = \alpha \eta_j$ with $\alpha = 1.1$ and $\eta_1 = 1$.

Careful selection of noise level κ

- Increasing κ gives more artificial noise, but also increases the learning rate of the M-step, which could potentially speed up EM algorithm convergence (see, e.g., Osoba et al., 2013).
- Carefully select amount of noise based on Monte Carlo simulations.

Monte Carlo Simulations

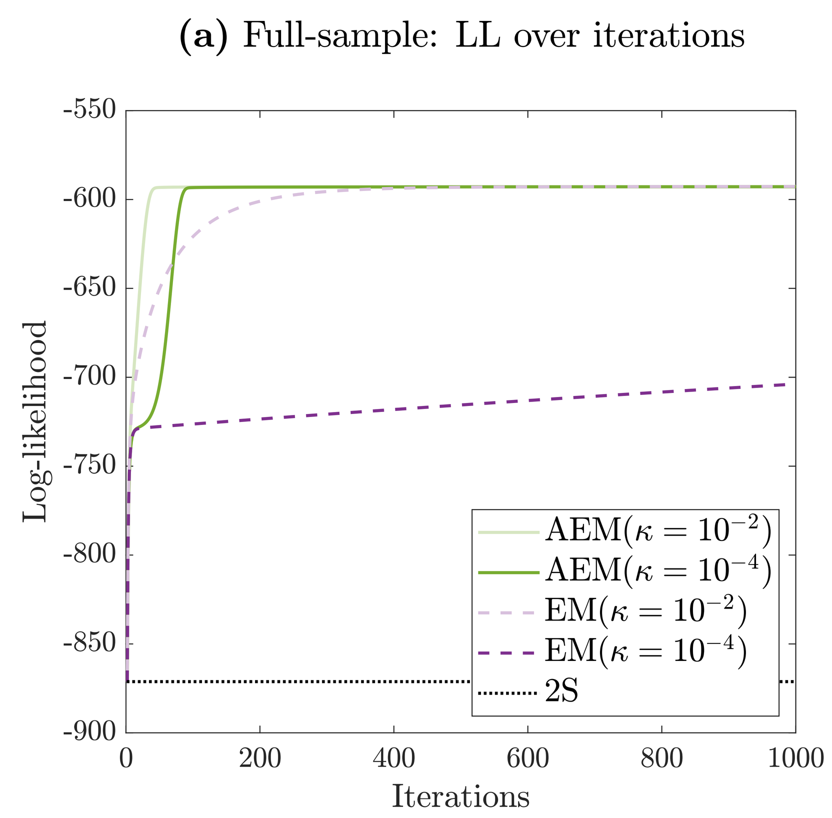
- Generate data from exact factor model à la Bańbura and Modugno (2014) and estimate low-noise DFM given in equation (1).
- Use two-step (2S) approach, EM algorithm and Adaptive EM algorithm for estimation with $\kappa = 10^{-4}$ and $\kappa = 10^{-2}$.
- Assess precision of parameter estimates with average RMSE and precision of factor estimates with average trace R^2 over 500 MC replications.
- Results for $T = 50$ and $N = 10$ (but similar for larger T and N):



- Extremely slow convergence of EM algorithm for estimation of $\boldsymbol{\Lambda}$.
  $\Longrightarrow$ Almost **no movement** from two-step (2S) initialization!
- Adaptive EM and slightly higher value of κ lead to much faster rate of convergence and thus more accurate estimates.
- Slow convergence of loadings also influences accuracy factor estimates.
- Results persist for other model (mis-)specifications.

Empirical Application

- Construct sequence of euro area GDP nowcasts/forecasts for 2006Q1 to 2022Q4 using macroeconomic dataset based on mixed-frequency DFM with serially correlated errors.
- Results for full-sample estimation and pseudo real-time nowcasting exercise based on small-scale model (i.e., $N = 10$):



(b) Relative RMSFE of GDP nowcasts compared to mean based on mixed-frequency DFM with $\kappa = 10^{-4}$

	2S	EM	AEM
Q(-1)M1	1.02	1.17	1.39
Q(-1)M2	1.02	1.11	1.01
Q(-1)M3	0.91	0.89	0.84
Q(0)M1	0.76	0.64	0.46
Q(0)M2	0.57	0.57	0.41
Q(0)M3	0.58	0.61	0.49
Q(+1)M1	0.62	0.66	0.53
Average	0.78	0.81	0.73

- AEM leads to **larger increments** and **faster convergence** of log-likelihood than EM, especially for small noise $\kappa = 10^{-4}$.
- AEM produces substantial **nowcast gains** compared to 2S and EM.

References

Bańbura, M. and M. Modugno (2014): "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160.

Osoba, O., S. Mitaim, and B. Kosko (2013): "The Noisy Expectation-Maximization Algorithm," *Fluctuation and Noise Letters*, 12, 1350012.

Petersen, K. B., O. Winther, and L. K. Hansen (2005): "On the Slow Convergence of EM and VBEM in Low-Noise Linear Models," *Neural Computation*, 17, 1921–1926.

Salakhutdinov, R. and S. Roweis (2003): "Adaptive overrelaxed bound optimization methods," in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, Washington, DC, USA: AAAI Press, ICML'03, 664–671.