TI 2023-018/III
Tinbergen Institute Discussion Paper

Slow Expectation-Maximization Convergence in Low-Noise Dynamic Factor Models

Daan Opschoor[1]
Dick van Dijk[1]

1 Erasmus University Rotterdam

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Slow Expectation-Maximization Convergence in Low-Noise Dynamic Factor Models*

Daan Opschoor† Dick van Dijk

Econometric Institute, Erasmus University Rotterdam

April 3, 2023

Abstract

This paper addresses the poor performance of the Expectation-Maximization (EM) algorithm in the estimation of low-noise dynamic factor models, commonly used in macroeconomic forecasting and nowcasting. We show analytically and in Monte Carlo simulations how the EM algorithm stagnates in a low-noise environment, leading to inaccurate estimates of factor loadings and latent factors. An adaptive version of EM considerably speeds up convergence, producing substantial improvements in estimation accuracy. Modestly increasing the noise level also accelerates convergence. A nowcasting exercise of euro area GDP growth shows gains up to 34% by using adaptive EM relative to the usual EM.

Keywords: Dynamic factor models, EM algorithm, artificial noise, convergence speed, nowcasting

JEL Classification: C32, C51, C53, E37

*We thank Matteo Barigozzi, Rutger-Jan Lange, Daniel Lewis, Sam van Meer, Minchul Shin and Terri van der Zwan for their helpful comments and suggestions, as well as participants at the 33th (EC)^2 Conference (2022, Paris) and seminar participants at Erasmus University Rotterdam (2022).

†Corresponding author. E-mail addresses: opschoor@ese.eur.nl (Daan Opschoor), djvandijk@ese.eur.nl (Dick van Dijk)

1 Introduction

Dynamic factor models have become a powerful tool for a wide range of macroeconomic and financial applications in data-rich environments, ranging from forecasting (Stock and Watson, 1999, 2002a,b; Ludvigson and Ng, 2009; Neely et al., 2014) and nowcasting (Giannone et al., 2008; Bańbura et al., 2011) to the construction of economic and financial activity indices (Stock and Watson, 1989, 1991; Aruoba et al., 2009; Brave and Butters, 2011) and uncertainty indices (Jurado et al., 2015; Scotti, 2016). In particular, dynamic factor models facilitate a straightforward approach to summarize the (co)variation of a large number of observed time series variables into a few latent common factors (see, for example, Poncela et al., 2021, for a recent survey on factor extraction in dynamic factor models). However, the number of parameters in the model becomes excessively large when the number of time series included increases.[1] This implies that conventional estimation based on direct numerical optimization of the likelihood in combination with the Kalman filter (see, among others, Engle and Watson, 1981; Stock and Watson, 1989) becomes unfeasible for large-scale dynamic factor models.

To overcome this issue, the Expectation-Maximization (EM) algorithm of Dempster et al. (1977) has become a popular alternative estimation approach in high-dimensional settings (see, among others, Quah and Sargent, 1993; Doz et al., 2012; Barigozzi and Luciani, 2022).[2] The EM algorithm has initially been adapted for dynamic factor models in state-space form by Shumway and Stoffer (1982) and Watson and Engle (1983). More recently, Bańbura and Modugno (2014) show that the EM algorithm is also easily modified to include serially correlated idiosyncratic components. Their approach even remains applicable under arbitrary patterns of missing data, which is particularly relevant for forecasting and nowcasting applications in which the included time series typically have different publication delays (the so-called 'ragged edge'), different sampling frequencies and different initial availability.

More specifically, Bańbura and Modugno (2014) treat the serially correlated idiosyn-

[1] The number of parameters increases linearly with the number of time series for an exact factor model (with cross-sectionally uncorrelated idiosyncratic components) and quadratically for an approximate factor model (with cross-sectionally correlated idiosyncratic components).

[2] Naturally, several alternative solutions have been proposed to deal with high-dimensional data in dynamic factor models (see, among others, Doz et al., 2011; Jungbacker and Koopman, 2015; Bräuning and Koopman, 2014). For a recent survey on high-dimensional dynamic factor models, see Lippi et al. (2022).

cratic components as additional latent states and introduce an artificial error term with small variance in the measurement equation. The latter is necessary in order to apply the EM algorithm in its usual form. Yet, it has been shown that the EM algorithm becomes inefficient in such a low-noise environment (Bermond and Cardoso, 1999; Petersen et al., 2005), causing extremely slow convergence, especially for the factor loading estimates. Unfortunately, this issue seems to have been overlooked by Bańbura and Modugno (2014) and subsequent applications of their approach (see, among others, Coroneo et al., 2016; Scotti, 2016; Alvarez et al., 2016; Bok et al., 2018; Barigozzi and Luciani, 2019; Cascaldi-Garcia et al., 2021; Caruso and Coroneo, 2023). Moreover, these low-noise issues could also arise more naturally whenever the series exhibit a strong factor structure with a high signal-to-noise ratio. An example of this situation concerns the term structure of interest rates, for which three factors explain almost all variation (Litterman and Scheinkman, 1991).

In this paper, we address this slow EM convergence issue in low-noise dynamic factor models in three different ways. First, we show both analytically and in Monte Carlo simulations how the EM algorithm fails in the estimation of the factor loadings. We demonstrate that the key issue concerns the learning rate in the M-step for the factor loadings, which is proportional to the variance of the error term in the measurement equation. Hence, low noise leads to slow convergence of the EM algorithm in the estimation of these loadings. Subsequently, our simulation study shows that the smoothed factors and other parameter estimates are also negatively affected by this slow convergence. We find that this failure of EM under low noise persists for different sample sizes and different model (mis)specifications.

Second, we demonstrate that the Adaptive Overrelaxed EM (AEM) algorithm of Salakhutdinov and Roweis (2003) is able to deal with these low-noise issues, as suggested by Petersen et al. (2005). The key feature of the AEM algorithm is that it boosts the parameter updates and thereby counters the low variance of the error term. Moreover, the AEM algorithm is a simple and straightforward extension of the conventional EM algorithm, making it just as easy to implement. Our Monte Carlo simulations show that the speed of convergence of the AEM algorithm is much faster than of the EM algorithm, with an average improvement in root mean squared error (RMSE) per iteration that is up to 883 times higher. Consequently, the AEM algorithm produces substantially more accurate factor loading estimates than the EM algorithm with up to 57% improvement in average RMSE. Furthermore, the smoothed factors based on the AEM algorithm are a better approximation of the true factors compared to standard EM, with an accuracy gain in average trace statistic of up to 23%. The other parameters are also more accurately estimated with AEM compared to EM, albeit to a lesser extent. In the Bańbura and Modugno (2014) approach, the variance of the artificial error term is treated as a hyperparameter, which ex ante is fixed typically at a very small value (such as 10^{-4}). We demonstrate that carefully choosing a somewhat larger but still modest level of artificial noise (for example, setting the variance equal to 10^{-2}) considerably speeds up convergence, improving the accuracy of both EM and AEM. Nonetheless, the adaptive augmentation of the EM algorithm remains complementary to the optimal level of noise as it leads to faster convergence for all noise levels than the standard EM algorithm.

Third, we conduct an empirical nowcasting exercise of euro area GDP growth based on a mixed-frequency dynamic factor model akin to Mariano and Murasawa (2003), which we estimate with both the EM and AEM algorithms. We find that the AEM algorithm is able to reach much higher log-likelihood values in much less iterations than EM, reconfirming the slower convergence of the standard algorithm. In addition, AEM produces more accurate nowcasts with improvements in root mean squared forecast error (RMSFE) up to 34% relative to the baseline EM.

Besides the Bańbura and Modugno (2014) approach, there exists at least one other way to estimate dynamic factor models with serially correlated idiosyncratic components. Specifically, Watson and Engle (1983) and Reis and Watson (2010) propose to include lags of the observed variables and latent factors in the measurement equation, which can then be estimated with the Expectation Conditional Maximization algorithm of Meng and Rubin (1993). The upside of this implementation is that its state dimension does not increase with N. By contrast, this dimension does increase in the framework of Bańbura and Modugno (2014), slowing down the filtering/smoothing recursions for large N. However, the downside of this alternative approach is that it is not directly compatible with arbitrary patterns of missing data (Jungbacker et al., 2011; Bańbura and Modugno, 2014). Although Jungbacker et al. (2011) propose an alternative state-space form to deal with missing data to overcome this deficiency, this comes at the cost of more complex time-varying state-space dimensions and system matrices, which are rather cumbersome to deal with in the EM algorithm.[3]

Our work also relates to the literature that extends the basic EM to speed up (global) convergence (see, among others, Liu et al., 1998; Huang et al., 2005; Varadhan and Roland, 2008; He and Liu, 2012). Notably, Osoba et al. (2013) derive the Noisy EM algorithm and show that careful additive and arbitrary noise injection could speed up the EM convergence, which corroborates our findings that modest noise levels speed up the (A)EM convergence. Nevertheless, our paper focuses on a specific case of slow EM convergence, namely the one under a low-noise environment. Therefore, we restrict the paper to the AEM algorithm of Salakhutdinov and Roweis (2003) only, as this adaptation is able to naturally counter the slow EM convergence under this specific noise setting.

The outline of the paper is as follows. Section 2 describes the low-noise dynamic factor model and its estimation based on the EM and AEM algorithms. Section 3 displays the Monte Carlo simulation results to assess the effect of low noise on the estimation performance. Section 4 shows the results related to the empirical nowcasting exercise. Section 5 summarizes our main conclusions.

2 Estimation of low-noise dynamic factor models

2.1 Low-noise dynamic factor model

Let $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ denote an N-dimensional vector with observed time series that is demeaned to have mean zero. We assume that $\boldsymbol{y}_t$ follows the factor model representation

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

for $t = 1, \ldots, T$, where $\boldsymbol{f}_t = (f_{1,t}, \ldots, f_{K,t})'$ is a $K \times 1$ vector with latent common factors, $\boldsymbol{\Lambda}$ is an $N \times K$ factor loading matrix and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})'$ is an $N \times 1$ vector with idiosyncratic components that are uncorrelated with $\boldsymbol{f}_t$ at all leads and lags. For now, we assume that $\boldsymbol{\varepsilon}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega}$ being a diagonal matrix, meaning that the $\boldsymbol{\varepsilon}_t$'s are cross-sectionally and serially uncorrelated and thus that $\boldsymbol{y}_t$ follows an exact factor model structure. Moreover, we assume that $\boldsymbol{f}_t$ follows a stationary vector autoregression

[3] See Grassi et al. (2015) for an empirical implementation and Holmes (2018) for more general derivations of the EM algorithm under deterministic time-varying parameter system matrices.

(VAR) of a finite order p, that is,

$$\boldsymbol{f}_t = \boldsymbol{\Phi}_1 \boldsymbol{f}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{f}_{t-p} + \boldsymbol{v}_t,$$

where $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p$ are $K \times K$ matrices with VAR coefficients. We assume that the state disturbance vector $\boldsymbol{v}_t \sim$ i.i.d. $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K)$ with $\boldsymbol{I}_K$ being a $K \times K$ identity matrix, where this covariance matrix restriction is only a normalization condition (see, for example, Doz et al., 2012). For simplicity, we set $p = 1$, but the case of $p > 1$ can easily be accommodated.

The exact dynamic factor model as specified above assumes that all cross-sectional dependence and time-series dependence in $\boldsymbol{y}_t$ is captured by the common factors $\boldsymbol{f}_t$. However, this assumption might be too restrictive and could be relaxed in two possible ways. First, allowing for cross-sectional dependence in $\boldsymbol{\varepsilon}_t$ results in a so-called approximate factor model (see, for example, Chamberlain and Rothschild, 1983; Fan et al., 2013; Bai and Liao, 2016) and could lead to a more efficient estimator of the latent factors (Barigozzi and Luciani, 2022).[4] Yet, estimating the full idiosyncratic covariance matrix becomes problematic for large N (Poncela et al., 2021). In fact, no version of the EM algorithm currently exists that is able to do so in a high-dimensional setting (Barigozzi and Luciani, 2022), making this extension not readily available to implement. Also, Luciani (2014) empirically shows that accounting for cross-sectional correlation does not lead to improvements in forecasting accuracy. Second, allowing for serial correlation in $\boldsymbol{\varepsilon}_t$ could lead to more efficient estimators of the factor loadings (Bai and Li, 2016; Barigozzi and Luciani, 2022). Moreover, modelling the dynamics in the idiosyncratic components could be beneficial in certain applications such as in the construction of coincident economic indicators (Stock and Watson, 1989, 1991; Mariano and Murasawa, 2003) and in forecasting and nowcasting (Stock and Watson, 2002b; Poncela et al., 2020), especially for ragged-edge data (Pinheiro et al., 2013; Bańbura and Modugno, 2014).

To explicitly model the autocorrelation of the idiosyncratic components, Bańbura and Modugno (2014) propose to treat the vector $\boldsymbol{\varepsilon}_t$ as additional latent state and to introduce an artificial noise term $\boldsymbol{e}_t$ with small variance.[5] Consequently, measurement equation (1)

[4] Doz et al. (2012) show that, under weak cross-sectional and time-series correlation, the factors can still be consistently estimated with EM in an exact dynamic factor model for $N, T \to \infty$.

[5] This approach is also used by Barigozzi and Luciani (2019) to deal with nonstationary idiosyncratic components in a nonstationary dynamic factor model.

can be rewritten as

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{I}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} + \boldsymbol{e}_t, \tag{2}$$

where $\boldsymbol{e}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \kappa \boldsymbol{I}_N)$ with κ a very small pre-fixed number (say, 10^{-4}). We continue to assume that $\boldsymbol{f}_t$ follows a stationary VAR(1) process and additionally that each element of $\boldsymbol{\varepsilon}_t$ follows a stationary univariate AR(1) process. The dynamics of $\tilde{\boldsymbol{f}}_t = (\boldsymbol{f}_t, \boldsymbol{\varepsilon}_t)'$ thus are given by

$$\begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix} \begin{pmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{\varepsilon}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\upsilon}_t \\ \boldsymbol{\nu}_t \end{pmatrix}, \tag{3}$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_N)$ and we assume that the error terms $\boldsymbol{\upsilon}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_K)$ and $\boldsymbol{\nu}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ are uncorrelated, with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$.

The reason that Bańbura and Modugno (2014) introduce the artificial error term $\boldsymbol{e}_t$ is to be able to properly define the complete data log-likelihood, because otherwise it is not possible to apply the EM algorithm in its usual form. However, in the next subsection, we show that this low-noise specification has severe implications for the convergence speed of the EM algorithm in the estimation of the factor loading matrix $\boldsymbol{\Lambda}$.

2.2 Failure of EM in a low-noise environment

Given measurement equation (2) and state equation (3), we want to estimate the unknown parameters, collected in $\boldsymbol{\Theta} = \{\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}$, and the latent states $\tilde{\boldsymbol{f}}_t$. Due to the fact that $\tilde{\boldsymbol{f}}_t$ is unobserved, it is generally not possible to find closed-form estimators for the parameters in $\boldsymbol{\Theta}$. At the same time, direct numerical optimization of the likelihood is computationally cumbersome, particularly for large N due to the excessive number of parameters. To handle this issue, the Expectation-Maximization (EM) algorithm of Dempster et al. (1977) has become a popular alternative estimation method, which has been adapted by Shumway and Stoffer (1982) and Watson and Engle (1983) for dynamic factor models in state-space form. The EM algorithm focuses on the joint log-likelihood of the complete data $\tilde{\boldsymbol{f}}_t$ and $\boldsymbol{y}_t$ and then iterates between estimating the latent states conditional on $\boldsymbol{\Theta}$ (E-step) and estimating the parameters conditional on the states (M-step).

More formally, the complete data log-likelihood is denoted as $\ell(\tilde{\boldsymbol{F}}, \boldsymbol{Y}; \boldsymbol{\Theta})$, where $\tilde{\boldsymbol{F}} = (\tilde{\boldsymbol{f}}_1, \ldots, \tilde{\boldsymbol{f}}_T)'$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$. The E-step is conducted by taking the expectation of the complete data log-likelihood conditional on the observed data and based on the j-th iteration of the parameter estimates, denoted as $\boldsymbol{\Theta}_j$, that is,

$$\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{\Theta}_j) = \mathbb{E}_{\boldsymbol{\Theta}_j}\Big(\ell(\tilde{\boldsymbol{F}}, \boldsymbol{Y}; \boldsymbol{\Theta})\big|\boldsymbol{Y}\Big), \tag{4}$$

which can be computed based on a pass of the Kalman smoother (see, for example, Shumway and Stoffer, 1982). Next, to update the parameter estimates, the M-step is conducted by maximizing the expected complete data log-likelihood with respect to $\boldsymbol{\Theta}$, that is,

$$\boldsymbol{\Theta}_{j+1} = \arg\max_{\boldsymbol{\Theta}} \mathcal{L}\big(\boldsymbol{\Theta}; \boldsymbol{\Theta}_j\big). \tag{5}$$

Analytic solutions to the maximization problem in equation (5) are given in Shumway and Stoffer (1982) and Watson and Engle (1983) for the system matrices in measurement equation (2) and a state equation corresponding to a VAR(1) process. Hence, by iterating between the E- and M-steps, we are able to estimate $\boldsymbol{\Theta}$ and $\tilde{\boldsymbol{F}}$, where Dempster et al. (1977) show that, under some regularity conditions, the EM algorithm converges towards a local maximum of the likelihood.

Based on the maximization in equation (5), we can derive the M-step of $\boldsymbol{\Lambda}$ as

$$\boldsymbol{\Lambda}_{j+1} = \left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\Theta}_j}\Big((\boldsymbol{y}_t - \boldsymbol{\varepsilon}_t)\boldsymbol{f}_t'\big|\boldsymbol{Y}\Big)\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\Theta}_j}\Big(\boldsymbol{f}_t\boldsymbol{f}_t'\big|\boldsymbol{Y}\Big)\right)^{-1}, \tag{6}$$

where $\mathbb{E}_{\boldsymbol{\Theta}_j}(\boldsymbol{f}_t|\boldsymbol{Y})$, $\mathbb{E}_{\boldsymbol{\Theta}_j}(\boldsymbol{\varepsilon}_t\boldsymbol{f}_t'|\boldsymbol{Y})$ and $\mathbb{E}_{\boldsymbol{\Theta}_j}(\boldsymbol{f}_t\boldsymbol{f}_t'|\boldsymbol{Y})$ can be obtained with the Kalman smoother. Similar expressions can be obtained for the other system matrices in the state-space representation of the dynamic factor model (see, for example, Bańbura and Modugno, 2014). Plugging in measurement equation (2) for the j-th EM parameter iteration, we obtain

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

Consequently, by decreasing the variance of $\boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0}, \kappa \boldsymbol{I}_N)$, that is, by letting $\kappa \to 0$, we get that $\boldsymbol{\Lambda}_{j+1} \to \boldsymbol{\Lambda}_j$. More formally, Bermond and Cardoso (1999) and Petersen et al. (2005) show that

$$\boldsymbol{\Lambda}_{j+1} = \boldsymbol{\Lambda}_j + \kappa \tilde{\boldsymbol{\Lambda}}_j + \mathcal{O}(\kappa^4), \tag{7}$$

where $\tilde{\boldsymbol{\Lambda}}_j$ is the first-order correction term, see the Technical Appendix of Petersen et al. (2005) for more details. In other words, the learning rate of the M-step for the factor loadings $\boldsymbol{\Lambda}$ is proportional to the noise level κ of the artificial error term. Hence, the convergence of the EM algorithm slows down for small values of κ and requires an excessive number of iterations in order to converge.[6]

2.3 Adaptive EM in a low-noise environment

To speed up the EM convergence in the low-noise setting, we advocate to employ the Adaptive Overrelaxed EM (AEM) algorithm of Salakhutdinov and Roweis (2003), following the suggestion of Petersen et al. (2005). The idea behind this AEM algorithm is to boost the parameter updates in iteration $j + 1$ by an adaptive factor η_j. Specifically, the AEM algorithm employs the usual E- and M-steps as given in equations (4) and (5). However, after obtaining the factor loading estimates $\boldsymbol{\Lambda}_{j+1}$ with (6), these are further updated using the adaptive scheme

$$\boldsymbol{\Lambda}_{j+1}^{AEM} = \boldsymbol{\Lambda}_j^{AEM} + \eta_j \Big(\boldsymbol{\Lambda}_{j+1} - \boldsymbol{\Lambda}_j^{AEM} \Big). \tag{8}$$

Combining this with equation (7), which implies that $\boldsymbol{\Lambda}_{j+1} - \boldsymbol{\Lambda}_j^{AEM}$ can be written as $\kappa \tilde{\boldsymbol{\Lambda}}_j^{AEM} + \mathcal{O}(\kappa^4)$, we get that

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

[6] Indeed, Coroneo et al. (2016) find that using initial or final estimates gives similar results, which they attribute to the fact that the two-step approach of Doz et al. (2011), used for initialization, and the maximum likelihood approach have similar properties under a strong factor structure (see, for example, Doz et al., 2011, 2012). We alternatively argue that this could be due to the low-noise specification in which a large number of iterations only changes the estimates by a very small margin. In fact, if the used tolerance level ϵ is too large, then the EM algorithm stops after only a few iterations as the changes are too small, rather than that the EM algorithm reaches a local optimum (see, for example, Alvarez et al., 2016, who indicate convergence after only 3-4 iterations).

This expression shows that an adaptive factor $\eta_j > 1$ is able to directly counter the low-noise level κ, increasing the learning rate relative to the conventional EM and thereby potentially speeding up convergence. As the other parameters are all part of the state equation, their updates are not directly affected by the low-noise setting in the measurement equation. Therefore, we do not apply this boosted M-step to those parameters.

The choice of η_j in equation (8) determines how much the learning rate of the M-step is boosted, where $\eta_j = 1$ for all j returns the standard EM algorithm. Although there exists an optimal boosting factor η_j^* with regard to the global rate of convergence of the algorithm, it is computationally difficult to obtain (Salakhutdinov and Roweis, 2003). Instead, Salakhutdinov and Roweis (2003) propose to set $\eta_{j+1} = \alpha\eta_j$ to gradually increase the boosting factor, where they set $\alpha = 1.1$ and initialize the boosting factor with $\eta_1 = 1$.[7] The only downside of using an adaptive learning rate with this specification is that an increase in the likelihood of $\boldsymbol{Y}$ is not guaranteed anymore. In case the likelihood does not improve in iteration $j + 1$, Salakhutdinov and Roweis (2003) propose to omit the adaptive update in (8) and set $\boldsymbol{\Lambda}_{j+1}^{AEM} = \boldsymbol{\Lambda}_{j+1}$ as obtained from the conventional M-step in equation (6). Due to the low-noise environment $\boldsymbol{\Lambda}_{j+1}$ is close to $\boldsymbol{\Lambda}_j^{AEM}$ (see again equation (7)) and satisfies monotonic increments of the likelihood. In addition, the boosting factor is re-set to $\eta_{j+1} = 1$, after which the algorithm continues.[8]

3 Monte Carlo simulations

3.1 Simulation set-up

To assess the effect of the low-noise specification on factor and parameter estimation based on the (A)EM algorithm, we conduct a Monte Carlo simulation in a similar fashion as Doz et al. (2012) and Bańbura and Modugno (2014). We simulate data from the following dynamic factor set-up:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K),$$

[7] Salakhutdinov and Roweis (2003) did not find their results to be very sensitive to the setting of α, as long as it is is close to but greater than unity.

[8] Alternatively, Yu (2012) show how to restrict η_j such that monotonic convergence of the likelihood is ensured. However, their restriction is too conservative in our setting as we often need $\eta_j \gg 1$. Hence, in the low-noise setting it is more beneficial to monitor the likelihood value over the iterations.

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Psi}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \text{i.i.d. } \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

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

$$\Lambda_{i,j} \sim \text{i.i.d. } \mathcal{N}(0,1), \qquad i = 1, \ldots N,\ j = 1, \ldots, K,$$

$$\Phi_{i,j} = \begin{cases} \phi & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases} \qquad i, j = 1, \ldots, K,$$

$$\Psi_{i,j} = \begin{cases} \psi & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases} \qquad i, j = 1, \ldots, N,$$

$$\Sigma_{i,j} = \tau^{|i-j|}(1 - \psi^2)\sqrt{\gamma_i \gamma_j}, \qquad i, j = 1, \ldots, N,$$

$$\gamma_i = \frac{\beta_i}{1 - \beta_i} \frac{1}{1 - \phi^2} \sum_{j=1}^{R} \Lambda_{i,j}^2, \qquad \beta_i \sim \text{i.i.d. } \mathcal{U}([u, 1 - u]),$$

where the subscript i, j denotes the (i, j)-th element of the corresponding matrix. The specifications of $\Phi_{i,j}$ and $\Psi_{i,j}$ imply that both the common factors as well as the idiosyncratic components follow univariate AR(1) processes, with persistence ϕ and ψ, respectively. The parameter τ governs the degree of cross-sectional dependence of the idiosyncratic components, where $\tau = 0$ corresponds to an exact factor model and $\tau > 0$ to an approximate factor model. Moreover, β_i governs the inverse signal-to-noise ratio of variable i, that is, it is equal to the variance of $\varepsilon_{i,t}$ divided by the variance of $y_{i,t}$. For our baseline simulation experiment, we set $K = 1$, $\phi = 0.7$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$. Furthermore, we consider cross-sectional dimensions $N = 10, 20, 50$ and sample sizes $T = 50, 100$.[9]

Given the generated data, we estimate the exact low-noise dynamic factor model given in equations (2) and (3) with the noise parameter fixed at $\kappa = 10^{-4}$, following the value that is used in Bok et al. (2018).[10] To determine the convergence of the EM algorithm, the stopping rule of Doz et al. (2012) enjoys substantial popularity, where, for a maximum

[9]We restrict the simulation set-up to max $T = 100$ and $N = 50$ for computational considerations. Because the state dimension increases with N, the filtering/smoothing recursions slow down and the simulations become computationally cumbersome for larger cross-sectional dimensions (see, for instance, Jungbacker et al., 2011).

[10]The value is not explicitly mentioned in the paper, but is given in their code (see `https://github.com/FRBNY-TimeSeriesAnalysis/Nowcasting`).

number of iterations J, the algorithm is stopped at the first iteration $j < J$ for which

$$\frac{|\ell(\boldsymbol{Y};\boldsymbol{\Theta}_j) - \ell(\boldsymbol{Y};\boldsymbol{\Theta}_{j-1})|}{\frac{1}{2}|\ell(\boldsymbol{Y};\boldsymbol{\Theta}_j) + \ell(\boldsymbol{Y};\boldsymbol{\Theta}_{j-1})|} < \epsilon, \tag{9}$$

where $\ell(\boldsymbol{Y};\boldsymbol{\Theta}_j)$ is the prediction error log-likelihood of $\boldsymbol{Y}$ computed at the j-th parameter iteration and ϵ is a pre-specified small tolerance level, which Doz et al. (2012) set equal to 10^{-4}. However, in our simulation setting, we simply conduct $J = 1,000$ iterations for both the EM and AEM algorithms to make our results insensitive to a specific tolerance level ϵ in the stopping rule.[11] We initialize the (A)EM algorithm with the two-step (2S) approach of Doz et al. (2012). Specifically, this 2S approach first estimates the loadings and common factors by means of principal component analysis, after which the idiosyncratic components are obtained as the residuals. The parameters in the state equation are estimated with OLS by using the factor and idiosyncratic component estimates (assuming they are the truth). Then, given the parameter estimates, a single pass of the Kalman smoother is used to get the final estimates of the latent factors and idiosyncratic components. For completeness, we also include this 2S approach in the simulation results.

To measure the precision of the parameter estimates in each Monte Carlo run, we follow Despois and Doz (2023) and compute the root mean squared errors (RMSE) for the different estimation methods $x \in \{2S, EM, AEM\}$, that is,

$$RMSE_{\Lambda}^{x} = \sqrt{\frac{1}{NK}\sum_{i=1}^{N}\sum_{j=1}^{K}(\Lambda_{i,j} - \hat{\Lambda}_{i,j}^{x})^2},$$

$$RMSE_{\phi}^{x} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\phi - \hat{\phi}_{i}^{x})^2},$$

$$RMSE_{\psi}^{x} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\psi - \hat{\psi}_{i}^{x})^2},$$

$$RMSE_{\Sigma}^{x} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Sigma_{i,i} - \hat{\Sigma}_{i,i}^{x})^2},$$

[11] Indeed, the stopping rule and tolerance level of Doz et al. (2012) employed by Bańbura and Modugno (2014) indicate convergence after only a few iterations, while it is clear from equation (7) that we need a substantial number of iterations for small κ.

where a lower RMSE indicates a more accurate estimation method.[12] To measure the precision of the common factor estimates in each Monte Carlo run, we follow Doz et al. (2012) and Bańbura and Modugno (2014) by using the trace R^2 of a (multivariate) regression of $\hat{\boldsymbol{F}}_x$ on $\boldsymbol{F}$ as proposed by Stock and Watson (2002a), that is,

$$R^2_{F,x} = \frac{\text{Tr}(\boldsymbol{F}'\hat{\boldsymbol{F}}_x(\hat{\boldsymbol{F}}_x'\hat{\boldsymbol{F}}_x)^{-1}\hat{\boldsymbol{F}}_x'\boldsymbol{F})}{\text{Tr}(\boldsymbol{F}'\boldsymbol{F})},$$

where $\hat{\boldsymbol{F}}_x$ is obtained by the Kalman smoother for $x \in \{2S, EM, AEM\}$ and a value of $R^2_{F,x}$ closer to one indicates a better approximation of the true factors. In a similar fashion, we compute the trace statistics for the idiosyncratic component estimates, resulting in $R^2_{E,x}$ for $x \in \{2S, EM, AEM\}$.

3.2 Simulation results

Table 1 shows the average RMSEs for the AEM parameter estimates and the average trace statistics for the AEM factor and idiosyncratic component estimates based on 500 Monte Carlo replications. Moreover, we show the relative RMSEs and relative trace statistics of the AEM algorithm compared to the ones of the 2S approach and the standard EM algorithm. A value smaller than one for the relative RMSEs indicates more accurate parameter estimates for the AEM algorithm compared to these benchmarks. Likewise, a value larger than one for the relative R^2's indicates more accurate factor and idiosyncratic component estimates for the AEM algorithm.

We find that the AEM algorithm produces more accurate estimates of the factor loadings than the standard EM algorithm for all sample sizes, with reductions in RMSEs ranging from a fairly modest 11% for $N = 50$ and $T = 100$ to an impressive 57% for $N = 10$ and $T = 100$. Overall, we see that larger cross-sections worsen the absolute RMSEs of the loading estimates from the AEM algorithm and make the outperformance compared to the EM algorithm less pronounced. This is due to the fact that the improvements in AEM accuracy over the iterations become slower for larger cross-sections, requiring more iterations to obtain convergence (see, for example, the convergence plots of the average RMSEs of the loadings over the $J = 1,000$ iterations in Figure 1). Nonetheless, the

[12] Since $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ are identified up to column permutations and sign changes, we follow the same identification scheme as Despois and Doz (2023) to switch the order and/or signs of the estimated factors and loadings accordingly.

Table 1: Monte Carlo results: RMSEs for model parameters and trace statistics for factor estimates and idiosyncratic component estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.35	0.50	0.76	0.29	0.49	0.77
	AEM/2S	0.50	0.64	0.87	0.42	0.61	0.88
	AEM/EM	0.50	0.64	0.88	0.43	0.62	0.89
ϕ	AEM	0.11	0.10	0.11	0.07	0.08	0.08
	AEM/2S	0.99	1.04	1.34	0.70	0.98	1.30
	AEM/EM	0.43	0.46	0.54	0.26	0.32	0.48
ψ	AEM	0.17	0.14	0.14	0.11	0.10	0.10
	AEM/2S	1.32	1.12	1.05	1.19	1.15	1.10
	AEM/EM	1.07	0.98	0.96	0.93	0.93	0.93
Σ	AEM	0.91	1.01	1.22	0.68	0.83	0.97
	AEM/2S	0.27	0.34	0.57	0.21	0.28	0.52
	AEM/EM	0.82	0.86	0.96	0.71	0.78	0.95
Panel B: Average and relative trace statistics							
R^2_F	AEM	0.90	0.95	0.97	0.92	0.95	0.98
	AEM/2S	1.36	1.18	1.06	1.35	1.14	1.03
	AEM/EM	1.23	1.15	1.10	1.23	1.17	1.08
R^2_E	AEM	0.96	0.98	1.00	0.96	0.98	0.99
	AEM/2S	1.00	0.99	1.00	1.00	0.99	1.00
	AEM/EM	1.01	1.00	1.00	1.01	1.00	1.00

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor and idiosyncratic component estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1{,}000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmark for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an exact factor model with $K = 1$, $\phi = 0.7$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$.

AEM algorithm still produces an average improvement in RMSE per iteration that is up to 883 times higher than the EM algorithm. The same observation applies to the time series dimension, albeit in the opposite direction (with improvements generally becoming larger for larger T, except for $N = 50$) and with considerably smaller effects compared to changes in N.

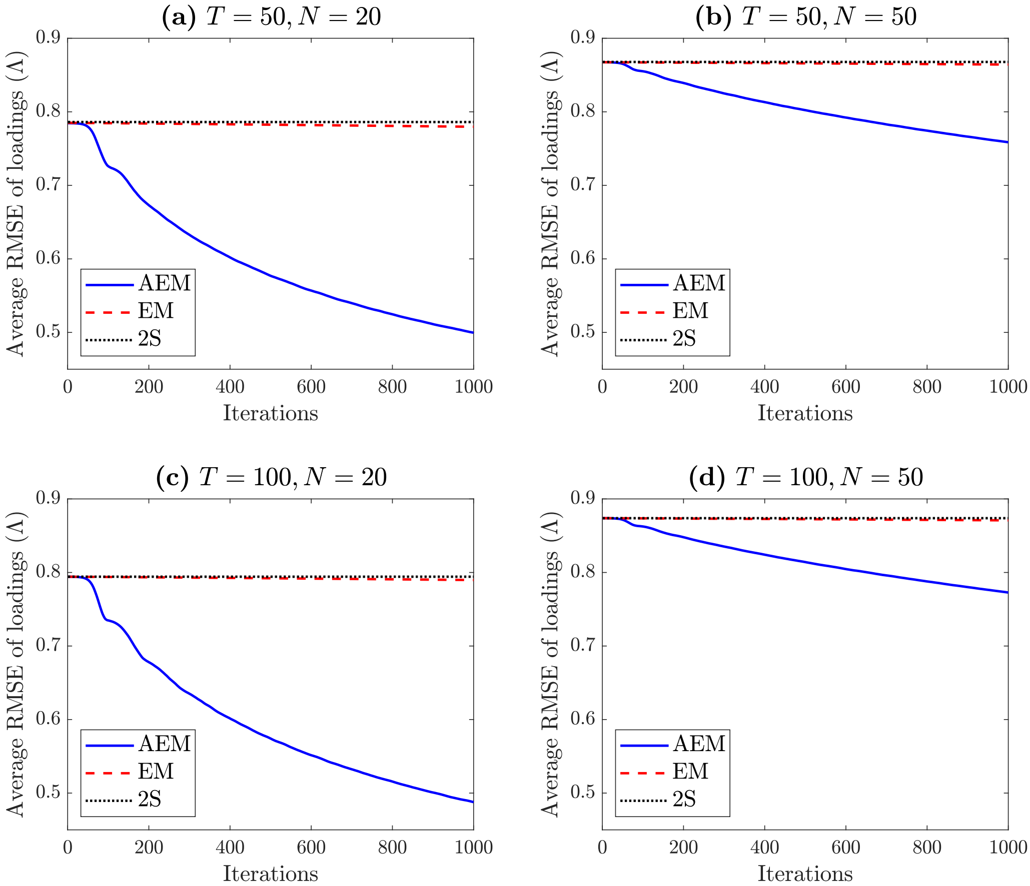

Figure 1: Convergence plots of average RMSEs of factor loading estimates ($\boldsymbol{\Lambda}$) based on the AEM algorithm, EM algorithm and 2S approach for various sample sizes and cross-sectional dimensions

It is also noteworthy that the estimation accuracy of the 2S approach for the loadings is comparable to the one of the baseline EM algorithm, highlighting that, even after 1,000 EM iterations, the factor loading estimates are still about the same as the values used for initialization. Indeed, this also becomes clear from the convergence plots in Figure 1. To quantify the movement of the (A)EM estimates away from the starting values, we compute root mean squared differences (RMSDs) between the (A)EM estimates and the 2S estimates. These RMSDs are about 0.13-0.48 for the AEM algorithm, whereas they are only about 0.01-0.02 for the EM algorithm, indicating that the 1,000 EM iterations result in only very minor changes in the factor loading estimates. This slow movement away from the initialization of the EM estimates clearly illustrates the low-noise issue that arises in the standard EM algorithm.

Moving to the other parameter estimates, we again find that the AEM algorithm is substantially more accurate than the EM algorithm, particularly for estimating the persistence of the latent factor. This results from the fact that inaccurate factor loading estimates lead to inaccurate factor estimates and, consequently, inaccurate factor persistence estimates. Indeed, the AEM algorithm produces a better approximation of the true factors than the EM algorithm for all sample sizes, where the increase in R^2 is about 23% for small N and 8-10% for large N. For large N, we even find that the trace statistics of the EM algorithm deteriorate compared to the 2S approach, implying that sticking with the 2S approach for large enough N results in more accurate factor estimates than by using the EM algorithm.

To illustrate this further, Figure 2 displays the convergence of the average trace statistics of the factor estimates (R_F^2) based on the different estimation methods over the $J = 1,000$ iterations. Here we clearly see that the trace statistics of the EM algorithm are well below the ones of the 2S approach for all cases (except $T = 50$ and $N = 20$), whereas the ones of the AEM algorithm all reach values above the 2S approach. In addition, the AEM algorithm reaches these higher values at a much faster pace than the EM algorithm, again illustrating the slow convergence of the EM algorithm in the low-noise

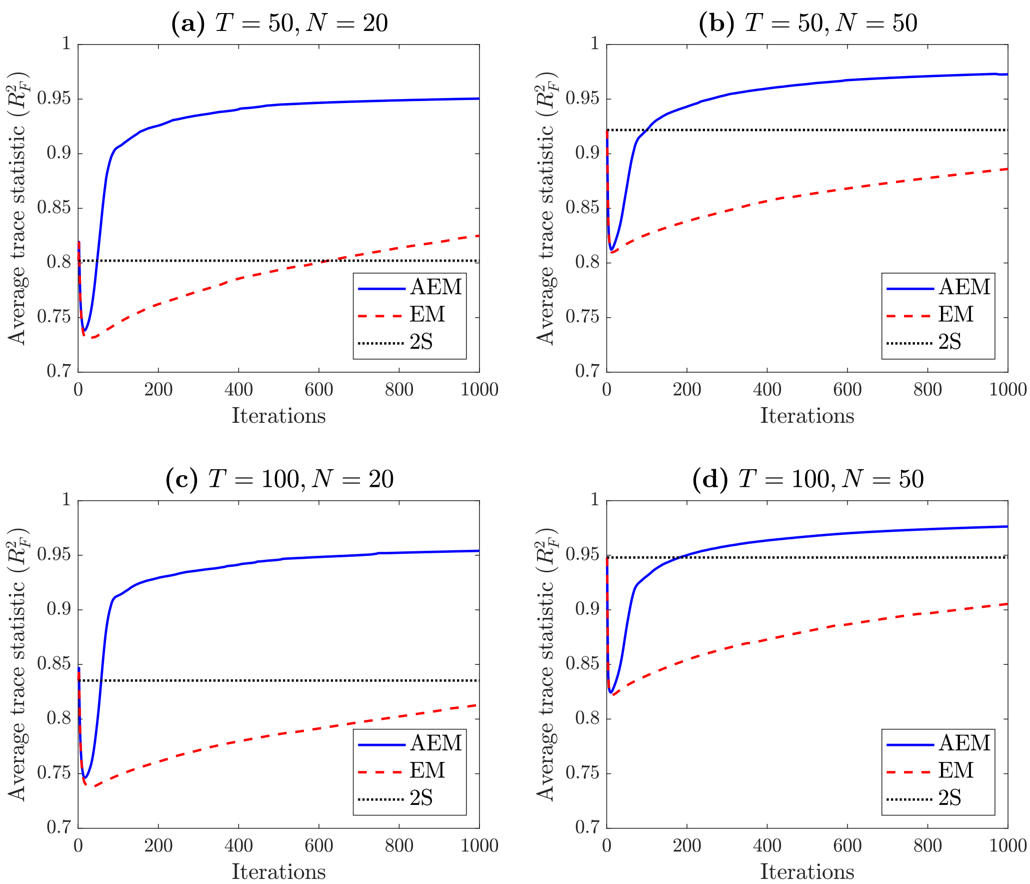

Figure 2: Convergence plots of average trace statistics of factor estimates (R_F^2) based on the AEM algorithm, EM algorithm and 2S approach for various sample sizes and cross-sectional dimensions

environment.

Finally, according to Table 1, the accuracy of the idiosyncratic component estimates is not affected by the low-noise setting or sample size, and only marginally by the cross-sectional dimension. Yet, the parameter estimates related to these components (ψ and $\boldsymbol{\Sigma}$) are generally more poorly estimated with the EM algorithm compared to the AEM algorithm, albeit to a lesser extent than for the factor estimates.

As a robustness check, we also explore other model (mis)specifications in Appendix A. In particular, we consider set-ups with cross-sectional dependence ($\tau = 0.5$), stronger factor persistence ($\phi = 0.9$), more factors ($K = 3$), no serial correlation in the idiosyncratic components ($\psi = 0$) and missing data issues. Overall, we find results that are qualitatively similar as the baseline results in Table 1, indicating that the superiority of the AEM algorithm over the EM algorithm persists across alternative model (mis)specifications.

3.3 Noise-level analysis

So far, we considered the setting where the noise variance is set equal to $\kappa = 10^{-4}$. To examine how the performance of the (A)EM algorithm varies across different levels of noise, we also conduct the Monte Carlo simulations for different values of $\kappa = 10^{z}$ with $z \in \{-6, \ldots, -1, 0, 1\}$. We use the same parameter settings as for Table 1, but we restrict ourselves to $N = 10$ and $T = 50$ due to computational costs.[13] Note that we do not include the results of the 2S approach anymore as its parameter estimates do not depend on the level of κ.

Table 2 shows the RMSEs and trace statistics for the AEM algorithm, also relative to the EM algorithm, based again on 500 Monte Carlo replications. We find that the accuracy of the factor loading estimates obtained with the AEM algorithm declines for low and high values of κ, where the minimum RMSE is attained around $\kappa = 10^{-1}$ or $\kappa = 10^{-2}$. The same holds for the factor persistence, albeit to a lesser extent. For the parameters related to the idiosyncratic components, the absolute RMSEs for the AEM algorithm are rather stable for small values of κ, whereas they start to increase for larger values. Similarly for both the factor estimates and idiosyncratic component estimates, we find

[13]We also experimented with different signal-to-noise ratios (that is, different specifications for β_i). The findings across different noise levels are generally robust to different signal-to-noise ratios, although weaker (stronger) signals make the results less (more) pronounced. Detailed results are available upon request.

Table 2: Monte Carlo results for different values of κ: RMSEs for model parameters and trace statistics for factor estimates and idiosyncratic component estimates

		κ							
		10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
Panel A: Average and relative RMSEs									
Λ	AEM	0.40	0.38	0.35	0.32	0.25	0.25	0.27	0.47
	AEM/EM	0.55	0.53	0.50	0.48	0.62	1.02	1.00	1.00
ϕ	AEM	0.11	0.11	0.11	0.10	0.09	0.09	0.09	0.14
	AEM/EM	0.31	0.31	0.43	0.53	0.91	1.00	0.99	1.00
ψ	AEM	0.16	0.16	0.17	0.17	0.18	0.23	0.31	0.37
	AEM/EM	1.17	1.11	1.07	1.03	1.02	1.00	1.00	1.00
Σ	AEM	0.97	0.93	0.91	0.98	0.98	0.93	1.48	3.90
	AEM/EM	0.79	0.77	0.82	0.84	1.02	1.01	1.00	1.00
Panel B: Average and relative trace statistics									
R^2_F	AEM	0.90	0.90	0.90	0.91	0.92	0.90	0.85	0.60
	AEM/EM	1.38	1.40	1.23	1.12	1.00	1.00	1.00	0.99
R^2_E	AEM	0.95	0.96	0.96	0.95	0.95	0.95	0.90	0.79
	AEM/EM	1.02	1.02	1.01	1.00	1.00	1.00	1.00	1.00

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor and idiosyncratic component estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with a variety of different values for κ. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmark for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The sample size and cross-sectional dimension are $T = 50$ and $N = 10$, respectively. The data is generated with an exact factor model with $K = 1$, $\phi = 0.9$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$.

stable absolute trace statistics for the AEM algorithm that hover around 0.90-0.92 and 0.95-0.96, respectively, for values of κ smaller than 10^{-1}. Yet, for $\kappa = 10^0$ and $\kappa = 10^1$, the trace statistics start to decrease, indicating poorer performance. Indeed, Barigozzi and Luciani (2019) argue that the larger the value of κ, the larger the misspecification of the model and thus the more it affects the factor estimates. Overall, we find that for small and moderate values of κ, the performance of the AEM algorithm is rather stable and not substantially influenced by the level of noise (except perhaps for the factor loadings), whereas for larger values of κ the performance deteriorates for all parameter and factor estimates.

Moving to the (relative) performance of the standard EM algorithm, Table 2 shows that increasing the level of κ makes the performance of the EM algorithm come closer to the one of the AEM algorithm for all parameter and factor estimates. Hence, using a slightly higher value of κ (say, 10^{-2}) results in more accurate EM estimates. On the other hand, a too high value of κ (say, 10^{1}) produces poor absolute RMSEs and trace statistics. Therefore, in practice, one should carefully set the arbitrary noise level of κ to be able to improve the accuracy of the (A)EM algorithm.

To illustrate the convergence speed of the different algorithms for different noise levels, Figure 3 shows the average RMSEs of the factor loading estimates over 1,000 iterations

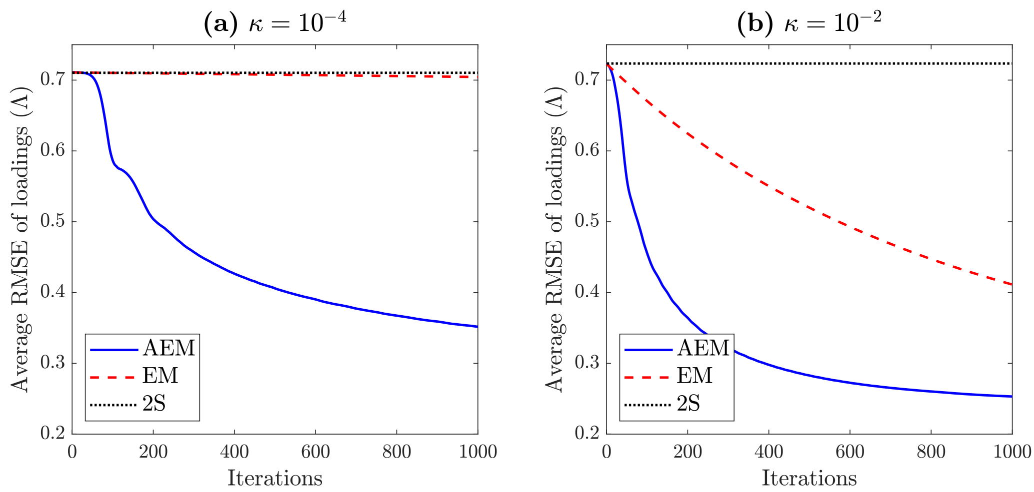

Figure 3: Convergence plots of average RMSEs of factor loading estimates ($\boldsymbol{\Lambda}$) based on the AEM algorithm, EM algorithm and 2S approach for various noise levels

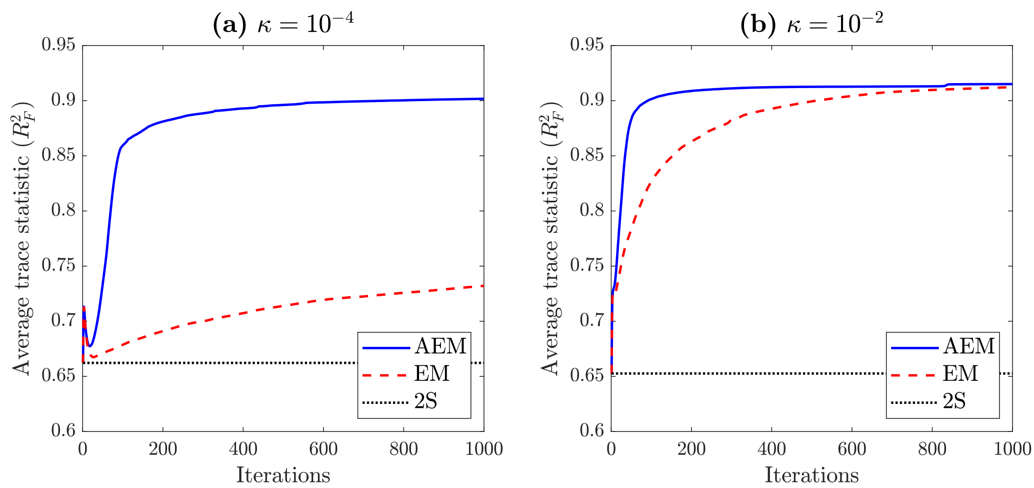

Figure 4: Convergence plots of average trace statistics of factor estimates (R_F^2) based on the AEM algorithm, EM algorithm and 2S approach for various noise levels

for $\kappa = 10^{-4}$ and $\kappa = 10^{-2}$. Clearly, the average RMSEs of the EM and AEM algorithms lie closer to each other after 1,000 iterations for $\kappa = 10^{-2}$ than for $\kappa = 10^{-4}$, although the accuracy and convergence speed is still much better for the AEM algorithm. Similar results are also found for the convergence plots of the average trace statistics of the factor estimates, which are given in Figure 4. The setting of $\kappa = 10^{-2}$ even seems to generate comparable factor estimates for EM and AEM after 1,000 iterations. The finding that higher levels of arbitrary noise speed up the convergence of the (A)EM algorithm corroborates with Osoba et al. (2013), who more generally show that careful additive noise injection can accelerate EM convergence. Overall, we conclude that using the AEM algorithm in combination with the right amount of arbitrary noise results in the fastest convergence, implying that these two manipulations of the traditional EM framework of Bańbura and Modugno (2014) are complementary in improving the estimation of low-noise dynamic factor models.

4 Empirical application

We conduct a nowcasting exercise of euro area GDP growth to examine the effect of the slow EM convergence in an empirical setting. We construct a similar euro area macroeconomic dataset as used by Bańbura and Rünstler (2011) and Bańbura and Modugno (2014), consisting of real economic, survey-based and financial variables. The resulting dataset contains 78 monthly series and one quarterly series, namely euro area GDP, and runs from January 1991 to December 2022. All variables are transformed into stationary time series by taking the natural logarithm and/or first differences. Similarly as Bańbura and Modugno (2014), the series are used to construct small ($N = 10$), medium ($N = 36$) and large ($N = 79$) datasets, making it possible to compare the estimation and nowcasting performance for different cross-sectional dimensions N. A complete description of all the series, transformations, publication delays and sources is given in Appendix B, together with the exact compositions of the small, medium and large datasets.

For each dataset, we estimate the exact factor model given in equations (2) and (3) with serial correlation in the idiosyncratic components and with the noise parameter fixed at either $\kappa = 10^{-4}$ or $\kappa = 10^{-2}$. Moreover, to handle the mixed-frequency nature of the data, we impose the temporal aggregation framework of Mariano and Murasawa (2003) to link the monthly factor and idiosyncratic component to the quarterly variable, just as in Bańbura and Modugno (2014). Both the EM and AEM algorithm are initialized with the 2S approach of Doz et al. (2011). All series are normalized to have zero mean and unit variance.

Before we move to the nowcasting exercise, we first show the log-likelihood values over the EM and AEM iterations when using the complete sample period to estimate the mixed-frequency dynamic factor model with the small, medium and large datasets in Figure 5. Clearly, for both values of κ and all N, the AEM algorithm leads to much larger increments and much faster convergence of the log-likelihood than the EM algorithm. For $\kappa = 10^{-4}$, the AEM algorithm generally converges somewhere in the range of 100-200 iterations, while the EM algorithm still has not converged after 1,000 iterations. In fact,

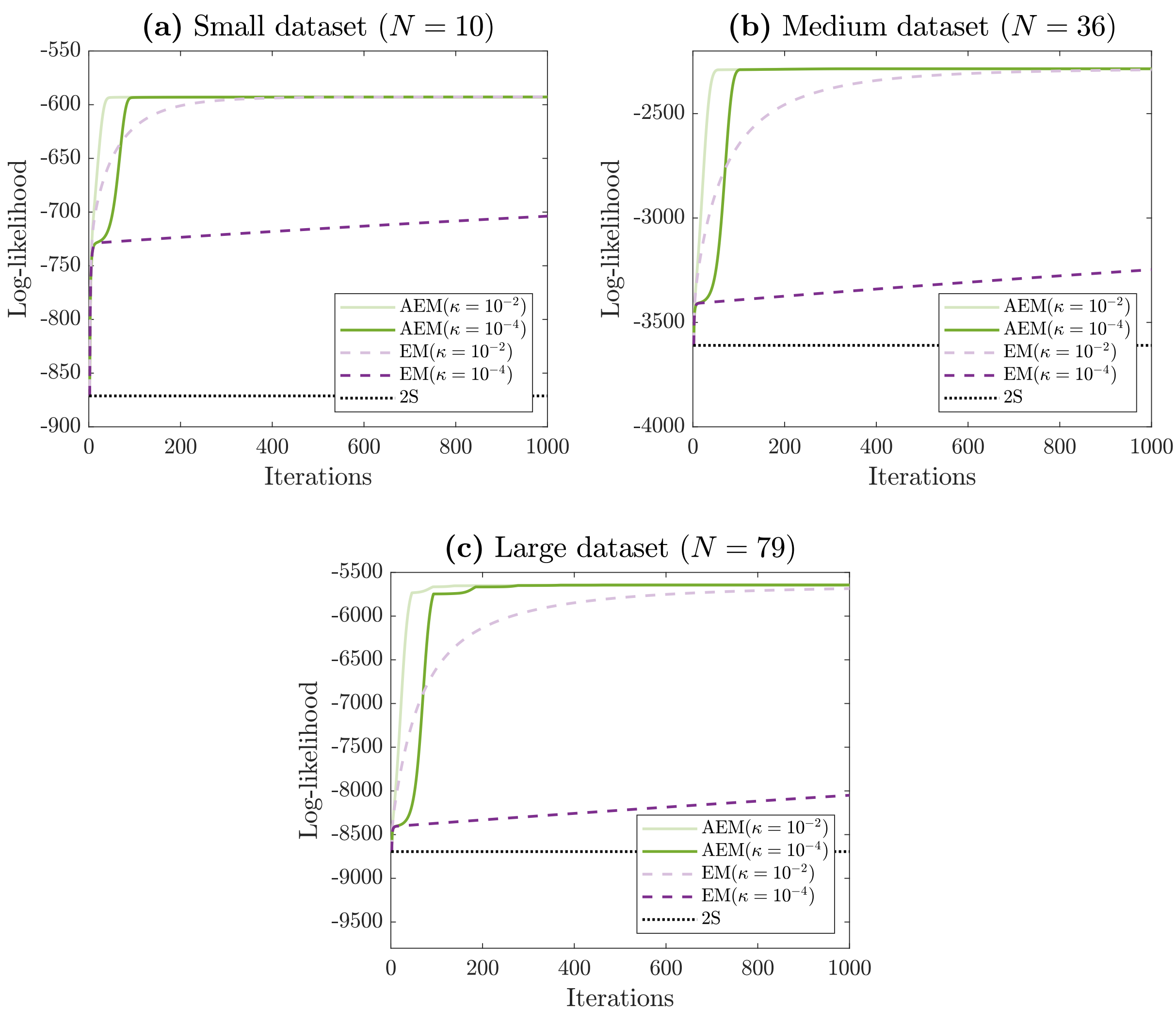

Figure 5: Log-likelihood values over 1,000 iterations of the EM and AEM algorithm in estimating a mixed-frequency dynamic factor model for various cross-sectional dimensions N

Figure C.1 shows that the EM algorithm has not even converged after 10,000 iterations, where the slope of the increments declines over the iterations. On the other hand, Figure 5 also illustrates that we could considerably speed up the convergence of both algorithms by increasing the artificial noise level to $\kappa = 10^{-2}$, although AEM still converges faster than the usual EM in that case. It is also useful to note that the initial increase in the likelihood over the first EM and AEM iterations that is observed under $\kappa = 10^{-4}$ is due to the M-steps of all but the factor loading parameter estimates. Indeed, Figure C.2 shows that keeping the estimates of $\boldsymbol{\Lambda}$ fixed in the M-step of the EM algorithm generates a similar initial increase in likelihood, where its subsequent likelihood values remain rather close to the ones of the EM algorithm in which $\boldsymbol{\Lambda}$ is updated. This again confirms the slow EM convergence of the loading estimates, where not updating them leads to similar results.

Next, we conduct an expanding-window nowcasting exercise for quarterly GDP growth, where at each point in time we take into account the publication delays of the series (which are given in Appendix B) and impose this ragged edge structure onto the data.[14] For each target quarter, we construct a similar sequence of nowcasts and forecasts as in Bańbura and Modugno (2014), starting in the first month of the previous quarter (that is, Q(-1)M1) up to the first month of the subsequent quarter (that is, Q(+1)M1), which leads to seven predictions for each quarter. We use the evaluation sample from 2006Q1 to 2022Q3, such that the first estimation sample runs from January 1991 to October 2005 to produce Q(-1)M1 for 2006Q1. Just as in the simulation study, we run a fixed number of iterations to make the comparison of the EM and AEM algorithm insensitive to the stopping rule.[15] Specifically, we employ $1,000$ iterations for the small dataset, 500 iterations for the medium dataset and 100 iterations for the large dataset, which is motivated by the fact that the filtering/smoothing recursions become slower for large N. Besides the nowcasts from EM and AEM, we also include nowcasts based on the 2S approach, a first-order autoregression for GDP growth, and the historical mean of past

[14] This exercise is pseudo real-time, though, as real-time vintages are not available for all series over the full period. The current vintage that is used is January 2023.

[15] In Appendix D.1 we show the nowcasting results for different values of ϵ in the stopping rule in equation (9), highlighting that these results are indeed sensitive to the chosen value of ϵ. For $\epsilon = 10^{-4}$, both the EM and AEM algorithm generally indicate convergence after only 20 iterations, leading to poor nowcasting performance. For $\epsilon = 10^{-5}$ and $\epsilon = 10^{-6}$, the number of iterations needed increases, especially during the financial crisis and covid pandemic, leading to pronounced nowcasting gains for AEM and only marginal gains for EM.

GDP growth.[16]

Table 3 shows the relative root mean squared forecast errors (RMSFEs) of the sequence of nowcasts for the various methods compared to the historical mean, as well as the absolute RMSFEs for the historical mean in the rightmost column. Panel A shows the nowcasting results for the complete out-of-sample period until 2022Q3, while Panel B excludes the covid pandemic and reports results until 2019Q4. For almost all models and estimation methods, we observe gains in accuracy when more information becomes available. Moreover, the mixed-frequency dynamic factor models generally outperform the AR(1) and historical mean benchmarks, irrespective of the estimation method, especially when more information becomes available. Zooming in on Panel A, we find that the

Table 3: Relative nowcasting performance of euro area GDP growth based on mixed-frequency dynamic factor models with $\kappa = 10^{-4}$

	Small			**Medium**			**Large**			**Benchmarks**	
	2S	**EM**	**AEM**	**2S**	**EM**	**AEM**	**2S**	**EM**	**AEM**	**AR**	**Mean**
Panel A: Evaluation period including covid pandemic (2006Q3 - 2022Q3)											
Q(-1)M1	1.02	1.17	1.39	1.29	1.93	1.87	1.15	1.64	5.08	1.07	223.3
Q(-1)M2	1.02	1.11	1.01	0.98	0.98	0.94	1.01	1.05	0.96	1.39	223.4
Q(-1)M3	0.91	0.89	0.84	0.86	0.81	0.78	0.90	0.88	0.80	1.39	223.4
Q(0)M1	0.76	0.64	0.46	0.63	0.57	0.50	0.70	0.63	1.24	1.39	223.4
Q(0)M2	0.57	0.57	0.41	0.64	0.66	0.44	0.68	0.61	0.41	1.19	223.9
Q(0)M3	0.58	0.61	0.49	0.62	0.65	0.48	0.64	0.58	0.45	1.19	223.9
Q(+1)M1	0.62	0.66	0.53	0.66	0.68	0.52	0.69	0.64	0.54	1.19	223.9
Average	0.78	0.81	0.73	0.81	0.90	0.79	0.82	0.86	1.35	1.26	223.6
Panel B: Evaluation period excluding covid pandemic (2006Q3 - 2019Q4)											
Q(-1)M1	0.98	0.94	0.82	0.96	0.84	0.88	0.97	0.86	0.87	1.13	71.0
Q(-1)M2	0.95	0.89	0.79	0.90	0.76	0.81	0.92	0.79	0.82	1.05	70.5
Q(-1)M3	0.84	0.73	0.66	0.83	0.72	0.77	0.84	0.73	0.76	1.05	70.5
Q(0)M1	0.76	0.68	0.61	0.80	0.74	0.82	0.80	0.73	0.81	1.05	70.5
Q(0)M2	0.63	0.60	0.54	0.68	0.68	0.73	0.68	0.67	0.73	0.98	69.7
Q(0)M3	0.52	0.49	0.43	0.61	0.61	0.66	0.59	0.58	0.57	0.98	69.7
Q(+1)M1	0.50	0.49	0.45	0.58	0.59	0.63	0.56	0.55	0.51	0.98	69.7
Average	0.74	0.69	0.62	0.77	0.71	0.76	0.77	0.70	0.73	1.03	70.2

Notes: This table displays the relative root mean squared forecast errors (RMFSEs) of nowcasting euro area GDP growth from 2006Q1 to 2022Q4 compared to a historical mean nowcast for which the absolute RMSFEs are shown in basis points. Panel A shows the results including the covid pandemic, while Panel B excludes this period. The small, medium and large mixed-frequency dynamic factor models are estimated with $\kappa = 10^{-4}$ based on either the two-step (2S) approach of Doz et al. (2011), the EM algorithm employed in Bańbura and Modugno (2014) or the AEM algorithm of Salakhutdinov and Roweis (2003). We conduct $J = 1,000$, $J = 500$ and $J = 100$ (A)EM iterations for the small, medium and large models, respectively. We also include a first-order autoregression for GDP growth as benchmark. For each target quarter, the nowcasts/forecasts construction dates range from the first month of the previous quarter (that is, Q(-1)M1) up to the first month of the subsequent quarter (that is, Q(+1)M1).

[16] We have chosen the number of lags in the autoregressive model with the Akaike and Schwarz information criteria, where one lag is always optimal for both criteria.

EM algorithm often performs worse than the 2S approach for small- and medium-scale models. By contrast, the AEM algorithm performs substantially better than both 2S and EM for almost all projections with improvements in accuracy up to 34%. Yet, the AEM nowcasts from the large-scale model made in Q(-1)M1 and Q(0)M1 (and Q(-1)M1 from the small-scale model) appear to be less accurate than those based on the usual EM algorithm. Closer inspection of the relevant sequences of nowcasts reveals that this is due entirely to the nowcasts made in the first month of the quarter in the covid pandemic, that is, April 2020 (see Figure D.2). For the large dataset, leaving the nowcasts constructed in April 2020 out of consideration results in a relative RMSFEs for Q(-1)M1 of 0.99 and 0.99 for EM and AEM, respectively, while they are 0.79 and 0.69 for Q(0)M1. These relative RMSFEs obviously are much more in line with the patterns observed in other months, confirming that AEM produces more accurate nowcasts than EM, especially for nowcasts made in the target quarter.[17]

Moving to Panel B that excludes the pandemic period altogether, we observe that the absolute RMSFEs of the historical mean decreases by a factor of three, just as for the other methods. Furthermore, the EM and AEM algorithms both perform better than the 2S approach, emphasizing the added value of increasing the number of iterations. For the small-scale model, we find that the AEM algorithm still performs better than EM for all projections, albeit less strong than for the complete out-of-sample period in Panel A, with gains in accuracy up to 13%. On the other hand, for the medium- and large-scale models, the EM algorithm seems to perform slightly better than AEM, although the differences are never larger than 10%. This result is largely driven by the financial crisis period (see Figure D.2), in which the slower EM convergence seems to be beneficial compared to the faster AEM convergence. In other words, for medium and large datasets, the initial factor loading estimates of the 2S approach, which are close to the ones of EM due to slow convergence, lead to more accurate nowcasts than the factor loading estimates based on maximum likelihood with the AEM algorithm. Table D.2 indeed shows that the nowcasting performance is similar for the EM algorithm when the factor loadings are kept fixed instead of estimated. Nonetheless, adopting the (A)EM algorithm for larger N does not necessarily lead to better nowcasting performance compared to a small-scale model, which concurs with the notion that more data is not always better (Boivin and

[17]For a more thorough analysis of macroeconomic behaviour during the covid pandemic, see, among others, Ng (2021), Maroz et al. (2021), Schorfheide and Song (2022) and Foroni et al. (2022)

Ng, 2006).

Finally, we also construct nowcasts based on a model with $\kappa = 10^{-2}$. Table 4 shows the corresponding relative RMSFEs compared to the historical mean. For both evaluation periods, the accuracy of the EM algorithm is much closer to the AEM algorithm, especially for the small and medium datasets. Meanwhile, the performance of the AEM algorithm is comparable to the performance of the models with $\kappa = 10^{-4}$ in Table 3. This highlights that AEM is less sensitive to κ, while using a slightly higher value of κ for EM leads to more accurate nowcasts, at least for the small-scale model.

Table 4: Relative nowcasting performance of euro area GDP growth based on mixed-frequency dynamic factor models with $\kappa = 10^{-2}$

	Small		**Medium**		**Large**	
	EM	**AEM**	**EM**	**AEM**	**EM**	**AEM**
Panel A: Evaluation period including covid pandemic (2006Q1 - 2022Q3)						
Q(-1)M1	1.39	1.39	1.89	1.86	1.89	10.90
Q(-1)M2	1.01	1.01	0.94	0.94	1.02	0.96
Q(-1)M3	0.84	0.84	0.78	0.78	0.84	0.78
Q(0)M1	0.46	0.46	0.51	0.50	0.52	2.13
Q(0)M2	0.41	0.41	0.47	0.44	0.44	0.47
Q(0)M3	0.49	0.49	0.50	0.49	0.47	0.42
Q(+1)M1	0.53	0.53	0.54	0.52	0.58	0.51
Average	0.73	0.73	0.80	0.79	0.82	2.31
Panel B: Evaluation period excluding covid pandemic (2006Q1 - 2019Q4)						
Q(-1)M1	0.83	0.82	0.87	0.88	0.82	0.87
Q(-1)M2	0.79	0.80	0.81	0.82	0.79	0.81
Q(-1)M3	0.66	0.66	0.77	0.77	0.73	0.75
Q(0)M1	0.61	0.61	0.81	0.82	0.77	0.84
Q(0)M2	0.55	0.54	0.73	0.73	0.69	0.78
Q(0)M3	0.43	0.43	0.66	0.66	0.59	0.51
Q(+1)M1	0.46	0.46	0.63	0.63	0.56	0.44
Average	0.62	0.62	0.75	0.76	0.71	0.71

Notes: This table displays the relative root mean squared forecast errors (RMFSEs) of nowcasting euro area GDP growth from 2006Q1 to 2022Q4 compared to a historical mean nowcast. Panel A shows the results including the covid pandemic, while Panel B excludes this period. The small, medium and large mixed-frequency dynamic factor models are estimated with $\kappa = 10^{-2}$ based on either the EM algorithm employed in Bańbura and Modugno (2014) or the AEM algorithm of Salakhutdinov and Roweis (2003). We conduct $J = 1{,}000$, $J = 500$ and $J = 100$ (A)EM iterations for the small, medium and large models, respectively. For each target quarter, the nowcasts/forecasts construction dates range from the first month of the previous quarter (that is, Q(-1)M1) up to the first month of the subsequent quarter (that is, Q(+1)M1).

5 Conclusion

In this paper we address the slow Expectation-Maximization (EM) convergence that emerges in the estimation of low-noise dynamic factor models. Specifically, we show analytically and with Monte Carlo simulations that the popular framework of Bańbura and Modugno (2014), developed to explicitly model the dynamics of idiosyncratic components by means of an artificial low-noise specification, slows down the convergence speed of the EM factor loading estimates. This follows from the fact that the learning rate of the loading estimates is proportional to the variance of the noise in the measurement equation, implying that low noise leads to slow convergence. As a result, the estimation accuracy of the factor loading estimates deteriorates, which in turn affects the estimates of the latent factors and other model parameters.

To remedy these slow EM convergence issues, we advocate to employ the Adaptive Overrelaxed EM (AEM) algorithm of Salakhutdinov and Roweis (2003), as also suggested by Petersen et al. (2005). This straightforward adaptation of the standard EM algorithm boosts the learning rate of the factor loading estimates, thus countering the adverse effects of the low-noise level. Our simulation study shows that the AEM algorithm substantially increases the convergence speed and leads to accuracy gains up to 57% for the loadings (in terms of average root mean squared error) and 23% for the factors (in terms of average trace statistic). At the same time, carefully choosing the appropriate level of artificial noise could lead to even faster convergence for both EM and AEM. In practice, the right choice of artificial noise could properly be determined based on Monte Carlo simulations, where this noise level seems to be unrelated to the scale of the data and the signal-to-noise ratio. Still, for all levels of noise, the AEM algorithm remains faster in convergence than the EM algorithm, making the adaptive extension complementary to using the right amount of artificial noise. Finally, we show in an empirical application that a mixed-frequency dynamic factor model estimated with the AEM algorithm is able to produce more accurate euro area GDP nowcasts than when it is estimated with the standard EM algorithm with decreases in root mean squared forecast error up to 34%.

References

Alvarez, R., M. Camacho, and G. Perez-Quiros (2016): "Aggregate versus disaggregate information in dynamic factor models," *International Journal of Forecasting*, 32, 680–694.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27, 417–427.

Bai, J. and K. Li (2016): "Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension," *The Review of Economics and Statistics*, 98, 298–309.

Bai, J. and Y. Liao (2016): "Efficient estimation of approximate factor models via penalized maximum likelihood," *Journal of Econometrics*, 191, 1–18.

Barigozzi, M. and M. Luciani (2019): "Quasi Maximum Likelihood Estimation of Non-Stationary Large Approximate Dynamic Factor Models," Tech. Rep. arXiv:1910.09841, arXiv.

——— (2022): "Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM algorithm," Tech. Rep. arXiv:1910.03821, arXiv.

Bańbura, M., D. Giannone, and L. Reichlin (2011): "Nowcasting," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, Oxford University Press, 63–90.

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.

Bańbura, M. and G. Rünstler (2011): "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP," *International Journal of Forecasting*, 27, 333–346.

Bermond, O. and J. F. Cardoso (1999): “Approximate likelihood for noisy mixtures,” in *Proceedings of the First International Workshop on Independent Component Analysis and Blind Source Separation*, Aussois, France, 325–330.

Boivin, J. and S. Ng (2006): “Are more data always better for factor analysis?” *Journal of Econometrics*, 132, 169–194.

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018): “Macroeconomic Nowcasting and Forecasting with Big Data,” *Annual Review of Economics*, 10, 615–643.

Brave, S. and R. Butters (2011): “Monitoring financial stability: a financial conditions index approach,” *Economic Perspectives*, 35, 22–43.

Bräuning, F. and S. J. Koopman (2014): “Forecasting macroeconomic variables using collapsed dynamic factor analysis,” *International Journal of Forecasting*, 30, 572–584.

Caruso, A. and L. Coroneo (2023): “Does Real-Time Macroeconomic Information Help to Predict Interest Rates?” *Journal of Money, Credit and Banking*, Forthcoming.

Cascaldi-Garcia, D., T. R. Ferreira, D. Giannone, and M. Modugno (2021): “Back to the Present: Learning about the Euro Area through a Now-casting Model,” International Finance Discussion Paper 1312.

Chamberlain, G. and M. Rothschild (1983): “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.

Coroneo, L., D. Giannone, and M. Modugno (2016): “Unspanned Macroeconomic Factors in the Yield Curve,” *Journal of Business & Economic Statistics*, 34, 472–485.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Despois, T. and C. Doz (2023): “Identifying and interpreting the factors in factor models via sparsity: Different approaches,” *Journal of Applied Econometrics*, Forthcoming.

Doz, C., D. Giannone, and L. Reichlin (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164, 188–205.

——— (2012): "A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94, 1014–1024.

Engle, R. and M. Watson (1981): "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, 76, 774–781.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 75, 603–680.

Foroni, C., M. Marcellino, and D. Stevanovic (2022): "Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis," *International Journal of Forecasting*, 38, 596–612.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55, 665–676.

Grassi, S., T. Proietti, C. Frale, M. Marcellino, and G. Mazzi (2015): "EuroMInd-C: A disaggregate monthly indicator of economic activity for the Euro area and member countries," *International Journal of Forecasting*, 31, 712–738.

He, Y. and C. Liu (2012): "The dynamic 'expectation–conditional maximization either' algorithm," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 313–336.

Holmes, E. E. (2018): "Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models," Technical Report.

Huang, H.-S., B.-H. Yang, and C.-N. Hsu (2005): "Triple jump acceleration for the EM algorithm," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*.

Jungbacker, B. and S. J. Koopman (2015): "Likelihood-based dynamic factor analysis for measurement and forecasting," *The Econometrics Journal*, 18, 1–21.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2011): "Maximum likelihood estimation for dynamic factor models with missing data," *Journal of Economic Dynamics and Control*, 35, 1358–1368.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

Lippi, M., M. Deistler, and B. Anderson (2022): "High-Dimensional Dynamic Factor Models: A Selective Survey and Lines of Future Research," *Econometrics and Statistics*.

Litterman, R. B. and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *The Journal of Fixed Income*, 1, 54–61.

Liu, C., D. B. Rubin, and Y. N. Wu (1998): "Parameter Expansion to Accelerate EM: The PX-EM Algorithm," *Biometrika*, 85, 755–770.

Luciani, M. (2014): "Forecasting with approximate dynamic factor models: The role of non-pervasive shocks," *International Journal of Forecasting*, 30, 20–29.

Ludvigson, S. C. and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, 5027–5067.

Mariano, R. S. and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

Maroz, D., J. H. Stock, and M. W. Watson (2021): "Comovement of Economic Activity During the Covid Recession," Tech. rep.

Meng, X.-L. and D. B. Rubin (1993): "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika*, 80, 267–278.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014): "Forecasting the Equity Risk Premium: The Role of Technical Indicators," *Management Science*, 60, 1772–1791.

Ng, S. (2021): "Modeling Macroeconomic Variations After COVID-19," Tech. Rep. arXiv:2103.02732, arXiv.

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.

Pinheiro, M., A. Rua, and F. Dias (2013): "Dynamic Factor Models with Jagged Edge Panel Data: Taking on Board the Dynamics of the Idiosyncratic Components," *Oxford Bulletin of Economics and Statistics*, 75, 80–102.

Poncela, P., E. Ruiz, and K. Miranda (2021): "Factor extraction using Kalman filter and smoothing: This is not just another survey," *International Journal of Forecasting*, 37, 1399–1425.

Poncela, P., E. Senra, and L. P. Sierra (2020): "Global vs Sectoral Factors and the Impact of the Financialization in Commodity Price Changes," *Open Economies Review*, 31, 859–879.

Quah, D. and T. Sargent (1993): "A Dynamic Index Model for Large Cross Sections," in *Business Cycles, Indicators, and Forecasting*, ed. by J. Stock and M. Watson, University of Chicago Press, 285–309.

Reis, R. and M. W. Watson (2010): "Relative Goods' Prices, Pure Inflation, and The Phillips Correlation," *American Economic Journal: Macroeconomics*, 2, 128–157.

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.

Schorfheide, F. and D. Song (2022): "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic," *International Journal of Central Banking*, Forthcoming.

Scotti, C. (2016): "Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises," *Journal of Monetary Economics*, 82, 1–19.

Shumway, R. H. and D. S. Stoffer (1982): "An approach to time series smoothing and forecasting using the em algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Stock, J. H. and M. W. Watson (1989): "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 4, 351–394.

——— (1991): "A probability model of the coincident economic indicators," in *Leading Economic Indicators: New Approaches and Forecasting Records*, ed. by G. H. Moore and K. Lahiri, Cambridge: Cambridge University Press, 63–90.

——— (1999): "Forecasting inflation," *Journal of Monetary Economics*, 44, 293–335.

——— (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162.

Varadhan, R. and C. Roland (2008): "Simple and Globally Convergent Methods for Accelerating the Convergence of Any EM Algorithm," *Scandinavian Journal of Statistics*, 35, 335–353.

Watson, M. W. and R. F. Engle (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, 23, 385–400.

Yu, Y. (2012): "Monotonically Overrelaxed EM Algorithms," *Journal of Computational and Graphical Statistics*, 21, 518–537.

Slow Expectation-Maximization Convergence in Low-Noise Dynamic Factor Models

Online Appendix

Daan Opschoor Dick van Dijk

April 3, 2023

Contents

A Alternative Monte Carlo simulation set-ups

A.1 Cross-sectional dependence

Table A.1: Monte Carlo results under cross-sectional dependence: RMSEs for model parameters and trace statistics for factor estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.40	0.52	0.76	0.34	0.50	0.76
	AEM/2S	0.55	0.65	0.88	0.48	0.63	0.88
	AEM/EM	0.55	0.66	0.88	0.48	0.63	0.88
ϕ	AEM	0.11	0.10	0.11	0.08	0.07	0.08
	AEM/2S	0.88	1.07	1.34	0.81	0.95	1.35
	AEM/EM	0.37	0.41	0.56	0.32	0.36	0.44
ψ	AEM	0.18	0.15	0.14	0.12	0.10	0.10
	AEM/2S	1.41	1.16	1.06	1.37	1.15	1.10
	AEM/EM	1.11	0.98	0.96	0.99	0.94	0.91
Σ	AEM	1.04	1.06	1.24	0.82	0.84	0.95
	AEM/2S	0.29	0.32	0.56	0.22	0.27	0.43
	AEM/EM	0.90	0.88	0.96	0.83	0.81	0.94
Panel B: Average and relative trace statistics							
R_F^2	AEM	0.85	0.94	0.97	0.87	0.95	0.97
	AEM/2S	1.38	1.23	1.06	1.35	1.17	1.04
	AEM/EM	1.23	1.18	1.09	1.24	1.16	1.09

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmarks for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an approximate factor model with $K = 1$, $\phi = 0.7$, $\psi = 0.5$, $\tau = 0.5$ and $u = 0.1$.

A.2 Stronger factor persistence

Table A.2: Monte Carlo results under stronger factor persistence: RMSEs for model parameters and trace statistics for factor estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.43	0.57	0.79	0.36	0.53	0.78
	AEM/2S	0.58	0.70	0.90	0.50	0.67	0.90
	AEM/EM	0.58	0.70	0.90	0.50	0.67	0.90
ϕ	AEM	0.09	0.08	0.08	0.07	0.06	0.06
	AEM/2S	0.47	0.49	0.69	0.38	0.42	0.68
	AEM/EM	0.43	0.53	0.67	0.34	0.39	0.51
ψ	AEM	0.16	0.15	0.14	0.11	0.10	0.10
	AEM/2S	1.17	1.12	1.05	1.06	1.06	1.07
	AEM/EM	1.05	1.05	1.06	0.92	0.97	1.03
Σ	AEM	2.27	2.99	3.17	1.82	1.82	2.37
	AEM/2S	0.24	0.32	0.47	0.17	0.25	0.37
	AEM/EM	0.89	0.92	0.95	0.84	0.87	0.94
Panel B: Average and relative trace statistics							
R_F^2	AEM	0.88	0.91	0.93	0.90	0.94	0.94
	AEM/2S	1.46	1.26	1.11	1.43	1.18	1.08
	AEM/EM	1.35	1.25	1.17	1.32	1.22	1.15

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmarks for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an exact factor model with $K = 1$, $\phi = 0.9$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$.

A.3 More factors

Table A.3: Monte Carlo results under more factors: RMSEs for model parameters and trace statistics for factor estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.78	0.73	0.69	0.68	0.65	0.61
	AEM/2S	0.89	0.80	0.74	0.80	0.73	0.67
	AEM/EM	0.89	0.80	0.74	0.80	0.73	0.67
ϕ	AEM	0.23	0.15	0.13	0.15	0.10	0.09
	AEM/2S	1.47	1.14	1.34	1.10	0.85	1.07
	AEM/EM	0.54	0.41	0.39	0.36	0.27	0.28
ψ	AEM	0.33	0.23	0.17	0.22	0.13	0.11
	AEM/2S	2.60	1.80	1.30	2.17	1.32	1.05
	AEM/EM	2.47	1.83	1.33	2.26	1.44	1.20
Σ	AEM	3.34	3.05	2.98	2.05	1.89	2.06
	AEM/2S	0.38	0.39	0.57	0.24	0.26	0.43
	AEM/EM	0.93	0.80	0.78	0.67	0.57	0.63
Panel B: Average and relative trace statistics							
R_F^2	AEM	0.69	0.86	0.95	0.73	0.88	0.96
	AEM/2S	1.17	1.23	1.18	1.27	1.28	1.14
	AEM/EM	1.34	1.51	1.55	1.55	1.73	1.64

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor estimates in the exact dynamic factor model as given in equations (2) and (3) with AR(1) dynamics for both the factors and idiosyncratic components based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmarks for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an exact factor model with $K = 3$, $\phi = 0.7$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$.

A.4 No serial correlation in idiosyncratic components

Table A.4: Monte Carlo results under no serial correlation in idiosyncratic components: RMSEs for model parameters and trace statistics for factor estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.37	0.54	0.80	0.32	0.52	0.79
	AEM/2S	0.52	0.69	0.92	0.46	0.67	0.92
	AEM/EM	0.53	0.69	0.92	0.46	0.67	0.92
ϕ	AEM	0.13	0.13	0.16	0.09	0.10	0.15
	AEM/2S	0.40	0.50	0.97	0.32	0.45	1.17
	AEM/EM	0.43	0.46	0.68	0.31	0.38	0.65
ψ	AEM	0.17	0.16	0.16	0.12	0.12	0.11
	AEM/2S	0.98	0.97	1.08	0.78	0.91	1.07
	AEM/EM	0.88	0.92	1.04	0.73	0.86	1.04
Σ	AEM	1.20	1.35	1.63	0.93	1.08	1.27
	AEM/2S	0.27	0.34	0.54	0.20	0.26	0.47
	AEM/EM	0.87	0.87	0.92	0.74	0.82	0.96
Panel B: Average and relative trace statistics							
R^2_F	AEM	0.90	0.94	0.93	0.92	0.94	0.93
	AEM/2S	1.35	1.18	1.04	1.33	1.15	1.04
	AEM/EM	1.23	1.19	1.10	1.24	1.19	1.10

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmarks for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an exact factor model with $K = 1$, $\phi = 0.7$, $\psi = 0$, $\tau = 0$ and $u = 0.1$.

A.5 Missing data

Table A.5: Monte Carlo results under missing data: RMSEs for model parameters and trace statistics for factor estimates

		$T = 50$			$T = 100$		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
Panel A: Average and relative RMSEs							
Λ	AEM	0.36	0.52	0.77	0.32	0.52	0.75
	AEM/2S	0.50	0.65	0.88	0.46	0.65	0.87
	AEM/EM	0.51	0.66	0.89	0.46	0.66	0.88
ϕ	AEM	0.10	0.11	0.12	0.08	0.09	0.10
	AEM/2S	1.08	1.26	1.41	0.92	1.27	1.34
	AEM/EM	0.43	0.51	0.58	0.36	0.40	0.48
ψ	AEM	0.21	0.19	0.18	0.14	0.12	0.12
	AEM/2S	1.12	1.00	0.91	1.04	0.88	0.81
	AEM/EM	1.12	1.02	0.97	1.01	0.93	0.90
Σ	AEM	1.05	1.22	1.46	0.84	0.97	1.10
	AEM/2S	0.29	0.41	0.59	0.22	0.29	0.55
	AEM/EM	0.81	0.88	0.96	0.74	0.79	0.91
Panel B: Average and relative trace statistics							
R^2_F	AEM	0.88	0.93	0.95	0.90	0.94	0.96
	AEM/2S	1.48	1.29	1.15	1.52	1.28	1.12
	AEM/EM	1.30	1.21	1.13	1.29	1.23	1.14

Notes: This table displays average root mean squared errors (RMSE) of the estimation of $\boldsymbol{\Lambda}$, ϕ, ψ and $\boldsymbol{\Sigma}$ and the average trace statistics of the factor estimates in the exact dynamic factor model as given in equations (2) and (3) based on the overrelaxed adaptive EM (AEM) algorithm of Salakhutdinov and Roweis (2003). The model is estimated with $\kappa = 10^{-4}$. We also include the relative RMSEs and relative trace statistics of the AEM algorithm compared to the two-step (2S) approach of Doz et al. (2011) and the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$ (A)EM iterations. The AEM algorithm is more (less) accurate compared to its benchmarks for a value lower (higher) than one for the relative RMSEs and a value higher (lower) than one for the relative trace statistics. The averages and relative statistics are based on 500 Monte Carlo simulation runs. The values T and N denote the sample size and cross-sectional dimension, respectively. The data is generated with an exact factor model with $K = 1$, $\phi = 0.7$, $\psi = 0.5$, $\tau = 0$ and $u = 0.1$, where 25% of the data is set as missing (these points are chosen at random).

B Description of data

Table B.1: Description of euro area macroeconomic dataset

	Group	Series	Composition			Transform		Publ. lag	Source
			S	M	L	Log	Diff.		
1	Real	GDP	✓	✓	✓	✓	✓	2	SDW
2	Real	IP Total			✓	✓	✓	2	SDW
3	Real	IP Total (excluding construction)	✓	✓	✓	✓	✓	2	SDW
4	Real	IP Total (excluding construction and energy)			✓	✓	✓	2	SDW
5	Real	IP Construction		✓	✓	✓	✓	2	SDW
6	Real	IP Intermediate goods industry		✓	✓	✓	✓	2	SDW
7	Real	IP Capital goods industry		✓	✓	✓	✓	2	SDW
8	Real	IP Durable consumer goods industry		✓	✓	✓	✓	2	SDW
9	Real	IP Non-durable consumer goods		✓	✓	✓	✓	2	SDW
10	Real	IP Energy		✓	✓	✓	✓	2	SDW
11	Real	IP Manufacturing			✓	✓	✓	2	SDW
12	Real	IP Manufacture of basic metals			✓	✓	✓	2	SDW
13	Real	IP Manufacture of chemicals and chemical products			✓	✓	✓	2	SDW
14	Real	IP Manufacture of electrical equipment			✓	✓	✓	2	SDW
15	Real	IP Manufacture of machinery and equipment			✓	✓	✓	2	SDW
16	Real	IP Manufacture of paper and paper products			✓	✓	✓	2	SDW
17	Real	IP Manufacture of rubber and plastic products			✓	✓	✓	2	SDW
18	Real	New passenger car registration	✓	✓	✓	✓	✓	1	SDW
19	Real	Industrial new orders (total)	✓	✓	✓	✓	✓	1	SDW
20	Real	Retail trade turnover (deflated, incl. fuel, except of motor vehicles and motorcycles)	✓	✓	✓	✓	✓	1	SDW
21	Real	Unemployment rate	✓	✓	✓		✓	1	SDW
22	Real	Extra euro area trade (export value)	✓	✓	✓	✓	✓	2	SDW
23	Real	Extra euro area trade (import value)		✓	✓	✓	✓	2	SDW
24	Real	Intra euro area trade (export value)			✓	✓	✓	2	SDW
25	Real	Intra euro area trade (import value)			✓	✓	✓	2	SDW
26	Real	US IP Total		✓	✓	✓	✓	1	FRED
27	Real	US Manufacturing and trade industry sales			✓	✓	✓	1	FRED
28	Real	US Unemployment rate			✓		✓	1	FRED
29	Real	US Employment level			✓	✓	✓	1	FRED

Table B.1: Continued

	Group	Series	Composition			Transform		Publ. lag	Source
			S	M	L	Log	Diff.		
30	Survey	Economic sentiment indicator	✓	✓	✓		✓	0	EC
31	Survey	Employment expectations indicator		✓	✓		✓	0	EC
32	Survey	Industrial confidence indicator		✓	✓		✓	0	EC
33	Survey	Industry survey: Production trend observed in recent months			✓		✓	0	EC
34	Survey	Industry survey: Assessment of order-book levels			✓		✓	0	EC
35	Survey	Industry survey: Assessment of export order-book levels			✓		✓	0	EC
36	Survey	Industry survey: Assessment of stocks of finished products			✓		✓	0	EC
37	Survey	Industry survey: Production expectations for the months ahead		✓	✓		✓	0	EC
38	Survey	Industry survey: Employment expectations for the months ahead		✓	✓		✓	0	EC
39	Survey	Services confidence indicator		✓	✓		✓	0	EC
40	Survey	Service survey: Business situation development over the past 3 months			✓		✓	0	EC
41	Survey	Service survey: Evolution of the demand over the past 3 months			✓		✓	0	EC
42	Survey	Service survey: Expectation of the demand over the next 3 months		✓	✓		✓	0	EC
43	Survey	Service survey: Evolution of the employment over the past 3 months			✓		✓	0	EC
44	Survey	Service survey: Expectations of the employment over the next 3 months		✓	✓		✓	0	EC
45	Survey	Consumer confidence indicator		✓	✓		✓	0	EC
46	Survey	Consumer survey: General economic situation over last 12 months			✓		✓	0	EC
47	Survey	Consumer survey: General economic situation over next 12 months			✓		✓	0	EC
48	Survey	Consumer survey: Unemployment expectations over next 12 months			✓		✓	0	EC
49	Survey	Consumer survey: Major purchases at present			✓		✓	0	EC
50	Survey	Consumer survey: Major purchases over next 12 months			✓		✓	0	EC
51	Survey	Retail trade confidence indicator		✓	✓		✓	0	EC
52	Survey	Retail survey: Business activity (sales) development over the past 3 months			✓		✓	0	EC
53	Survey	Retail survey: Orders expectations over the next 3 months			✓		✓	0	EC
54	Survey	Retail survey: Business activity expectations over the next 3 months			✓		✓	0	EC
55	Survey	Retail survey: Employment expectations over the next 3 months			✓		✓	0	EC
56	Survey	Construction confidence indicator		✓	✓		✓	0	EC
57	Survey	Construction survey: Building activity development over the past 3 months			✓		✓	0	EC
58	Survey	Construction survey: Evolution of your current overall order books			✓		✓	0	EC
59	Survey	Construction survey: Employment expectations over the next 3 months			✓		✓	0	EC
60	Survey	US Consumer sentiment index		✓	✓		✓	1	FRED
61	Financial	Money aggregate M3 (index of notional stocks)		✓	✓	✓	✓	1	SDW

Table B.1: Continued

	Group	Series	Composition			Transform		Publ. lag	Source
			S	M	L	Log	Diff.		
62	Financial	3-month interest rate (Euribor)		✓	✓		✓	0	SDW
63	Financial	1-year government bond yield			✓		✓	0	SDW
64	Financial	2-year government bond yield			✓		✓	0	SDW
65	Financial	5-year government bond yield			✓		✓	0	SDW
66	Financial	10-year government bond yield		✓	✓		✓	0	SDW
67	Financial	Nominal effective exchange rate (Euro against Narrow EER)		✓	✓		✓	0	SDW
68	Financial	Real effective exchange rate CPI deflated (Euro against Narrow EER)			✓		✓	0	SDW
69	Financial	Real effective exchange rate producer prices deflated (Euro against Narrow EER)			✓		✓	0	SDW
70	Financial	Exchange rate USD/EUR		✓	✓		✓	0	SDW
71	Financial	Exchange rate GBP/EUR			✓		✓	0	SDW
72	Financial	Exchange rate YEN/EUR			✓		✓	0	SDW
73	Financial	Eurostoxx 50 index	✓	✓	✓	✓	✓	0	INV
74	Financial	Gold price			✓	✓	✓	0	INV
75	Financial	Brent crude oil price		✓	✓	✓	✓	0	INV
76	Financial	Global price index of all commodities	✓	✓	✓	✓	✓	1	FRED
77	Financial	US S&P500 composite index		✓	✓	✓	✓	0	FRED
78	Financial	US 3-month Treasury bill			✓		✓	0	FRED
79	Financial	US 10-year Treasury rate			✓		✓	0	FRED

Notes: This table describes the details of each series in the constructed euro area macroeconomic dataset. Specifically, it indicates the group it belongs to (that is, real, survey or financial), the dataset composition it corresponds to (that is, small (S), medium (M) or large (L)), the data transformation that is conducted on the original series (that is, taking the natural logarithm (log) and/or the first differences (diff.)), the publication lag in months, and the data source. The sources are the European Central Bank Statistical Data Warehouse (SDW), the Federal Reserve Economic Data (FRED), the European Comission (EC) and Investing.com (INV).

C Additional log-likelihood convergence plots

C.1 Log-likelihood convergence over 10,000 iterations

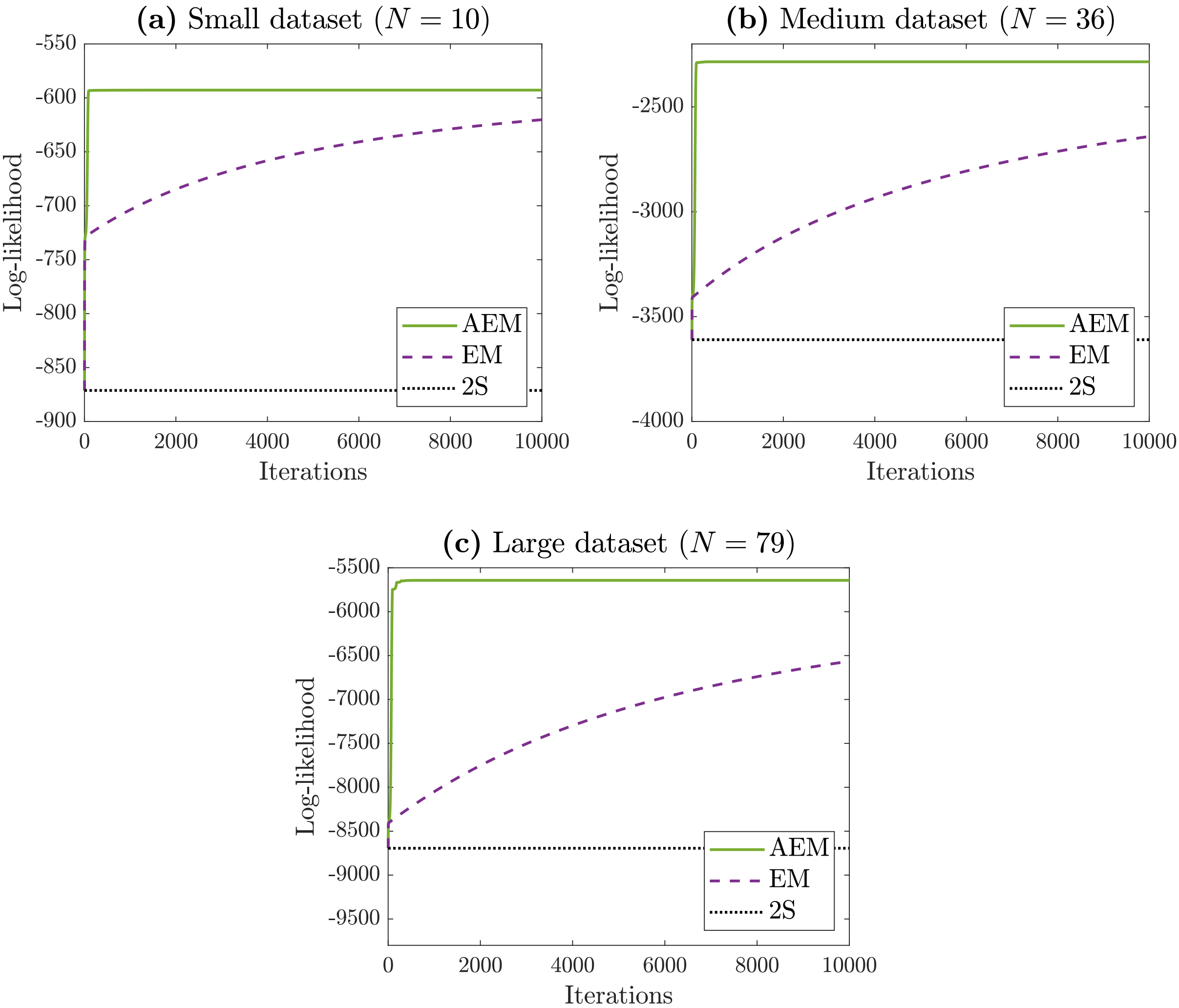

Figure C.1: Log-likelihood values over 10,000 iterations of the EM and AEM algorithm in estimating a mixed-frequency dynamic factor model with $\kappa = 10^{-4}$ for various cross-sectional dimension sizes N

C.2 Log-likelihood convergence under fixed factor loadings

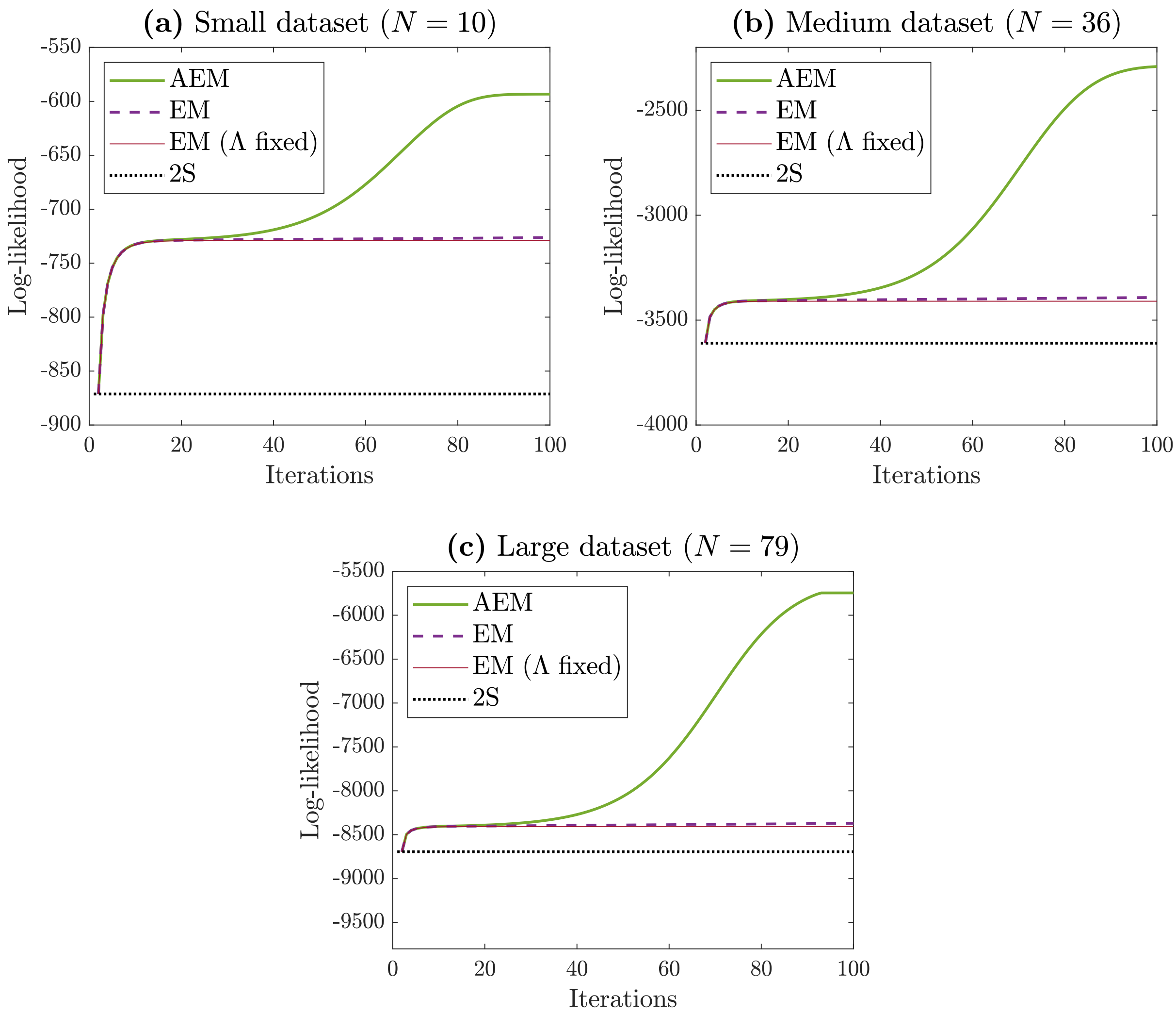

Figure C.2: Log-likelihood values over 100 iterations of the EM and AEM algorithm (including the EM under fixed factor loadings $\boldsymbol{\Lambda}$) in estimating a mixed-frequency dynamic factor model with $\kappa = 10^{-4}$ for various cross-sectional dimension sizes N

D Additional nowcasting results

D.1 Nowcasting results for different convergence thresholds

Table D.1: Relative nowcasting performance of euro area GDP growth based on a small mixed-frequency dynamic factor model for various values of ϵ in the stopping rule

	$\epsilon = 10^{-4}$		$\epsilon = 10^{-5}$		$\epsilon = 10^{-6}$	
	EM	**AEM**	**EM**	**AEM**	**EM**	**AEM**
Panel A: Evaluation period including covid pandemic (2006Q1 - 2022Q3)						
Q(-1)M1	1.14	1.14	1.17	1.39	1.17	1.39
Q(-1)M2	1.11	1.11	1.11	1.01	1.11	1.01
Q(-1)M3	0.91	0.86	0.90	0.85	0.89	0.84
Q(0)M1	0.67	0.60	0.64	0.46	0.64	0.46
Q(0)M2	0.50	0.44	0.58	0.40	0.57	0.40
Q(0)M3	0.56	0.47	0.62	0.46	0.61	0.49
Q(+1)M1	0.59	0.50	0.66	0.50	0.66	0.53
Average	0.78	0.73	0.81	0.72	0.81	0.73
Panel B: Evaluation period excluding covid pandemic (2006Q1 - 2019Q4)						
Q(-1)M1	0.95	0.95	0.95	0.82	0.94	0.82
Q(-1)M2	0.90	0.90	0.90	0.79	0.89	0.79
Q(-1)M3	0.75	0.75	0.75	0.66	0.73	0.66
Q(0)M1	0.69	0.70	0.69	0.61	0.68	0.61
Q(0)M2	0.62	0.63	0.62	0.55	0.60	0.55
Q(0)M3	0.51	0.53	0.51	0.43	0.49	0.43
Q(+1)M1	0.51	0.52	0.51	0.46	0.49	0.46
Average	0.71	0.71	0.71	0.62	0.69	0.62

Notes: This table displays the relative root mean squared forecast errors (RMFSEs) of nowcasting euro area GDP growth from 2006Q1 to 2022Q4 compared to a historical mean nowcast for different values of ϵ in the stopping rule in equation (9) with a maximum number of $J = 1,000$ (A)EM iterations. Panel A shows the results including the covid pandemic, while Panel B excludes this period. The small mixed-frequency dynamic factor models is estimated with $\kappa = 10^{-4}$ based on either the EM algorithm employed in Bańbura and Modugno (2014) or the AEM algorithm of Salakhutdinov and Roweis (2003). For each target quarter, the nowcasts/forecasts construction dates range from the first month of the previous quarter (that is, Q(-1)M1) up to the first month of the subsequent quarter (that is, Q(+1)M1).

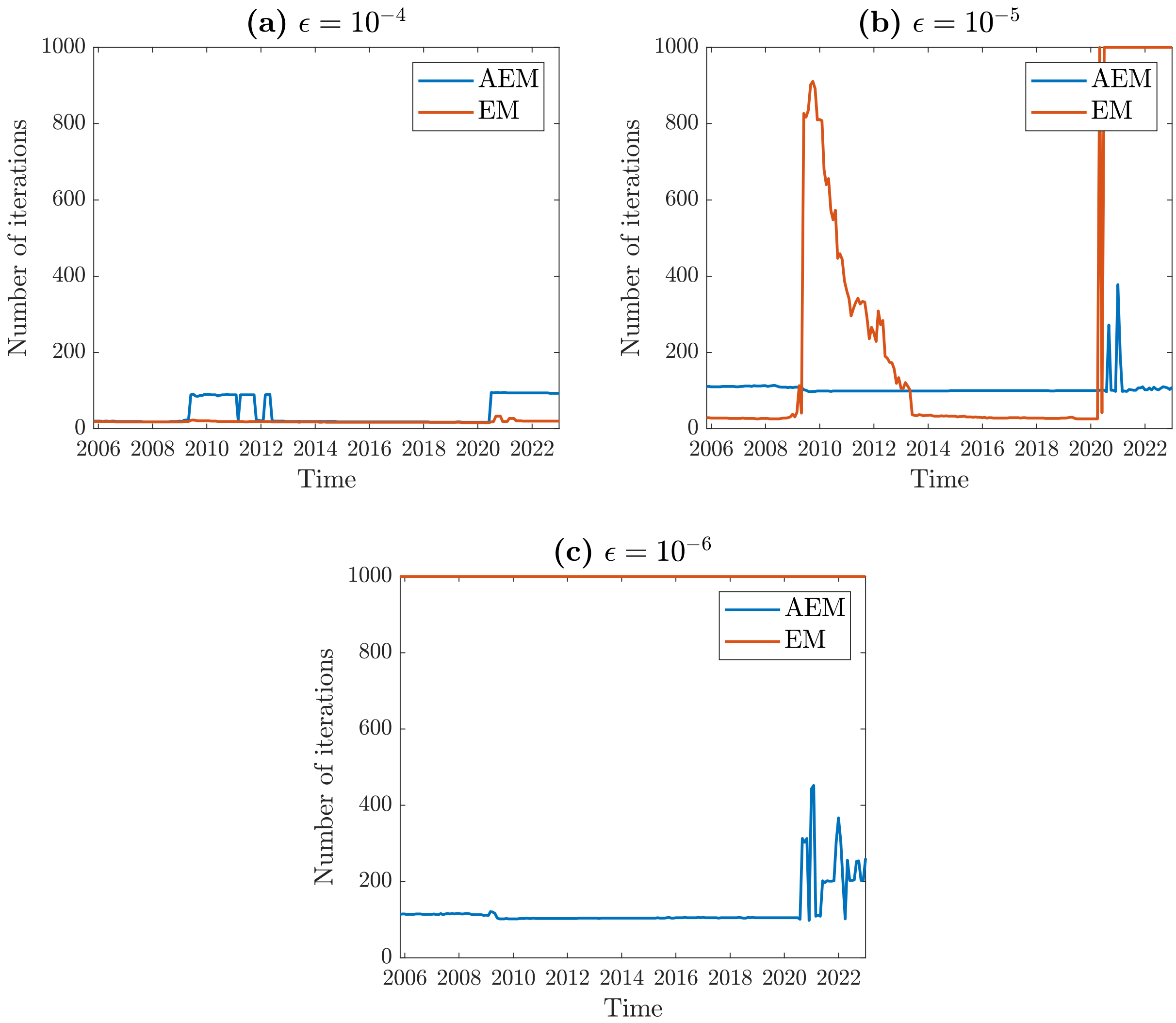

Figure D.1: Number of iterations until convergence for different values of ϵ in the stopping rule in the expanding-window estimation of the small mixed-frequency dynamic factor model estimated with $\kappa = 10^{-4}$

D.2 Cumulative sum of squared error difference plots

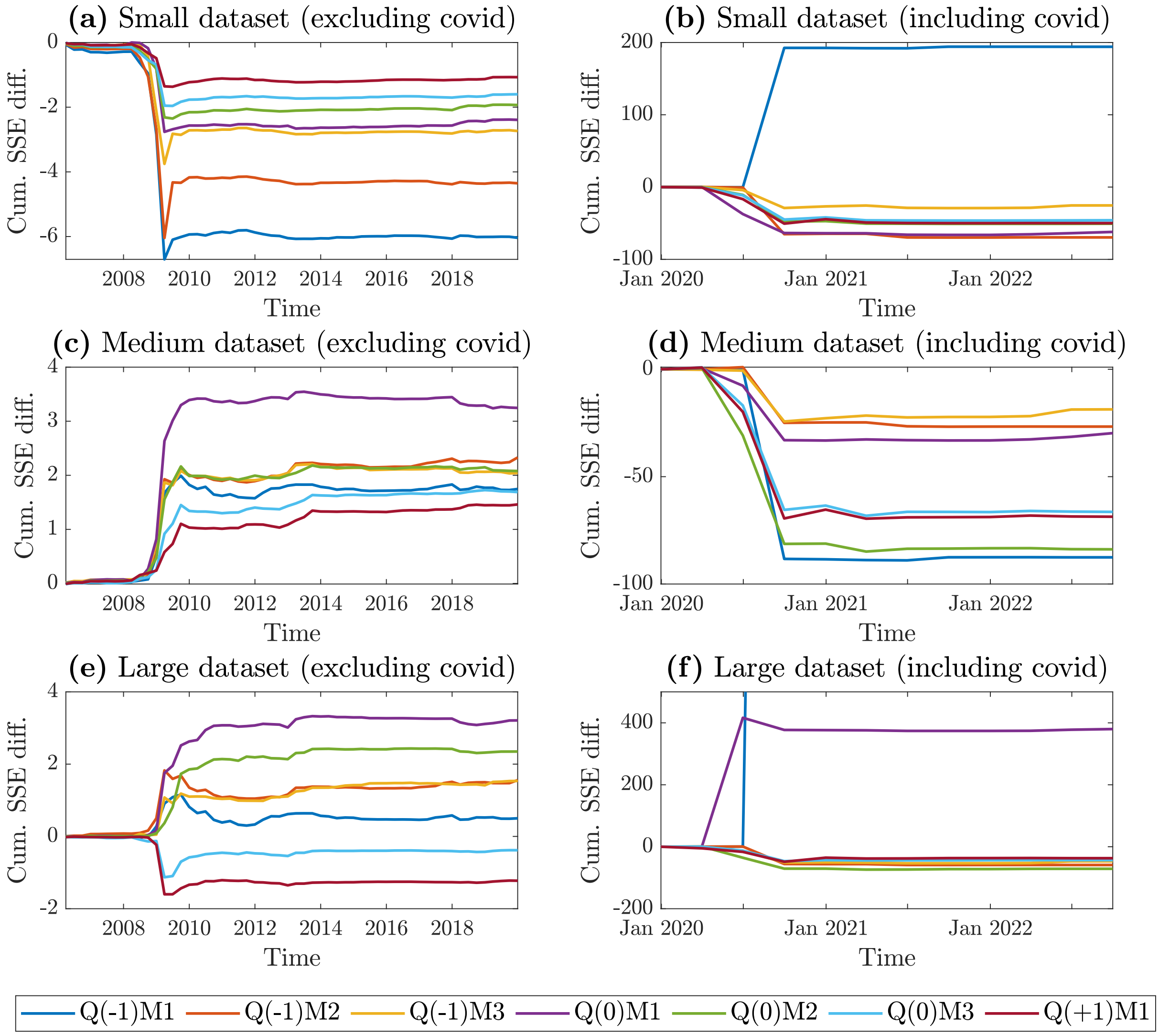

Figure D.2: Cumulative sum of squared forecast error (SSE) difference plots between the AEM and EM algorithm. A positive (negative) value indicates that EM produces more (less) accurate nowcasts than the AEM.

D.3 Nowcast results under fixed factor loadings

Table D.2: Relative nowcasting performance of euro area GDP growth based on the EM algorithm with either fixed or estimated factor loadings

	Small		**Medium**		**Large**	
	EM ($\boldsymbol{\Lambda}$ est.)	**EM ($\boldsymbol{\Lambda}$ fixed)**	**EM ($\boldsymbol{\Lambda}$ est.)**	**EM ($\boldsymbol{\Lambda}$ fixed)**	**EM ($\boldsymbol{\Lambda}$ est.)**	**EM ($\boldsymbol{\Lambda}$ fixed)**
Panel A: Evaluation period including covid pandemic (2006Q1 - 2022Q3)						
Q(-1)M1	1.17	1.14	1.93	1.93	1.64	1.64
Q(-1)M2	1.11	1.11	0.98	0.98	1.05	1.05
Q(-1)M3	0.89	0.90	0.81	0.81	0.88	0.88
Q(0)M1	0.64	0.67	0.57	0.57	0.62	0.63
Q(0)M2	0.57	0.60	0.66	0.67	0.69	0.70
Q(0)M3	0.61	0.63	0.65	0.66	0.66	0.67
Q(+1)M1	0.66	0.67	0.68	0.69	0.72	0.72
Average	0.81	0.82	0.90	0.90	0.89	0.90
Panel B: Evaluation period excluding covid pandemic (2006Q1 - 2019Q4)						
Q(-1)M1	0.94	0.95	0.84	0.85	0.85	0.86
Q(-1)M2	0.89	0.90	0.76	0.77	0.78	0.79
Q(-1)M3	0.73	0.75	0.72	0.72	0.72	0.73
Q(0)M1	0.68	0.69	0.74	0.74	0.73	0.73
Q(0)M2	0.60	0.62	0.68	0.68	0.67	0.67
Q(0)M3	0.49	0.51	0.61	0.61	0.58	0.58
Q(+1)M1	0.49	0.51	0.59	0.59	0.55	0.56
Average	0.69	0.70	0.71	0.71	0.70	0.70

Notes: This table displays the relative root mean squared forecast errors (RMFSEs) of nowcasting euro area GDP growth from 2006Q1 to 2022Q4 compared to a historical mean nowcast based on the EM algorithm with either fixed or estimates factor loadings $\boldsymbol{\Lambda}$. Panel A shows the results including the covid pandemic, while Panel B excludes this period. The small, medium and large mixed-frequency dynamic factor models are estimated with $\kappa = 10^{-4}$ based on the EM algorithm employed in Bańbura and Modugno (2014). We conduct $J = 1,000$, $J = 500$ and $J = 100$ (A)EM iterations for the small, medium and large models, respectively. For each target quarter, the nowcasts/forecasts construction dates range from the first month of the previous quarter (that is, Q(-1)M1) up to the first month of the subsequent quarter (that is, Q(+1)M1).

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.

Doz, C., D. Giannone, and L. Reichlin (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164, 188–205.

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.