MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR MODELS ON DATA SETS WITH ARBITRARY PATTERN OF MISSING DATA

by Marta Bańbura
and Michele Modugno

EUROPEAN CENTRAL BANK

EUROSYSTEM

MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR MODELS ON DATA SETS WITH ARBITRARY PATTERN OF MISSING DATA.[1]

by Marta Bańbura[2]
and Michele Modugno[3]

In 2010 all ECB
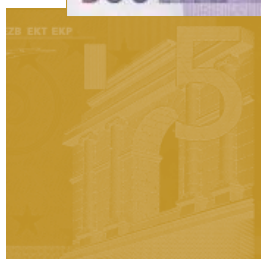publications
feature a motif
taken from the
€500 banknote.

CONTENTS

Abstract

In this paper we propose a methodology to estimate a dynamic factor model on data sets with an arbitrary pattern of missing data. We modify the Expectation Maximisation (EM) algorithm as proposed for a dynamic factor model by Watson and Engle (1983) to the case with general pattern of missing data. We also extend the model to the case with serially correlated idiosyncratic component. The framework allows to handle efficiently and in an automatic manner sets of indicators characterized by different publication delays, frequencies and sample lengths. This can be relevant e.g. for young economies for which many indicators are compiled only since recently. We also show how to extract a model based *news* from a statistical data release within our framework and we derive the relationship between the *news* and the resulting forecast revision. This can be used for interpretation in e.g. nowcasting applications as it allows to determine the sign and size of a *news* as well as its contribution to the revision, in particular in case of simultaneous data releases. We evaluate the methodology in a Monte Carlo experiment and we apply it to nowcasting and backdating of euro area GDP.

Keywords: Factor Models, Forecasting, Large Cross-Sections, Missing data, EM algorithm.

JEL classification: C53, E37.

Non-technical summary

In this paper we propose a methodology to estimate a dynamic factor model on data sets with an arbitrary pattern of missing data.

Dynamic factor models have found many applications in econometrics such as forecasting, structural analysis or construction of economic activity indicators. The underlying idea of such models is that the co-movement of (possibly many) observed series can be summarised by means of few unobserved factors. In this paper we adopt a version of dynamic factor model which implements factors as unobserved states. Hence Kalman filter and smoother apparatus can be used to estimate the unobserved factors and missing observations. Such dynamic factor models have been recently implemented e.g. at various central banks as short-term forecasting tools since they allow to exploit dynamic relationships when extracting information from incomplete cross-sections at the end of the sample, which arise due to publication delays and nonsynchronous data releases.

The estimation approach we propose here is based on Expectation-Maximisation (EM) algorithm. It is a general algorithm that offers a solution to problems for which incomplete or latent data yield the likelihood intractable or difficult to deal with. The essential idea is to write the likelihood as if the data were complete and to "fill in" the missing data in the expectation step. In case of the dynamic factor model considered here, the estimation problem is reduced to a sequence of simple steps, each of which essentially involves a pass of the Kalman smoother and two multivariate regressions. We show how to adapt the EM algorithm for factor model to the case with a general pattern of missing data. We also propose how to model the dynamics of idiosyncratic (series-specific) components.

Our approach allows to handle efficiently and in an automatic manner data sets with an arbitrary pattern of data availability. It is well suited for data sets including e.g. series of different sample lengths. Therefore, our framework can be particularly relevant for the euro area or other young economies for which many series have been compiled only since recently (e.g. euro area Purchasing Managers' Surveys). It could be also used to incorporate financial indicators with shorter history (e.g. share prices of particular institutions or series from euro area Bank Lending Survey). Moreover, as the series measured at a lower frequency can be interpreted as "high frequency" indicators with missing data, mixed frequency data sets can be easily handled. This can be important for two reasons: first, the information in the indicators sampled at a lower frequency (e.g. consumption, employment) can be used to extract the factors; second, the forecasts or interpolations of the former can be easily obtained.

We also discuss how to impose parameter restrictions in our framework, hence it can be used to estimate such models as e.g. Factor Augmented Vector Auto Regressions (FAVARs) or factor models with block structure. Flexibility with respect to data availability allows to apply the framework e.g. to estimate VARs or FAVARs on mixed frequency data or use these models in real-time forecasting applications.

Additional contribution of the paper is that we show how to extract a model based *news* from a statistical data release within our framework and we derive the relationship between the *news* and the resulting forecast revision. This can be of interest for understanding and interpreting forecast revisions in e.g. nowcasting applications in which there is a continuous inflow of new information and forecasts are frequently updated. It allows us to determine the sign and size of a *news* as well as its contribution to the revision, in particular in case of simultaneous data releases. For example, it enables us to produce statements like "the forecast was revised up by ... because of higher than expected release of ...".

We evaluate our methodology on both simulated and euro area data. In a Monte Carlo study we consider

different model specifications, sample sizes and fractions of missing data. We evaluate the precision in estimating the space spanned by the common factors as well as forecast accuracy. We compare these with alternative approaches based on EM algorithm. In the empirical application, we use the methodology for nowcasting and backdating of the euro area GDP using monthly and quarterly indicators. We consider specifications of different cross-sectional sizes, from small scale model with around 15 variables to large scale specification with around 100 series. Our approach can deal with such features of the data set as "ragged edge" caused by delayed and non-synchronous data releases, mixed frequency and varying series length. We compare the forecast accuracy of these specifications with that of univariate benchmarks as well as of another factor model implementation. We also illustrate how the *news* in the consecutive releases of different groups of variables revise the GDP forecast for the fourth quarter of 2008. Overall the results indicate that our methodology provides reliable results and is easy to implement and computationally inexpensive. In particular, it is feasible for large cross-sections.

1 Introduction

In this paper we propose a methodology to estimate a dynamic factor model on data sets with an arbitrary pattern of missing data.

Starting with seminal papers of Geweke (1977) and Sargent and Sims (1977), dynamic factor models have found many applications in econometrics such as forecasting, structural analysis or construction of economic activity indicators.[1] The underlying idea of a factor model is that (dynamic) co-movement of (possibly many) observed series can be summarised be few unobserved factors. Due to latency of the factors, maximum likelihood estimators cannot, in general, be obtained explicitly. Small scale dynamic factor models have been traditionally estimated by optimisation algorithms both in frequency (Geweke, 1977; Sargent and Sims, 1977; Geweke and Singleton, 1980) and in time domain (Engle and Watson, 1981; Stock and Watson, 1989; Quah and Sargent, 1992). For example, Engle and Watson (1981) write a dynamic factor model in a state space representation, apply Kalman filter to compute the likelihood and use an optimisation method to find maximum likelihood estimates of the parameters. An alternative approach has been proposed by Watson and Engle (1983), who have adapted the Expectation-Maximisation (EM) algorithm of Dempster, Laird, and Rubin (1977) to the case of dynamic factor model.[2]

We build on the dynamic factor model representation of Watson and Engle (1983) and, like this study, adopt the EM approach for maximum likelihood estimation. One contribution of the paper is to derive the steps of EM algorithm for a general pattern of missing data. While EM algorithm has been designed as a general approach to deal with latent and missing data, in the context of dynamic factor model, it has been usually applied to deal only with latency of the factors under the assumption that there are no missing values in the observables. The only exception is the paper by Shumway and Stoffer (1982), who show how to implement the EM algorithm for a state space representation with missing data, however only in the case in which the matrix linking the states and the observables is known. Here we deal with a general case. In addition, we propose how to model the serial correlation of the idiosyncratic component. Approaches proposed elsewhere (e.g. Reis and Watson, 2007; Jungbacker and Koopman, 2008) are not feasible in case of a general pattern of missing data.

With respect to a popular non-parametric method based on principal components,[3] maximum likelihood approach as adopted here has several advantages. First, it can deal with general pattern of missing data. Second, it provides framework for imposing restrictions on the parameters. Finally, it is more efficient for small samples.

Hence, the methodology proposed in this paper allows to handle efficiently and in an automatic manner data sets with an arbitrary pattern of data availability. It is well suited for data sets including e.g. series of different sample lengths. Therefore, our framework can be particularly relevant for the euro area or other young economies for which many series have been compiled only since recently (e.g. euro area Purchasing Managers' Surveys). It could be also used to incorporate financial indicators with shorter history (e.g. share prices of particular institutions or series from euro area Bank Lending Survey). Moreover, as the series measured at a lower frequency can be interpreted as "high frequency" indicators with missing data, mixed

---

[1] see e.g. Engle and Watson (1981); Watson and Engle (1983); Stock and Watson (1989); Quah and Sargent (1992); Bernanke and Boivin (2003); Forni, Hallin, Lippi, and Reichlin (2003, 2005); Giannone, Reichlin, and Sala (2004); Marcellino, Stock, and Watson (2003); Stock and Watson (1999, 2002a,b); Altissimo, Cristadoro, Forni, Lippi, and Veronese (2006);

[2] EM algorithm was originally proposed by Dempster, Laird, and Rubin (1977) as a general iterative solution for maximum likelihood estimation in problems with missing or latent data. It has been adapted to a variety of problems, such as e.g. mixture models, regime switching models, linear models with missing or truncated data, see e.g. McLachlan and Krishnan (1996) for an overview.

[3] see e.g. Connor and Korajczyk (1986, 1988, 1993); Forni and Reichlin (1996, 1998); Stock and Watson (2002a); Forni, Hallin, Lippi, and Reichlin (2000); Bai (2003); Giannone, Reichlin, and Small (2008);

frequency data sets can be easily handled. This can be important for two reasons: first, the information in the indicators sampled at a lower frequency (e.g. consumption, employment) can be used to extract the factors; second, the forecasts or interpolations of the former can be easily obtained.

Furthermore, since Factor Augmented VARs (FAVARs, see e.g. Bernanke, Boivin, and Eliasz, 2005) or factor models with a block structure (e.g. Kose, Otrok, and Whiteman, 2003) are restricted versions of a general model studied here, the methodology we propose can be used to estimate such models, in particular, in the presence of missing data (e.g. on mixed frequency or real-time data sets). We discuss how to impose such restrictions within our framework.[4]

Finally, the methodology is computationally feasible for large data sets. Maximum likelihood approach, in general, has been long considered infeasible for data sets in which the size of cross-section is large. Therefore, non-parametric methods based on principal components have been applied. Recently, Doz, Giannone, and Reichlin (2006) have proved that, as the size of the cross-section goes to infinity, one can obtain consistent estimates of the factors by maximum likelihood (also in case of weak cross and serial correlation in the idiosyncratic component). In a Monte Carlo study they have used EM algorithm for the estimation and shown that it is reliable and computationally inexpensive also in the case of large cross-sections.[5]

Additional contribution of the paper is that we show how to extract a model based *news* from a statistical data release within our framework and we derive the relationship between the *news* and the resulting forecast revision.[6] The derivations can be easily adopted to any model that can be cast in a state space form. This can be of interest for understanding and interpreting forecast revisions in e.g. nowcasting applications in which there is a continuous inflow of new information and forecasts are frequently updated. It allows us to determine the sign and size of a *news* as well as its contribution to the revision, in particular in case of simultaneous data releases. For example, it enables us to produce statements like "the forecast was revised up by ... because of higher than expected release of ...".

We evaluate the performance of the methodology both on simulated and on euro area data.

In a Monte Carlo simulation experiment we consider different model specifications, sample sizes and fractions of missing data. We evaluate the precision in estimating the space spanned by the common factors as well as forecast accuracy. We compare these with the results obtained when using the EM algorithms proposed by Stock and Watson (2002b) and by Rubin and Thayer (1982) (the latter is a special case of the algorithm derived in this paper).

In the empirical application, we use the methodology for real-time forecasting and backdating of the euro area GDP using monthly and quarterly indicators. We consider specifications of different cross-sectional sizes, from small scale model with around 15 variables to large scale specification with around 100 series. Our approach can deal with such features of the data set as "ragged edge",[7] mixed frequency and varying series length (e.g. Purchasing Managers' Surveys are available only later in the sample). We compare

---

[4]EM algorithm has been recently applied to estimate models in the spirit of FAVAR by Bork, Dewachter, and Houssa (2009) and Bork (2009). Applications to other types of restricted factor models include Reis and Watson (2007) and Modugno and Nikolaou (2009); the former impose restrictions in order to identify the pure inflation, the latter in order to forecast the yield curve using the Nelson-Siegel exponential components framework.

[5]Jungbacker and Koopman (2008) show that a simple transformation of the state space representation can yield substantial computational gains for likelihood evaluation. They show that, on one hand, this can be used to speed up the EM iterations and, on the other hand, direct maximisation of the likelihood by optimisation methods becomes feasible also for large cross-sections.

[6]Note that the *news* concept considered here is defined with respect to the model and not market expectations. It is also different from news vs. noise concept considered by Giannone, Reichlin, and Small (2008).

[7]"Ragged edge" arises in real-time applications and means that there is a varying number of missing observations at the end of the sample as different series are subject to different publication delays and are released at different points in time.

the forecast accuracy of these specifications with that of univariate benchmarks as well as of the model of Bańbura and Rünstler (2010), who adopt the methodology of Giannone, Reichlin, and Small (2008) to the case of euro area.

Giannone, Reichlin, and Small (2008) have proposed a factor model framework, which allows to deal with "ragged edge" and exploit information from large data sets in a timely manner. They have applied it to nowcasting of US GDP from a large number of monthly indicators. While Giannone, Reichlin, and Small (2008) can handle the "ragged edge" problem, it is not straightforward to apply their methodology to mixed frequency panels with series of different lengths or, in general, to any pattern of missing data.[8] In addition, as the estimation is based on principal components, it could be inefficient for small samples.

Other papers related to ours include Camacho and Perez-Quiros (2008) who obtain real-time estimates of the euro area GDP from monthly indicators from a small scale model applying the mixed frequency factor model approach of Mariano and Murasawa (2003). Schumacher and Breitung (2008) forecast German GDP from large number of monthly indicators using the EM approach proposed by Stock and Watson (2002b). Proietti (2008) estimates a factor model for interpolation of GDP and its main components and shows how to incorporate relevant accounting and temporary constraints. Angelini, Henry, and Marcellino (2006) propose methodology for backdating and interpolation based on large cross-sections. In contrast to theirs, our method exploits the dynamics of the data and is based on maximum likelihood which allows for imposing restrictions and is more efficient for smaller cross-sections.

The paper is organized as follows. Section 2 presents the model, discusses the estimation and explains how the *news* content can be extracted. Section 3 provides the results of the Monte Carlo experiment. Section 4 describes the empirical application. Section 5 concludes. The technical details and data description are provided in the Appendix.

2 Econometric framework

Let $y_t = [y_{1,t}, y_{2,t}, \ldots, y_{n,t}]'$, $t = 1, \ldots, T$ denote a stationary n-dimensional vector process standardised to mean 0 and unit variance. We assume that y_t admits the following factor model representation:

$$y_t = \Lambda f_t + \epsilon_t \,, \tag{1}$$

where f_t is a $r \times 1$[9] vector of (unobserved) common factors and $\epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{n,t}]'$ is the idiosyncratic component, uncorrelated with f_t at all leads and lags. The $n \times r$ matrix Λ contains factor loadings. $\chi_t = \Lambda f_t$ is referred to as the common component. It is assumed that ϵ_t is normally distributed and cross-sectionally uncorrelated, i.e. y_t follows an exact factor model. We also shortly discuss validity of the approach in the case of an approximate factor model, see below. What concerns the dynamics of the idiosyncratic component we consider two cases: ϵ_t is serially uncorrelated or it follows an AR(1) process.

Further, it is assumed that the common factors f_t follow a stationary VAR process of order p:

$$f_t = A_1 f_{t-1} + A_2 f_{t-2} + \cdots + A_p f_{t-p} + u_t \,, \qquad u_t \sim \text{i.i.d. } \mathcal{N}(0, Q) \,, \tag{2}$$

where $A_1, \ldots, A_p$ are $r \times r$ matrices of autoregressive coefficients. We collect the latter into $A = [A_1, \ldots, A_p]$.

---

[8] Their estimation approach consists of two steps. First, the parameters of the state space representation of the factor model are obtained using a principal components based procedure applied to a truncated data set (without missing data). Second, Kalman filter is applied on the full data set in order to obtain factor estimates and forecasts using all available information.

[9] For identification it is required that $2r + 1 \leq n$, see e.g. Geweke and Singleton (1980).

2.1 Estimation

As f_t are unobserved, the maximum likelihood estimators of the parameters of model (1)-(2), which we collect in θ, are in general not available in closed form. On the other hand, a direct numerical maximisation of the likelihood is computationally demanding, in particular for large n due to the large number of parameters.[10]

In this paper we adopt an approach based on the Expectation-Maximisation (EM) algorithm, which was proposed by Dempster, Laird, and Rubin (1977) as a general solution to problems for which incomplete or latent data yield the likelihood intractable or difficult to deal with. The essential idea of the algorithm is to write the likelihood as if the data were complete and to iterate between two steps: in the Expectation step we "fill in" the missing data in the likelihood, while in the Maximisation step we re-optimise this expectation. Under some regularity conditions, the EM algorithm converges towards a local maximum of the likelihood (or a point in its ridge, see also below).

To derive the EM steps for the model described above, let us denote the joint log-likelihood of y_t and f_t, $t = 1, \ldots, T$ by $l(Y, F; \theta)$, where $Y = [y_1, \ldots, y_T]$ and $F = [f_1, \ldots, f_T]$. Given the available data $\Omega_T \subseteq Y$,[11] EM algorithm proceeds in a sequence of two alternating steps:

1. E-step - the expectation of the log-likelihood conditional on the data is calculated using the estimates from the previous iteration, $\theta(j)$:

$$L(\theta, \theta(j)) = \mathbb{E}_{\theta(j)}\big[l(Y, F; \theta)|\Omega_T\big];$$

2. M-step - the parameters are re-estimated through the maximisation of the expected log-likelihood with respect to θ:

$$\theta(j+1) = \arg\max_\theta L(\theta, \theta(j)). \tag{3}$$

Watson and Engle (1983) and Shumway and Stoffer (1982) show how to derive the maximisation step (3) for models similar to the one given by (1)-(2). As a result the estimation problem is reduced to a sequence of simple steps, each of which essentially involves a pass of the Kalman smoother and two multivariate regressions. Doz, Giannone, and Reichlin (2006) show that the EM algorithm is a valid approach for the maximum likelihood estimation of factor models for large cross-sections as it is robust, easy to implement and computationally inexpensive. Watson and Engle (1983) assume that all the observations in y_t are available ($\Omega_T = Y$). Shumway and Stoffer (1982) derive the modifications for the missing data case but only with known Λ. We provide the EM steps for the general case with missing data.

In the main text, we set for simplicity $p = 1$ ($A = A_1$), the case of $p > 1$ is discussed in the Appendix. We first consider the case of serially uncorrelated ϵ_t:

$$\epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, R), \tag{4}$$

where R is a diagonal matrix. In that case $\theta = \{\Lambda, A, R, Q\}$ and the maximisation of (3) results in the

---

[10]Recently, Jungbacker and Koopman (2008) have shown how to reduce the computational complexity related to estimation and smoothing if the number of observables is much larger than the number of factors.

[11]$\Omega_T \subseteq Y$ because some observations in y_t can be missing.

following expressions for $\Lambda(j+1)$ and $A(j+1)$:[12]

$$\Lambda(j+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[y_t f_t'|\Omega_T\right]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_t f_t'|\Omega_T\right]\right)^{-1}, \tag{5}$$

$$A(j+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_t f_{t-1}'|\Omega_T\right]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_{t-1} f_{t-1}'|\Omega_T\right]\right)^{-1}. \tag{6}$$

Note that these expressions resemble the ordinary least squares solution to the maximum likelihood estimation for (auto-) regressions with complete data with the difference that the sufficient statistics are replaced by their expectations.

The $(j+1)$-iteration covariance matrices are computed as the expectations of sums of squared residuals conditional on the updated estimates of Λ and A:[13]

$$R(j+1) = \text{diag}\left(\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[\left(y_t - \Lambda(j+1)f_t\right)\left(y_t - \Lambda(j+1)f_t\right)'|\Omega_T\right]\right) \tag{7}$$

$$= \text{diag}\left(\frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[y_t y_t'|\Omega_T\right] - \Lambda(j+1)\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_t y_t'|\Omega_T\right]\right)\right)$$

and

$$Q(j+1) = \frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_t f_t'|\Omega_T\right] - A(j+1)\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_{t-1} f_t'|\Omega_T\right]\right). \tag{8}$$

When y_t does not contain missing data, we have that

$$\mathbb{E}_{\theta(j)}\left[y_t y_t'|\Omega_T\right] = y_t y_t' \quad \text{and} \quad \mathbb{E}_{\theta(j)}\left[y_t f_t'|\Omega_T\right] = y_t \mathbb{E}_{\theta(j)}\left[f_t'|\Omega_T\right]. \tag{9}$$

Finally, the conditional moments of the latent factors, $\mathbb{E}_{\theta(j)}\left[f_t|\Omega_T\right]$, $\mathbb{E}_{\theta(j)}\left[f_t f_t'|\Omega_T\right]$, $\mathbb{E}_{\theta(j)}\left[f_{t-1} f_{t-1}'|\Omega_T\right]$ and $\mathbb{E}_{\theta(j)}\left[f_t f_{t-1}'|\Omega_T\right]$, can be obtained through the Kalman smoother for the state space representation:

$$\begin{aligned}
y_t &= \Lambda(j)f_t + \epsilon_t, & \epsilon_t &\sim \text{i.i.d. } \mathcal{N}(0, R(j)), \\
f_t &= A(j)f_{t-1} + u_t, & u_t &\sim \text{i.i.d. } \mathcal{N}(0, Q(j)),
\end{aligned} \tag{10}$$

see Watson and Engle (1983).

However, when y_t contains missing values we can no longer use (9) when developing the expressions (5) and (7). Let W_t be a diagonal matrix of size n with i^{th} diagonal element equal to 0 if $y_{i,t}$ is missing and equal to 1 otherwise. As shown in the Appendix, $\Lambda(j+1)$ can be obtained as

$$\text{vec}\big(\Lambda(j+1)\big) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\left[f_t f_t'|\Omega_T\right] \otimes W_t\right)^{-1} \text{vec}\left(\sum_{t=1}^{T} W_t y_t \mathbb{E}_{\theta(j)}\left[f_t'|\Omega_T\right]\right). \tag{11}$$

Intuitively, W_t works as a selection matrix, so that only the available data are used in the calculations. Analogously, the expression (7) becomes

$$\begin{aligned}
R(j+1) = \text{diag}\Bigg(\frac{1}{T}\sum_{t=1}^{T}\Big(&W_t y_t y_t' W_t' - W_t y_t \mathbb{E}_{\theta(j)}\left[f_t'|\Omega_T\right]\Lambda(j+1)'W_t - W_t\Lambda(j+1)\mathbb{E}_{\theta(j)}\left[f_t|\Omega_T\right]y_t'W_t \\
&+ W_t\Lambda(j+1)\mathbb{E}_{\theta(j)}\left[f_t f_t'|\Omega_T\right]\Lambda(j+1)'W_t + (I-W_t)R(j)(I-W_t)\Big)\Bigg).
\end{aligned} \tag{12}$$

Again, only the available data update the estimate. $I - W_t$ in the last term "selects" the entries of $R(j)$ corresponding to the missing observations. For example, when for some t all the observations in y_t are missing, the period t contribution to $R(j+1)$ would be $R(j)/T$.

When applying the Kalman filter on the state space representation (10), in case some of the observations in y_t are missing, the corresponding rows in y_t and $\Lambda(j)$ (and the corresponding rows and columns in $R(j)$) are skipped (cf. Durbin and Koopman, 2001).

It is easy to see that with $W_t \equiv I$, (11) and (12) coincide with the "complete data" expressions obtained by plugging (9) into (5) and (7).

Static factor model

Note that the static factor model is a special case of the representation considered above in which $A = 0$. EM algorithm for a static factor model (without missing data) has been derived by Rubin and Thayer (1982). In the Appendix we show that the EM steps of Rubin and Thayer (1982) can be derived from the general expressions for $\Lambda(j+1)$ and $R(j+1)$ as given by formulas (5) and (7), where the conditional expectations can be derived explicitly. We also discuss the modification of the expressions of Rubin and Thayer (1982) to the missing data case.

Note that this approach is different from the EM based method proposed by Stock and Watson (2002b) to compute the principal components from data sets with missing observations. In the latter case, the objective function is proportional to the expected log-likelihood under the assumption of fixed factors and homoscedastic idiosyncratic component.

The performance of these different approaches for different model specifications and different fractions of missing data is compared in the Monte Carlo study in Section 3.

Approximate factor model

As argued in e.g. Stock and Watson (2002a) or Doz, Giannone, and Reichlin (2006) the assumption of no cross-correlation in the idiosyncratic component could be too restrictive, in particular in the case of large cross-sections. Following Chamberlain and Rothschild (1983), factor models with weakly cross-correlated idiosyncratic component are often referred to as approximate.

Doz, Giannone, and Reichlin (2007) show that, under the approximate factor model (with possibly serially correlated idiosyncratic errors), as $n, T \to \infty$ the factors can be consistently estimated by quasi maximum likelihood, where the miss-specified model is the exact factor model (with uncorrelated idiosyncratic error) described above (see Doz, Giannone, and Reichlin, 2006, for the technical details). Consequently, the estimators considered above are asymptotically valid also in the case of the approximate factor model.[14]

In the Monte Carlo simulations in Section 3 we study the performance of different methods also in the presence of serial and cross-correlations of the idiosyncratic component.

Restrictions on the parameters

One of the advantages of the maximum likelihood approach proposed here, with respect to non-parametric methods based on principal components, is that it allows imposing restrictions on the parameters in a relatively straightforward manner.

---

[14]Stock and Watson (2002a) prove similar result for factor estimators based on principal components.

Bork (2009) and Bork, Dewachter, and Houssa (2009) show how to modify the M-step of Watson and Engle (1983) in order to impose restrictions of the form $H_\Lambda \text{vec}(\Lambda) = \kappa_\Lambda$ for the model given by (1)-(2). Straightforward adaptation of their expressions to the missing data case results in the restricted estimate given by

$$
\begin{aligned}
\text{vec}\big(\Lambda_r(j+1)\big) &= \text{vec}\big(\Lambda_u(j+1)\big) + \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t' | \Omega_T \big] \otimes R(j) \right) H_\Lambda' \times \\
&\times \left(H_\Lambda \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t' | \Omega_T \big] \otimes R(j) \right) H_\Lambda' \right)^{-1} \left(\kappa_\Lambda - H_\Lambda \text{vec}\big(\Lambda_u(j+1)\big) \right),
\end{aligned}
\tag{13}
$$

where $\Lambda_u(j+1)$ is the unrestricted estimate given by expression (11). Restrictions on the parameters in the transition equation: $H_A \text{vec}(A) = \kappa_A$, can be imposed in an analogous manner, see Bork (2009).[15]

This type of restrictions are relevant for a number of models, such as e.g.:

- Factor Augmented VAR (FAVAR) models as proposed by Bernanke, Boivin, and Eliasz (2005). Bork (2009) has recently shown how to estimate this type of model by EM algorithm.

- Mixed frequency models - for example, the approach of Mariano and Murasawa (2003) to joint modelling of monthly and quarterly variables requires imposing restrictions on the factor loadings of the latter. We impose this type of restriction in the empirical application in Section 4. Giannone, Reichlin, and Simonelli (2009) apply the EM approach to estimate a mixed frequency VAR.

- Factor models with a block structure - there are several applications, in which (some) factors are specific to a subset of variables considered. For example, Kose, Otrok, and Whiteman (2003) consider global and region-specific factors. Belviso and Milani (2006) extract factors from blocks of variables representing a single concept (e.g. real activity, inflation, money). While these two papers adopt Bayesian approach, Bańbura, Giannone, and Reichlin (2010b) apply the methodology proposed in this paper to extract real and nominal factors. This type of models implies zero restriction on some factor loadings and/or autoregressive parameters in the factor VAR, which can be imposed either by using the formula (13) or by estimating each block of Λ or A separately (see Bańbura, Giannone, and Reichlin, 2010b).

The methodology presented here can be applied to estimate these types of models in the presence of missing data. It could be e.g. used to estimate mixed-frequency VARs or FAVARs or to apply these models to forecasting in real-time.

Identification

The likelihood of the model given by (1)-(2) and (4) is invariant to any invertible linear transformation of the factors. In other words, for any invertible matrix M, the parameters $\theta = \{\Lambda, A, R, Q\}$ and $\theta_M = \{\Lambda M^{-1}, MAM^{-1}, R, MQM'\}$ are observationally equivalent and hence θ is not identifiable from the data. As argued in Dempster, Laird, and Rubin (1977), in this case EM algorithm will converge to a particular θ_M in the ridge of the likelihood function (and not move indefinitely between different points in the ridge). Therefore, for forecasting applications, this lack of identifiability is not an issue, as one is interested in the space spanned by the factors and not in the factors themselves.

---

[15]Shumway and Stoffer (1982) show how to impose restrictions on A of the form $AF = G$. This type of restrictions is, however, less general and e.g. does not allow to restrict only selected equations.

In order to achieve identifiability of θ, one needs to choose a particular normalisation or, in other words, restrict the parameter space. For example, Proietti (2008) or Jungbacker and Koopman (2008) restrict Λ as:[16]

$$\Lambda = \left[\begin{array}{c} I_r \\ \Lambda^* \end{array} \right],$$

where Λ^* is $(n-r) \times r$ unrestricted matrix. In order to impose such restriction one could either use formula (13) or modify the updating formula (11) as:

$$\Lambda(j+1) = \left[\begin{array}{c} I_r \\ \Lambda^*(j+1) \end{array} \right],$$

$$\mathrm{vec}\big(\Lambda^*(j+1)\big) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t' | \Omega_T\big] \otimes W_t^* \right)^{-1} \mathrm{vec}\left(\sum_{t=1}^{T} W_t^* y_t^* \mathbb{E}_{\theta(j)}\big[f_t' | \Omega_T\big] \right),$$

where $y_t^* = [y_{r+1,t}, \ldots, y_{n,t}]'$ and W_t^* is obtained from W_t by removing the first r rows and columns.[17]

Modelling the serial correlation in the idiosyncratic component

The EM steps discussed above were derived under the assumption of no serial correlation in the idiosyncratic component. As mentioned, such estimates are asymptotically valid even when this assumption is violated. However in certain applications, like e.g. forecasting, it could be advantageous to model the idiosyncratic dynamics, cf. e.g. Stock and Watson (2002b). Such strategy might improve the forecasts for two reasons: first, we could forecast the idiosyncratic component; second, we could improve the efficiency of the common factor estimates in small samples or in real-time applications in which the cross-sections at the end of the sample are incomplete.

There are different approaches to modelling of the idiosyncratic serial correlation. For example, Reis and Watson (2007) include lags of the observables into the measurement equation and alternate between two steps - they estimate the coefficients on the lags conditional on the remaining parameters and vice versa. Jungbacker and Koopman (2008) propose to use the Kalman smoother to estimate the (auto-) regression parameters as additional states in an augmented state space form. Those approaches are however not appropriate in the case of arbitrary missing data pattern. Instead, we propose to represent the idiosyncratic component by an AR(1) process and to add it to the state vector.

More precisely, we assume that $\epsilon_{i,t}$, $i = 1, \ldots, n$ in (1) can be decomposed as:

$$\begin{aligned} \epsilon_{i,t} &= \tilde{\epsilon}_{i,t} + \xi_{i,t}, & \xi_{i,t} &\sim \text{ i.i.d. } \mathcal{N}(0, \kappa), \\ \tilde{\epsilon}_{i,t} &= \alpha_i \tilde{\epsilon}_{i,t-1} + e_{i,t}, & e_{i,t} &\sim \text{ i.i.d. } \mathcal{N}(0, \sigma_i^2), \end{aligned} \tag{14}$$

where both $\xi_t = [\xi_{1,t}, \ldots, \xi_{n,t}]'$ and $\tilde{\epsilon}_t = [\tilde{\epsilon}_{1,t}, \ldots, \tilde{\epsilon}_{n,t}]'$ are cross-sectionally uncorrelated and κ is a very small number.[18] Combining (1), (2) and (14) results in the new state space representation:

$$\begin{aligned} y_t &= \tilde{\Lambda} \tilde{f}_t + \xi_t, & \xi_t &\sim \mathcal{N}(0, \tilde{R}), \\ \tilde{f}_t &= \tilde{A} \tilde{f}_{t-1} + \tilde{u}_t, & \tilde{u}_t &\sim \mathcal{N}(0, \tilde{Q}), \end{aligned} \tag{15}$$

where

$$\tilde{f}_t = \left[\begin{array}{c} f_t \\ \tilde{\epsilon}_t \end{array} \right], \; \tilde{u}_t = \left[\begin{array}{c} u_t \\ e_t \end{array} \right], \; \tilde{\Lambda} = [\; \Lambda \;\; I \;], \; \tilde{A} = \left[\begin{array}{cc} A & 0 \\ 0 & \mathrm{diag}(\alpha_1, \cdots, \alpha_n) \end{array} \right], \; \tilde{Q} = \left[\begin{array}{cc} Q & 0 \\ 0 & \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2) \end{array} \right],$$

---

[16]This restriction is based on the theoretical results in Geweke and Singleton (1980), who also propose an alternative normalisation, see also Camba-Mendez, Kapetanios, Smith, and Weale (2001). As shown in Heaton and Solo (2004), under certain assumptions these restrictions could be partly redundant, but this issue is beyond the scope of this paper.

[17]In order to avoid the problem of weak identification, in practise the first r series should be selected so as to have a relatively large and sufficiently different common components.

[18]This allows us to write the likelihood analogously to the exact factor model case, see the Appendix.

$e_t = [e_{1,t}, \ldots, e_{n,t}]'$ and $\tilde{R}$ is a fixed diagonal matrix with κ on the diagonal.

It follows, that the expressions for $A(j+1)$ and $Q(j+1)$ remain as above while the one for $\Lambda(j+1)$ needs to be modified as follows:

$$\mathrm{vec}\big(\Lambda(j+1)\big) = \left(\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[f_t f_t'|\Omega_T\big] \otimes W_t\right)^{-1} \mathrm{vec}\left(\sum_{t=1}^{T} W_t y_t \mathbb{E}_{\tilde{\theta}(j)}\big[f_t'|\Omega_T\big] + W_t \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_t f_t'|\Omega_T\big]\right),$$

with $\tilde{\theta} = \{\tilde{\Lambda}, \tilde{A}, \tilde{Q}\}$, see the Appendix for the derivations. Furthermore, the $(j+1)$-iteration estimates of the autoregressive parameters of the idiosyncratic component are given by:

$$
\begin{aligned}
\alpha_i(j+1) &= \left(\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_{i,t}\tilde{\epsilon}_{i,t-1}|\Omega_T\big]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_{i,t-1}^2|\Omega_T\big]\right)^{-1}, \\
\sigma_i^2(j+1) &= \frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_{i,t}^2|\Omega_T\big] - \alpha_i(j+1)\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_{i,t-1}\tilde{\epsilon}_{i,t}|\Omega_T\big]\right).
\end{aligned}
$$

The conditional moments involving $\tilde{\epsilon}_t$ can be obtained from the Kalman smoother on the augmented state space given by (15).

Note that augmenting the state vector by the idiosyncratic component increases the dimension of the former. This slows down the Kalman filter but has not caused any computational problems in our applications. Jungbacker, Koopman, and van der Wel (2009) show how to speed up the Kalman filter recursions by alternating between the representation of Reis and Watson (2007) and the one given by (15) depending on the availability of the data in y_t. Depending on the fraction of missing data, this can lead to substantial computational gains, however comes at the cost of more complex, time-varying state space representation.

Initial parameter values and stopping rule

In order to obtain initial values for the parameters, $\theta(0)$, we replace the missing observations in y_t by draws from $\mathcal{N}(0,1)$ distribution and we apply the methodology of Giannone, Reichlin, and Small (2008). First, we estimate Λ and F by applying principal components analysis to the covariance matrix of Y. Second, we obtain A and Q by estimating VAR on $\hat{F}$, obtained in the previous step. Depending on the version of the model, we estimate R or α_i and σ_i^2 from the residuals $\hat{\epsilon}_t = y_t - \hat{\Lambda}\hat{f}_t$, (see also the discussion in Doz, Giannone, and Reichlin, 2006, Section 4).

Concerning the stopping rule, we follow Doz, Giannone, and Reichlin (2006) and stop the iterations when the increase in likelihood between two consecutive steps is small. More precisely, let $l(\Omega_T; \theta)$ denote the log-likelihood of the data conditional on parameter θ (which can be obtained from Kalman filter) and $c_j = \frac{l(\Omega_T;\theta(j)) - l(\Omega_T;\theta(j-1))}{(|l(\Omega_T;\theta(j))| + |l(\Omega_T;\theta(j-1))|)/2}$. We stop after iteration J when c_J is below the threshold of 10^{-4}.

2.2 Forecasting, backdating and interpolation

Given the estimates of the parameters $\hat{\theta}$ and the data set Ω_T we can obtain the conditional expectations for the missing observations from:

$$\mathbb{E}_{\hat{\theta}}[y_{i,t}|\Omega_T] = \hat{\Lambda}_{i\cdot}\mathbb{E}_{\hat{\theta}}[f_t|\Omega_T] + \mathbb{E}_{\hat{\theta}}[\epsilon_{i,t}|\Omega_T], \qquad y_{i,t} \notin \Omega_T,$$

where $\hat{\Lambda}_{i\cdot}$ denotes the ith row of $\hat{\Lambda}$. $\mathbb{E}_{\hat{\theta}}[f_t|\Omega_T]$ and $\mathbb{E}_{\hat{\theta}}[\epsilon_{i,t}|\Omega_T]$ are obtained by applying Kalman filter and smoother to the state representation (10) or (15). In the former case, $\mathbb{E}_{\hat{\theta}}[\epsilon_{i,t}|\Omega_T] = 0$.

Depending on the purpose of the application (and the pattern of missing data), these conditional expectation can be used to obtain e.g.:

- *Forecasts*

  They are readily available from the Kalman filter. One of the appeals of the framework in the real-time context is that it allows to exploit the dynamic relationships when extracting the information from incomplete cross-sections at the end of the sample. This is one of the advantages over static methods, which sometimes have to discard data at the end of the sample, when the fraction of missing data is too large to reliably extract the factors based only on static correlations (cf. Section 3 or Doz, Giannone, and Reichlin, 2007). In addition, explicit modelling of dynamics within the model and the fact that it can be cast in a state space representation allow to extract model based news from statistical data releases and to link it to resulting forecast revision, see the next section.

- *Back data*

  If, for example, series i is available only as of period $t_i > 1$, Kalman smoother can be used to obtain the back data for this series: $\mathbb{E}_{\hat{\theta}}(y_{i,t}|\Omega_T)$, $t < t_i$, conditional on the information in other series and estimated correlations, see an example in Section 4.5.

- *Interpolations*

  A low-frequency series can be considered as a partially observed high-frequency variable. For example, in the empirical application, we treat quarterly variables as monthly series observed only in the third month of each quarter, i.e. with missing data in the first and second month of each quarter. Kalman smoother can be applied to obtain expectations for the "missing" months conditional on the information in the monthly series and taking into account the estimated dynamic relationships. Therefore, the methodology can be a valid alternative to standard interpolation techniques such as e.g. Chow and Lin (1971), (see also Angelini, Henry, and Marcellino, 2006; Proietti, 2008, for recent methodologies based on large data sets).

2.3 News in data releases and forecast revisions

When forecasting in real-time, one faces a continues inflow of information as new figures for various predictors are released non-synchronously and with different degree of delay. Therefore, in such applications, we seldom perform a single prediction for the reference period but rather a sequence of forecasts, which are updated when new data arrive. Intuitively, only the *news* or the "unexpected" component from released data should revise the forecast, hence, extracting the *news* and linking it to the resulting forecast revision is key for understanding and interpreting the latter. This section introduces the concept of model based *news* in data releases, shows how to extract it for the model described above and finally derives the relationship between the *news* and the forecast revision.

We denote by Ω_v a vintage of data corresponding to a particular statistical release date v.[19] Let us consider two consecutive data vintages, Ω_v and Ω_{v+1}. The information sets Ω_v and Ω_{v+1} can differ for two reasons: first, Ω_{v+1} contains some newly released figures, $\{y_{i_j,t_j}, \ j = 1, \ldots, J_{v+1}\}$, which were not available in Ω_v; second, some of the data might have been revised. However, in what follows we abstract from data revisions and therefore we have:

$$\Omega_v \subset \Omega_{v+1} \qquad \text{and} \qquad \Omega_{v+1}\backslash\Omega_v = \{y_{i_j,t_j}, j = 1, \ldots, J_{v+1}\},$$

hence the information set is "expanding". Note that since different types of data are characterised by different publication delays, in general we will have $t_j \neq t_l$ for some $j \neq l$.

---

[19]We do not index the data vintages by t since statistical data releases usually occur at a higher frequency due to their non-synchronicity. For example, we will have several releases of monthly data within a month, corresponding to different groups of indicators, such as e.g. industrial production (released around mid-month) or surveys (released shortly before the end of month).

Let us now look at the two consecutive forecast updates, $\mathbb{E}\left[y_{k,t_k}|\Omega_v\right]$ and $\mathbb{E}\left[y_{k,t_k}|\Omega_{v+1}\right]$, for a variable of interest, k, in period t_k. In this section we abstract from the problem of parameter uncertainty and to simplify the notation we drop the subscript θ. The new figures, $\{y_{i_j,t_j}, \; j = 1, \ldots, J_{v+1}\}$, will in general contain some new information on y_{k,t_k} and consequently lead to a revision of its forecast. From the properties of conditional expectation as an orthogonal projection operator, it follows that:

$$\underbrace{\mathbb{E}\left[y_{k,t_k}|\Omega_{v+1}\right]}_{\text{new forecast}} = \underbrace{\mathbb{E}\left[y_{k,t_k}|\Omega_v\right]}_{\text{old forecast}} + \underbrace{\mathbb{E}\left[y_{k,t_k}|I_{v+1}\right]}_{\text{revision}}, \qquad (16)$$

where

$$I_{v+1} = [I_{v+1,1} \ldots I_{v+1,J_{v+1}}]', \qquad I_{v+1,j} = y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right], \; j = 1, \ldots, J_{v+1}.$$

I_{v+1} represents the part of the release $\{y_{i_j,t_j}, \; j = 1, \ldots, J_{v+1}\}$, which is "orthogonal" to the information already contained in Ω_v. In other words, it is the "unexpected", with respect to the model, part of the release. Therefore, we label I_{v+1} as the *news*. Note that it is the *news* and not the release itself that leads to forecast revision. In particular, if the new numbers in Ω_{v+1} are exactly as predicted, given the information in Ω_v, or in other words "there is no *news*", the forecast will not be revised.

We can further develop the expression for the revision as:

$$\mathbb{E}\left[y_{k,t_k}|I_{v+1}\right] = \mathbb{E}\left[y_{k,t_k}I_{v+1}'\right]\mathbb{E}\left[I_{v+1}I_{v+1}'\right]^{-1}I_{v+1}.$$

In order to find $\mathbb{E}\left[y_{k,t_k}I_{v+1}'\right]$ and $\mathbb{E}\left[I_{v+1}I_{v+1}'\right]$ under the assumption that the data generating process is given by (1)-(2) and (4), let us first note that[20]

$$
\begin{aligned}
y_{k,t_k} &= \Lambda_{k\cdot}f_{t_k} + \epsilon_{k,t_k} &&\text{and} \\
I_{v+1,j} &= y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right] = \Lambda_{i_j\cdot}\left(f_{t_j} - \mathbb{E}\left[f_{t_j}|\Omega_v\right]\right) + \epsilon_{i_j,t_j}.
\end{aligned}
$$

Consequently, j^{th} element of $\mathbb{E}\left(y_{k,t_k}I_{v+1}'\right)$ and the element in j^{th} row and l^{th} column of $\mathbb{E}\left(I_{v+1}I_{v+1}'\right)$ are given by

$$
\begin{aligned}
\mathbb{E}\left(y_{k,t_k}I_{v+1,j}\right) &= \Lambda_{k\cdot}\mathbb{E}\left[\left(f_{t_k} - \mathbb{E}\left[f_{t_k}|\Omega_v\right]\right)\left(f_{t_j} - \mathbb{E}\left[f_{t_j}|\Omega_v\right]\right)'\right]\Lambda_{i_j\cdot}' &&\text{and} \\
\mathbb{E}\left(I_{v+1,j}I_{v+1,l}\right) &= \Lambda_{i_j\cdot}\mathbb{E}\left[\left(f_{t_j} - \mathbb{E}\left[f_{t_j}|\Omega_v\right]\right)\left(f_{t_l} - \mathbb{E}\left[f_{t_l}|\Omega_v\right]\right)'\right]\Lambda_{i_l\cdot}' + 1_{\{j=l\}}R_{jj},
\end{aligned}
$$

where R_{jj} is the jth element of the diagonal of the residual covariance matrix R. The expectations $\mathbb{E}\left[\left(f_{t_j} - \mathbb{E}\left[f_{t_j}|\Omega_v\right]\right)\left(f_{t_l} - \mathbb{E}\left[f_{t_l}|\Omega_v\right]\right)'\right]$ can be obtained from the Kalman smoother, see the Appendix for more details on the derivations.

As a result, we can find a vector $B_{v+1} = [b_{v+1,1}, \cdots, b_{v+1,J_{v+1}}]$ such that the following holds:

$$\underbrace{\mathbb{E}\left[y_{k,t_k}|\Omega_{v+1}\right] - \mathbb{E}\left[y_{k,t_k}|\Omega_v\right]}_{\text{revision}} = B_{v+1}I_{v+1} = \sum_{j=1}^{J_{v+1}} b_{v+1,j}\underbrace{\left(y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right]\right)}_{\text{news}}. \qquad (17)$$

In other words, the revision can be decomposed as a weighted average of the *news* in the latest release. What matters for the revision is both the size of the *news* as well as its relevance for the variable of interest, as represented by the associated weight $b_{v+1,j}$.

Formula (17) can be considered as a generalisation of the usual Kalman filter update equation (see e.g. Harvey, 1989, eq. 3.2.3a) to the case in which "new" data arrive in a non-synchronous manner.

---

[20]For the case with the idiosyncratic component following an AR(1) process, f_t and Λ should be simply replaced by respectively $\tilde{f}_t$ and $\tilde{\Lambda}$.

Relationship (17) enables us to trace sources of forecast revisions.[21] More precisely, in the case of a simultaneous release of several (groups of) variables it is possible to decompose the resulting forecast revision into contributions from the *news* in individual (groups of) series, see the illustration in Section 4.4.[22] In addition, we can produce statements like e.g. "after the release of industrial production, the forecast of GDP went up because the indicators turned out to be (on average) higher than expected".[23]

3   Monte Carlo evidence

In this section, we perform a Monte Carlo experiment in order to assess how the estimation methodology described above performs in finite sample for different fractions of missing data.

We follow Doz, Giannone, and Reichlin (2006) and generate the data from the following (approximate) factor model:

$$
\begin{aligned}
y_t &= \chi_t + \epsilon_t = \Lambda_0 f_t + \cdots + \Lambda_s f_{t-s} + \epsilon_t\,, \\
f_t &= A f_{t-1} + u_t\,, & u_t &\sim \text{i.i.d. } \mathcal{N}(0, I_r)\,, \\
\epsilon_t &= D \epsilon_{t-1} + v_t\,, & v_t &\sim \text{i.i.d. } \mathcal{N}(0, \Phi)\,,
\end{aligned}
$$

$t = 1, \ldots, T$, where

$$
\begin{aligned}
\Lambda_{ij,k} &\sim \text{ i.i.d. } \mathcal{N}(0,1)\,, & i &= 1,\ldots,n,\ j = 1,\ldots,r,\ k = 0,\ldots,s\,, \\
A_{ij} &= \left\{ \begin{array}{ll} \rho, & i = j \\ 0, & i \neq j \end{array} \right.\,, & D_{ij} &= \left\{ \begin{array}{ll} \alpha, & i = j \\ 0, & i \neq j \end{array} \right.\,, \\
\Phi_{i,j} &= \tau^{|i-j|}(1-\alpha^2)\sqrt{\gamma_i \gamma_j}\,, & \gamma_i &= \frac{\beta_i}{1-\beta_i}\frac{1}{1-\rho^2}\sum_{k=0}^{s}\sum_{j=1}^{r}\Lambda_{ij,k}^2\,, \quad \beta_i \sim \text{i.i.d. } U\big([u, 1-u]\big)\,.
\end{aligned}
$$

Parameters α and τ govern the degree of, respectively, serial- and cross-correlation of the idiosyncratic component. $\tau > 0$ violates the assumption of diagonal spectral density matrix of the idiosyncratic component required for exact factor model, however the condition of weak cross-correlation (for an approximate factor model) is satisfied, see e.g. Doz, Giannone, and Reichlin (2006). For $s > 0$ the relationship between the factors and the observables is dynamic. It may arise in case of lead-lag relationships between the observables. Such model has a representation given by (1)-(2) with Q of reduced rank, see e.g. Bai and Ng (2007). Parameter β_i governs the signal to noise ratio for variable i. More precisely $\beta_i = \frac{\text{Var}(\epsilon_{it})}{\text{Var}(y_{it})}$. Similar process was used in the Monte Carlo experiment of Stock and Watson (2002a) (with a different pattern of idiosyncratic cross-correlation).

We generate the data for different cross-section size n, sample length T, number of factors r and different values of ρ, α, τ and s. We also consider the case in which the number of factors $\hat{r}$ as input into the estimation procedure is larger than the true number of factors r (input into the data generating process).

---

[21]Note, that the contribution from the *news* is equivalent to the change in the overall contribution of the series to the forecast (the measure proposed in Bańbura and Rünstler, 2010) when the correlations between the predictors are not exploited in the model. Otherwise, those measures are different, see the Appendix for the details. In particular, there can be a change in the overall contribution of a variable even if no new information on this variable was released. Therefore *news* is a better suited tool for analysing the sources of forecasts revisions.

[22]If the release concerns only one group or one series, the contribution of its *news* is simply equal to the change in the forecast.

[23]This holds of course for the indicators with positive entries in $b_{v+1,j}$.

Estimating the space spanned by the factors

In this experiment we generate the data from the process described above and subsequently we set a certain fraction of the data as missing (we choose the data points randomly). We consider the cases of 0, 10, 25 and 40% of missing data. Subsequently, we estimate the model using the EM algorithm described above under the assumption of lack of serial correlation in the idiosyncratic component (assumption (4)) and run the Kalman smoother to estimate the factors (we label this approach as BM). We also compare the results of the methodology described in this paper with the ones obtained using the algorithm of Rubin and Thayer (1982) (labelled as RT) and of Stock and Watson (2002b) (labelled as SW). As mentioned above, one of the key differences between these approaches and the one advocated in this paper is that the former do not model the dynamics of the common factors.

To assess the precision of the estimates of the factors we follow Stock and Watson (2002a) and Doz, Giannone, and Reichlin (2006) and use the trace R^2 of the regression of the estimated factors on the true ones:

$$\frac{\text{Trace}\big(F'\hat{F}\big(\hat{F}'\hat{F}\big)^{-1}\hat{F}'F\big)}{\text{Trace}\big(F'F\big)},$$

where $\hat{F} = \mathbb{E}_{\hat{\theta}}[F|\Omega_T]$. This measure is smaller than 1 and tends to 1 with the increasing canonical correlation between the estimated and the true factors.

Tables 1 and 2 present the average trace statistics over 500 Monte Carlo replications for the number of factors $r = 1$ and $r = 3$, respectively. First section of the tables reports the trace statistics for the BM approach. The remaining two sections report the trace statistics of BM relative to the trace statistics of RT and SW approaches (BM/RT and BM/SW respectively). Ratio larger than 1 indicates that BM estimates are on average more precise. For better readability, we highlight the ratios lower than 0.95 in green, higher than 1.05 but lower than 1.1 in orange and higher than 1.1 in red.

Let us first look at the trace statistics for the BM approach. We can see that the space spanned by the estimated factors converges to the true one with increasing T and n. The finite sample precision, however, depends on the fraction of missing data, the number of factors and other parameters of the data generating process. The estimates are less precise for more persistent factors ($\rho = 0.9$ vs $\rho = 0.5$), for larger number of factors ($r = 3$ vs $r = 1$) and for a miss-specified model ($d, \tau > 0$) in small sample. The estimation accuracy decreases with increasing fraction of missing data, however the losses are not that large, especially for $n \geq 50$. Finally, the procedure is rather robust to a miss-specified number of factors.

As for the comparison with RT and SW approaches, they are in most of the cases outperformed by BM (the ratios are mostly larger than 1). The largest gains for BM occur, in general, for smaller samples, larger fraction of missing data, more persistent factors and more dimensional factor space. In addition, BM gains a lot in relative accuracy for a "truly" dynamic model, in which observables load the factors and their lags ($s = 1$). Finally, among the "static" approaches, RT seems to perform better than SW.

As for the model given by (15), in which the idiosyncratic component is modelled as AR(1) process the trace statistics are similar as reported above. This suggests that if we are only interested in estimating the factors we do not win much by accounting for the serial correlation in the idiosyncratic component (as long as it is not too strong). Table 3 below reports the average over i of mean absolute estimation error of the idiosyncratic autoregressive parameter α_i for $n = 25$, $\rho = 0.7$, $\alpha = 0.7$, $\tau = 0$, $\beta \sim U[0.1\ 0.9]$, $s = 0$, $\hat{r} = r = 3$ and different values of T. We consider panels with no missing data and with 20% fraction of missing values. We can see that the estimates converge towards the true values as the sample size increases. In addition, the estimates based on the data with missing values are slightly less accurate.

Table 1: Monte Carlo analysis, trace R-square for the factor estimates, $r = 1$

		BM				BM/RT				BM/SW			
n	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
		$\rho = 0.7$, $\alpha=0$, $\tau = 0$, $\beta\sim U[0.1\ 0.9]$, $s=0$, r_hat = r											
10	50	0.84	0.84	0.82	0.80	1.01	1.01	1.01	1.03	1.02	1.02	1.03	1.05
10	100	0.89	0.89	0.87	0.85	1.01	1.01	1.01	1.02	1.02	1.03	1.04	1.06
25	50	0.88	0.88	0.87	0.86	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
25	100	0.92	0.92	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.02
50	50	0.89	0.89	0.89	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
50	100	0.94	0.93	0.93	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
100	50	0.90	0.90	0.89	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	100	0.94	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\rho = 0.7$, $\alpha=0.5$, $\tau = 0.5$, $\beta\sim U[0.1\ 0.9]$, $s=0$, r_hat = r											
10	50	0.81	0.81	0.79	0.77	1.00	1.00	1.00	1.01	0.99	1.00	1.01	1.05
10	100	0.86	0.86	0.84	0.82	1.00	1.00	1.01	1.01	0.99	1.00	1.01	1.03
25	50	0.88	0.88	0.88	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
25	100	0.93	0.92	0.92	0.91	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01
50	50	0.90	0.90	0.89	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	100	0.94	0.94	0.94	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
100	50	0.91	0.91	0.90	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	100	0.95	0.95	0.95	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\rho = 0.5$, $\alpha=0$, $\tau = 0$, $\beta\sim U[0.1\ 0.9]$, $s=0$, r_hat = r											
10	50	0.89	0.88	0.86	0.82	1.00	1.00	1.00	1.01	1.02	1.02	1.02	1.03
10	100	0.92	0.91	0.89	0.86	1.00	1.00	1.01	1.01	1.02	1.03	1.03	1.05
25	50	0.93	0.92	0.92	0.90	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
25	100	0.95	0.95	0.94	0.93	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
50	50	0.94	0.94	0.93	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
50	100	0.96	0.96	0.96	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
100	50	0.95	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	100	0.97	0.97	0.97	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\rho = 0.9$, $\alpha=0$, $\tau = 0$, $\beta\sim U[0.1\ 0.9]$, $s=0$, r_hat = r											
10	50	0.69	0.69	0.68	0.66	1.02	1.03	1.04	1.07	1.04	1.04	1.06	1.09
10	100	0.80	0.80	0.79	0.78	1.02	1.02	1.03	1.06	1.04	1.04	1.06	1.09
25	50	0.72	0.72	0.71	0.70	1.00	1.00	1.01	1.01	1.01	1.01	1.02	1.02
25	100	0.82	0.82	0.82	0.81	1.00	1.00	1.01	1.01	1.01	1.01	1.02	1.02
50	50	0.73	0.73	0.72	0.72	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01
50	100	0.83	0.83	0.83	0.83	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01
100	50	0.73	0.73	0.73	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	100	0.84	0.84	0.83	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\rho = 0.7$, $\alpha=0$, $\tau = 0$, $\beta\sim U[0.1\ 0.9]$, $s=0$, r_hat = r+1											
10	50	0.84	0.84	0.82	0.78	1.00	1.01	1.01	1.01	1.01	1.02	1.03	1.09
10	100	0.89	0.89	0.87	0.84	1.01	1.01	1.01	1.02	1.02	1.03	1.04	1.09
25	50	0.88	0.88	0.87	0.86	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
25	100	0.92	0.92	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.02
50	50	0.89	0.89	0.88	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
50	100	0.94	0.93	0.93	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
100	50	0.89	0.89	0.89	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	100	0.94	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\rho = 0.7$, $\alpha=0$, $\tau = 0$, $\beta\sim U[0.1\ 0.9]$, $s=1$, r_hat = r											
25	50	0.88	0.87	0.87	0.85	1.03	1.03	1.04	1.06	1.04	1.05	1.06	1.09
25	100	0.93	0.92	0.92	0.91	1.03	1.03	1.04	1.05	1.04	1.04	1.06	1.08
50	50	0.89	0.89	0.88	0.88	1.01	1.01	1.02	1.03	1.02	1.02	1.03	1.04
50	100	0.94	0.94	0.93	0.93	1.01	1.01	1.02	1.02	1.02	1.02	1.03	1.03
100	50	0.89	0.89	0.89	0.89	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02
100	100	0.94	0.94	0.94	0.94	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02

Notes: Table reports average trace R-square for the factor estimates. *BM* refers to the estimation method studied in this paper, *RT* to the algorithm proposed by Rubin and Thayer (1982) and *SW* refers to the algorithm of Stock and Watson (2002). We report the trace R-square for BM, as well as its ratio to the trace R-square of *RT* and *SW*. *0%*, *10%*, *25%* and *40%* refer to the fraction of missing data. The number of factors is $r = 1$. T and n refer to the sample and cross-section size, respectively. s is the number of lags of the factors included in the measurement equation. The parameters ρ, α, τ and β govern the persistence of the factors, the degree of serial- and cross-correlation of the idiosyncratic component and its relative variance, respectively. r_hat is the number of factors with which the models are estimated.

Table 2: Monte Carlo analysis, trace R-square for the factor estimates, $r = 3$

		BM				BM/RT				BM/SW			
n	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
$\rho = 0.7,\ \alpha=0,\ \tau = 0,\ \beta\sim U[0.1\ 0.9],\ s=0,\ r\_hat = r$													
10	50	0.67	0.64	0.58	0.50	1.05	1.06	1.07	1.09	1.07	1.11	1.18	1.38
10	100	0.75	0.73	0.68	0.62	1.08	1.09	1.12	1.17	1.12	1.15	1.27	1.51
25	50	0.82	0.81	0.78	0.73	1.01	1.01	1.02	1.05	1.04	1.04	1.07	1.12
25	100	0.88	0.87	0.85	0.82	1.01	1.02	1.02	1.04	1.04	1.05	1.06	1.11
50	50	0.86	0.86	0.84	0.82	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.04
50	100	0.91	0.91	0.90	0.89	1.00	1.00	1.00	1.01	1.02	1.02	1.02	1.03
100	50	0.88	0.88	0.87	0.86	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
100	100	0.93	0.93	0.92	0.92	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$\rho = 0.7,\ \alpha=0.5,\ \tau = 0.5,\ \beta\sim U[0.1\ 0.9],\ s=0,\ r\_hat = r$													
10	50	0.60	0.58	0.56	0.51	1.03	1.03	1.05	1.10	1.01	1.03	1.11	1.34
10	100	0.63	0.62	0.60	0.56	1.04	1.05	1.07	1.11	1.02	1.04	1.14	1.38
25	50	0.77	0.77	0.74	0.70	0.99	0.99	1.00	1.01	1.01	1.02	1.03	1.07
25	100	0.84	0.83	0.81	0.78	1.00	1.00	1.01	1.02	1.02	1.02	1.04	1.08
50	50	0.85	0.85	0.83	0.81	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.03
50	100	0.91	0.90	0.89	0.88	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.02
100	50	0.88	0.87	0.87	0.86	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
100	100	0.93	0.93	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$\rho = 0.5,\ \alpha=0,\ \tau = 0,\ \beta\sim U[0.1\ 0.9],\ s=0,\ r\_hat = r$													
10	50	0.69	0.65	0.58	0.47	1.01	1.01	1.00	0.98	1.03	1.04	1.12	1.23
10	100	0.75	0.72	0.67	0.58	1.03	1.03	1.05	1.05	1.07	1.09	1.19	1.39
25	50	0.86	0.85	0.81	0.75	1.00	1.01	1.01	1.02	1.03	1.03	1.05	1.08
25	100	0.90	0.89	0.86	0.82	1.00	1.01	1.01	1.02	1.03	1.04	1.05	1.08
50	50	0.91	0.90	0.89	0.86	1.00	1.00	1.00	1.00	1.01	1.02	1.02	1.03
50	100	0.94	0.93	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.02	1.02	1.03
100	50	0.93	0.92	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
100	100	0.96	0.95	0.95	0.94	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$\rho = 0.9,\ \alpha=0,\ \tau = 0,\ \beta\sim U[0.1\ 0.9],\ s=0,\ r\_hat = r$													
10	50	0.58	0.56	0.53	0.49	1.12	1.14	1.19	1.26	1.13	1.17	1.32	1.59
10	100	0.71	0.70	0.67	0.64	1.14	1.17	1.22	1.33	1.17	1.22	1.37	1.79
25	50	0.66	0.66	0.64	0.61	1.03	1.04	1.06	1.10	1.05	1.07	1.10	1.17
25	100	0.78	0.77	0.76	0.75	1.03	1.03	1.05	1.08	1.05	1.06	1.09	1.14
50	50	0.69	0.69	0.68	0.66	1.01	1.01	1.02	1.03	1.02	1.03	1.04	1.06
50	100	0.80	0.79	0.79	0.78	1.01	1.01	1.01	1.02	1.02	1.02	1.03	1.04
100	50	0.70	0.70	0.70	0.69	1.00	1.00	1.00	1.01	1.01	1.01	1.02	1.02
100	100	0.81	0.81	0.80	0.80	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.02
$\rho = 0.7,\ \alpha=0,\ \tau = 0,\ \beta\sim U[0.1\ 0.9],\ s=0,\ r\_hat = r+1$													
10	50	0.69	0.66	0.60	0.52	1.05	1.06	1.05	1.04	1.05	1.09	1.24	1.44
10	100	0.76	0.74	0.70	0.63	1.08	1.10	1.12	1.15	1.10	1.15	1.31	1.65
25	50	0.82	0.81	0.78	0.73	1.01	1.01	1.02	1.03	1.03	1.03	1.05	1.10
25	100	0.88	0.87	0.85	0.82	1.01	1.01	1.02	1.03	1.03	1.04	1.06	1.10
50	50	0.85	0.85	0.84	0.82	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.04
50	100	0.91	0.91	0.90	0.88	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.03
100	50	0.87	0.87	0.86	0.85	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
100	100	0.93	0.93	0.92	0.91	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$\rho = 0.7,\ \alpha=0,\ \tau = 0,\ \beta\sim U[0.1\ 0.9],\ s=1,\ r\_hat = r$													
25	50	0.81	0.79	0.74	0.66	1.09	1.10	1.12	1.16	1.12	1.14	1.18	1.28
25	100	0.87	0.86	0.83	0.78	1.10	1.12	1.15	1.20	1.15	1.18	1.24	1.40
50	50	0.87	0.86	0.83	0.78	1.04	1.05	1.07	1.10	1.08	1.09	1.12	1.15
50	100	0.91	0.91	0.90	0.87	1.04	1.05	1.07	1.10	1.07	1.09	1.12	1.18
100	50	0.89	0.88	0.88	0.86	1.02	1.02	1.03	1.05	1.03	1.04	1.06	1.10
100	100	0.93	0.93	0.92	0.91	1.02	1.02	1.03	1.04	1.03	1.04	1.05	1.08

Notes: Notes for Table 1 apply with the difference that the number of factors is $r = 3$.

Table 3: Mean absolute estimation error for the idiosyncratic autoregressive parameter

T	50	100	200	500	1000
No missing data	0.127	0.075	0.047	0.028	0.020
20% missing data	0.135	0.079	0.050	0.030	0.021

Notes: Table reports the average over i of mean absolute estimation error of α_i for different ratios of missing data for data simulated from a factor model. T refers to the sample size, the size of cross-section n is equal to 25. Further, $\rho = 0.7$, $\alpha = 0.7$, $\tau = 0$, $\beta \sim U[0.1\ 0.9]$, $s = 0$ and $\hat{r} = r = 3$.

Forecasting

In this exercise we evaluate the three approaches in terms of forecast accuracy. In order to mimic data availability patterns typically encountered in real-time forecasting, we assume a different pattern of missing data than in the previous exercise.

Specifically, we are interested in forecasting $y_{1,T}$ and we consider following four data availability patterns:
- *hor 1*: 20% of data points at time T are missing (including $y_{1,T}$)
- *hor 2*: 20% and 40% of data points at time $T-1$ and T respectively, are missing (including $y_{1,T-1}$ and $y_{1,T}$)
- *hor 3*: 20%, 40% and 60% of data points at time $T-2$, $T-1$ and T respectively, are missing (including $y_{1,T-2}$, $y_{1,T-1}$ and $y_{1,T}$)
- *hor 4*: 20%, 40%, 60% and 80% of data points at time $T-3$, $T-2$, $T-1$ and T respectively, are missing (including $y_{1,T-3}$, $y_{1,T-2}$, $y_{1,T-1}$ and $y_{1,T}$).

We label these availability patterns as *hor 1*, ..., *hor 4* as they can be associated with an (increasing) forecast horizon for $y_{1,T}$. Note that with decreasing forecast horizon the data set is "expanding" in the sense discussed in Section 2.3.

We measure the forecast accuracy relative to the accuracy of the unfeasible forecast based on true common component $\chi_{1,t}$. Specifically, Table (4) reports

$$1 - \frac{(\chi_{1,T} - \mathbb{E}_{\hat{\theta}}[y_{1,T}|\Omega_T])^2}{\text{Var}(\chi_{1,t})} = 1 - \frac{(\chi_{1,T} - \hat{\chi}_{1,T})^2}{\text{Var}(\chi_{1,t})}, \tag{18}$$

where $\hat{\chi}_{1,T} = \mathbb{E}_{\hat{\theta}}[\chi_{1,T}|\Omega_T]$. This measure is smaller than 1 and tends to 1 as the estimated forecast approaches the unfeasible one. We also present the forecast accuracy statistics of *BM* relative to that of *RT* and *SW* approaches (*BM/RT* and *BM/SW* respectively). Again, ratio larger than 1 indicates that the *BM* forecasts are on average more accurate. We apply the same highlighting principle as in the previous exercise. '−' entries correspond to the cases in which (18) is negative (the variance of the forecast error is larger than the variance of the common component). We consider the case of $r = 3$ and the same parameterisations for the data generating process as in the previous exercise.

Starting with the results for the *BM* approach, again, forecast accuracy increases with increasing sample length and cross-section size. For $n = 10$ and large fraction of missing data the forecasts are rather inaccurate (especially for the miss-specified model with $d, \tau > 0$). For $n = 100$, on the other hand, we are relatively close to the unfeasible forecast. In this cases accuracy losses due to missing data are not that large either. In contrast to the results of the previous exercise, more persistent factors result in more accurate forecasts. Accuracy losses due to incorrect number of factors are larger but still limited.

Table 4: Monte Carlo analysis, forecast accuracy relative to unfeasible forecast, r = 3

		BM				BM/RT				BM/SW			
n	T	hor 1	hor 2	hor 3	hor 4	hor 1	hor 2	hor 3	hor 4	hor 1	hor 2	hor 3	hor 4
ρ = 0.7, α=0, τ = 0, β=0.5, s=0, r_hat = r													
10	50	0.53	0.40	0.31	0.13	1.08	1.05	1.15	1.51	-	-	-	-
10	100	0.66	0.58	0.50	0.31	1.09	1.11	1.18	1.34	4.65	-	-	-
25	50	0.76	0.70	0.63	0.45	1.02	1.03	1.08	1.25	1.07	1.12	1.80	-
25	100	0.81	0.77	0.71	0.54	1.01	1.03	1.06	1.18	1.05	1.08	1.27	-
50	50	0.82	0.79	0.74	0.60	1.00	1.01	1.02	1.08	1.01	1.02	1.04	1.72
50	100	0.88	0.85	0.81	0.70	1.00	1.01	1.02	1.07	1.01	1.02	1.04	1.33
100	50	0.86	0.83	0.80	0.72	1.00	1.00	1.01	1.03	1.00	1.00	1.01	1.09
100	100	0.91	0.89	0.87	0.80	1.00	1.00	1.01	1.01	1.00	1.00	1.01	1.04
ρ = 0.7, α=0.5, τ = 0.5, β=0.5, s=0, r_hat = r													
10	50	0.21	0.15	0.00	-	0.91	0.75	-	-	-	-	-	-
10	100	0.37	0.33	0.22	0.08	1.05	0.98	1.08	0.93	-	-	-	-
25	50	0.47	0.40	0.31	0.10	0.95	0.93	1.00	0.70	0.95	1.04	3.65	-
25	100	0.65	0.59	0.51	0.38	1.00	1.00	1.04	1.11	1.00	1.04	1.43	-
50	50	0.65	0.58	0.50	0.38	0.98	0.97	0.98	1.04	0.97	0.96	0.99	1.79
50	100	0.78	0.75	0.69	0.57	0.99	0.99	1.01	1.05	0.99	0.99	1.02	1.36
100	50	0.71	0.68	0.61	0.54	0.99	0.99	0.99	0.99	0.97	0.97	0.98	1.03
100	100	0.84	0.83	0.79	0.73	1.00	1.00	1.00	1.01	0.99	0.99	1.00	1.02
ρ = 0.5, α=0, τ = 0, β=0.5, s=0, r_hat = r													
10	50	0.50	0.40	0.31	0.18	1.01	1.01	0.93	1.01	-	-	-	-
10	100	0.63	0.53	0.44	0.28	1.05	1.06	1.04	1.17	1.91	-	-	-
25	50	0.73	0.66	0.59	0.42	1.01	1.02	1.03	1.19	1.07	1.15	2.14	-
25	100	0.80	0.74	0.66	0.48	1.01	1.02	1.02	1.11	1.04	1.09	1.32	-
50	50	0.81	0.75	0.70	0.58	1.01	1.01	1.02	1.06	1.01	1.03	1.08	1.63
50	100	0.88	0.83	0.77	0.65	1.00	1.01	1.01	1.05	1.01	1.02	1.05	1.39
100	50	0.86	0.84	0.79	0.73	1.00	1.00	1.01	1.02	1.00	1.01	1.02	1.05
100	100	0.92	0.90	0.87	0.79	1.00	1.00	1.00	1.02	1.00	1.00	1.01	1.03
ρ = 0.9, α=0, τ = 0, β=0.5, s=0, r_hat = r													
10	50	0.59	0.48	0.37	0.22	1.20	1.29	1.30	6.13	-	-	-	-
10	100	0.72	0.63	0.56	0.43	1.28	1.37	1.57	4.95	2.78	-	-	-
25	50	0.75	0.69	0.64	0.52	1.07	1.10	1.25	2.03	1.19	1.29	-	-
25	100	0.82	0.77	0.73	0.62	1.04	1.07	1.16	1.71	1.08	1.16	1.51	-
50	50	0.80	0.76	0.72	0.63	1.02	1.03	1.09	1.23	1.05	1.07	1.21	3.11
50	100	0.88	0.84	0.80	0.71	1.01	1.02	1.06	1.20	1.02	1.03	1.09	1.59
100	50	0.84	0.81	0.78	0.72	1.00	1.00	1.02	1.07	1.02	1.02	1.05	1.14
100	100	0.91	0.89	0.87	0.81	1.00	1.00	1.01	1.06	1.00	1.00	1.01	1.07
ρ = 0.7, α=0, τ = 0, β=0.5, s=0, r_hat = r+1													
10	50	0.38	0.27	0.18	0.07	1.07	1.09	1.26	-	-	-	-	-
10	100	0.57	0.50	0.41	0.23	1.13	1.15	1.44	1.91	-	-	-	-
25	50	0.62	0.57	0.49	0.34	1.01	1.00	1.08	1.25	1.50	-	-	-
25	100	0.75	0.70	0.65	0.52	1.03	1.03	1.09	1.21	1.40	3.14	-	-
50	50	0.73	0.71	0.65	0.54	1.00	1.00	1.05	1.14	1.14	1.52	-	-
50	100	0.84	0.82	0.77	0.67	1.01	1.01	1.03	1.06	1.07	1.19	1.57	-
100	50	0.80	0.79	0.74	0.69	1.00	1.00	0.99	1.03	1.07	1.18	2.81	-
100	100	0.88	0.87	0.85	0.79	1.00	0.99	1.00	1.03	1.02	1.10	1.43	14.21
ρ = 0.7, α=0, τ = 0, β=0.5, s=1, r_hat = r													
25	50	0.46	0.40	0.39	0.34	1.17	1.30	1.74	4.45	-	-	-	-
25	100	0.71	0.67	0.62	0.52	1.18	1.23	1.36	2.16	2.78	-	-	-
50	50	0.68	0.62	0.56	0.52	1.17	1.20	1.34	1.83	1.49	2.56	-	-
50	100	0.81	0.77	0.72	0.67	1.09	1.13	1.22	1.58	1.19	1.39	6.46	-
100	50	0.73	0.70	0.67	0.62	1.07	1.10	1.13	1.25	1.13	1.23	1.51	-
100	100	0.85	0.84	0.81	0.76	1.04	1.06	1.08	1.18	1.06	1.10	1.17	2.59

Notes: Table reports average forecast accuracy relative to an unfeasible forecast over the Monte Carlo simulations. *BM* refers to the estimation method studied in this paper, *RT* to the algorithm proposed by Rubin and Thayer (1982) and *SW* refers to the algorithm of Stock and Watson (2002). We report the relative forecast accuracy for BM, as well as its ratio to the corresponding statistics for RT and SW. *hor 1, hor 2, ..., hor 4* refer to the (decreasing) pattern of end-of-sample data availability as described in the main text. The number of factors *r = 3*. *T* and *n* refer to the sample and cross-section size, respectively. *s* is the number of lags of the factors included in the measurement equation. The parameters *ρ, α, τ* and *β* govern the persistence of the factors, the degree of serial- and cross-correlation of the idiosyncratic component and its relative variance, respectively. *r_hat* is the number of factors with which the models are estimated. - means that the variance of the forecast error was larger than the variance of the common component.

As for the comparison with RT and SW approaches, they are outperformed by BM, apart from the case with $d, \tau > 0$ in which RT performs best. Again, the largest improvements in forecast accuracy for BM occur for smaller samples, more persistent factors, larger fraction of missing data and a "truly" dynamic model. In particular, as the forecast horizon increases, so do the accuracy gains of the BM approach over the "static" ones. This shows the importance of exploiting the dynamics in case of incomplete cross-section at the end of the sample. In these cases SW yields rather poor forecasts, with the variance of the forecast error larger than the variance of the common component.

4 Empirical application

In this section we employ the methodology developed in Section 2 for two applications: nowcasting and backdating of euro area GDP.

4.1 Data set

We evaluate the methodology on panels with different size of the cross-section, corresponding to different level of (sectoral) disaggregation of various macro-economic concepts. Sectoral information can provide additional or more robust signal for the variable of interest. Moreover, it is sometimes required to provide a more detailed interpretation of the results. On the other hand, it can lead to model mis-specification in small samples by introducing idiosyncratic cross-correlation. We evaluate robustness of the results to expanding the information set by sectoral information, by considering the following data set compositions:

- *Small* - contains the main indicators of real activity on the total economy, such as industrial production, orders, retail sales, unemployment, European Commission Economic Sentiment Indicator, Purchasing Manager Index, GDP or employment (14 series in total). It also contains financial series such as stock prices index or prices of raw materials.

- *Medium* - in addition to the series contained in *Small* specification, it includes more disaggregated information on industrial production, more disaggregated survey information and national accounts data. This composition contains most of the real key economic indicators reported in monthly reports of the European Commission (46 series in total).

- *Large* - apart from the indicators contained in *Medium*, this specification includes series from the large euro area factor model described in Bańbura and Rünstler (2010) and ECB (2008) (101 series).

The data set contains monthly and quarterly variables. The series observed on a daily basis are converted to monthly frequency by taking monthly averages. The detailed description including the list of the series in each specification, their availability and applied transformations is provided in the Appendix. The data set contains the figures as available on the 15th of October 2009.

4.2 Modelling monthly and quarterly series

Before moving to the applications let us explain how we combine the information from monthly and quarterly variables. In this we follow Mariano and Murasawa (2003) and assume that the frequency of the model is monthly and for each quarterly variable we construct a partially observed monthly counterpart.

Let us illustrate this on the example of GDP. We construct a partially observed monthly GDP (3-month on 3-month) growth rate as:

$$\bar{y}_{1,t}^Q = \begin{cases} \log(GDP_t) - \log(GDP_{t-3}), & t = 3, 6, 9, ... \\ \text{missing}, & \text{otherwise}, \end{cases}$$

where GDP_t denotes the level of GDP observed in month t. In this, we follow the convention that quarterly observations are "assigned" to the third month of each quarter. Further, we use the approximation of Mariano and Murasawa (2003)

$$\bar{y}_{1,t}^Q = (1 + L + L^2)^2 y_{1,t}^Q = y_{1,t}^Q + 2y_{1,t-1}^Q + 3y_{1,t-2}^Q + 2y_{1,t-3}^Q + y_{1,t-4}^Q, \tag{19}$$

where $y_{1,t}^Q$ denotes the unobserved month-on-month GDP growth rate. Finally, we assume that $y_{1,t}^Q$ admits the same factor model representation (1) as the monthly variables, with loadings $\Lambda_{1,Q}$. Combining (1) and (19) results in the following representation for $\bar{y}_{1,t}^Q$:

$$\bar{y}_{1,t}^Q = (1 + L + L^2)^2 (\Lambda_{1,Q} f_t + \varepsilon_{1,t}^Q) = \bar{\Lambda}_{1,Q} \left[f_t' \ f_{t-1}' \ \cdots \ f_{t-4}' \right]' + \bar{\varepsilon}_{1,t}^Q,$$

where $\bar{\Lambda}_{1,Q} = [\Lambda_{1,Q} \ 2\Lambda_{1,Q} \ 3\Lambda_{1,Q} \ 2\Lambda_{1,Q} \ \Lambda_{1,Q}]$ is a (restricted) matrix of loadings on factors and their lags. In an analogous manner we construct $\bar{y}_{2,t}^Q, ..., \bar{y}_{n_Q,t}^Q$ for the remaining $n_Q - 1$ quarterly variables. The details of the resulting joint state space representations are provided in the Appendix.

4.3 Forecast evaluation

We start by evaluating our methodology in nowcasting, which is understood as forecasting the present, the very near future and the very recent past, see e.g. Bańbura, Giannone, and Reichlin (2010b). For a variable such as GDP this is a relevant exercise since, while it is the main indicator of the state of the economy, it is released with a substantial delay (around six weeks in the euro area). In the mean-time it can be forecast using more timely, typically monthly variables.

An important feature of nowcasting models is that they should be able to incorporate the most up-to-date information, which due to non-synchronous releases and publication delays results in an irregular pattern of missing observations at the end of the sample ("ragged edge"). Another source of missing observations is the mixed frequency nature of the data set, as explained above. Finally, several series in the data set, namely Purchasing Managers' Surveys, exhibit missing data at the beginning of the sample. Our methodology can deal with such different patterns of data availability in an automatic manner.

Details of the exercise

We evaluate the average precision of the nowcasts for the three data set compositions in a recursive out-of-sample exercise, replicating at each point of the forecast evaluation sample the real-time data availability pattern specific to that point in time.[24] More precisely, in each month we follow the availability pattern specific to the middle of the month (after the data on industrial production are released). For example, in mid-February the last available figure on industrial production would refer to December of the previous year, while for survey data, which are much more timely, there would be already numbers for January. Accordingly, in the middle of each month the publication lag for industrial production and surveys is two and one month, respectively. Consequently when we evaluate the model in e.g. March, the data for industrial production "end" in January, while for surveys they are available up to February. The same

---

[24]The real-time vintages are not available for all the variables of interest and whole evaluation period, therefore the exercise is "pseudo real-time". That is, we use the final figures as of October 2009, but we observe the real-time data availability.

mechanism is applied to all the variables, taking into account their respective (stylised) publication delays, as reported in the data table in the Appendix. The procedure for quarterly variables follows a similar logic modified to take into account the quarterly frequency of the releases, see e.g. Bańbura and Rünstler (2010) for more formal explanation.

For each reference quarter we produce a sequence of projections, starting with the forecast based on the information available in the first month of the preceding quarter, seven months ahead of the GDP flash release. The second forecast is produced with the information that would be available one month later and the last forecast is based on the information available in the first month of the following quarter, 1 month before the flash release. We denote projections based on the information in preceding, current and following quarter (with respect to the forecast reference quarter) as $Q(-1)$, $Q(0)$ and $Q(+1)$ respectively.[25] Forecasts made in the first, second and third month of a quarter are referred to as $M1$, $M2$ and $M3$ respectively. For example, a forecast made in the first month of preceding quarter ($Q(-1)M1$) means that we project e.g. the second quarter relying on the information available in January (i.e. first month of the first quarter); the third quarter using the information available in April, etc.[26]

For the measure of prediction accuracy we choose the Root Mean Squared Forecast Error (RMSFE). The evaluation sample is 2000 Q1 to 2007 Q4. The recent period including recession has been excluded from the evaluation sample because of the extreme values of the GDP in this period, which could bias the results towards the models that were accurate in this particular quarters. The estimation sample starts in January 1993. We choose a recursive estimation which means that the sample length increases each time that more information becomes available.

We run the out-of-sample forecast evaluation for specifications including 1 to 5 factors ($r = 1, 2, \ldots, 5$) and 1 or 2 lags in the VAR ($p = 1, 2$).[27]

We evaluate the forecasts for the *Small*, *Medium* and *Large* data set compositions, both under assumption of serially uncorrelated or AR(1) idiosyncratic component, see the Appendix for the respective state space representations. For reference, we also consider univariate benchmarks: autoregressive model with number of lags chosen by AIC and a sample mean of the GDP growth rate. Finally, we reproduce the forecasts from the factor model proposed by Bańbura and Rünstler (2010) who apply the methodology of Giannone, Reichlin, and Small (2008) to the euro area.

Results of the forecast evaluation

Table 5 presents the results for the different forecast horizons, from the first month of preceding quarter, $Q(-1)M1$, till the first month of the following quarter, $Q(+1)M1$. *Average* gives the average forecast error for all considered horizons. What regards the number of factors, the best parameterisations ex-post and equally weighted forecast combinations over all parameterisations are presented. *AR* and *Mean* refer to results from the univariate benchmarks and *BR* refers to the model of Bańbura and Rünstler (2010).

We can see that all the factor models perform much better than the univariate benchmarks, with largest improvements for shortest forecast horizons. This confirms the importance of relying on timely information contained in monthly indicators (cf. e.g. Giannone, Reichlin, and Small, 2008; Bańbura and Rünstler, 2010).

---

[25]The number in the parenthesis reflects the "shift" with respect to the reference quarter. For example, $Q(-1)$ means that we forecast the reference quarter using the information available in the preceding (-1) quarter.

[26]As GDP is assumed to be "observed" in the third month of the corresponding quarter, a forecast made in the first month of preceding quarter will correspond to a 5-month forecast horizon; a forecast made in the second month of preceding quarter to a 4-month horizon, etc.; with the forecast made in the first month of following quarter corresponding to a -1-month forecast horizon, cf. Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin (2008);

[27]Increasing p to 3 has resulted in a deterioration of the forecast accuracy;

Table 5: Root Mean Squared Forecast Errors for GDP, 2000-2007

Idio	Small		Medium		Large		Benchmarks		
	Uncorr	AR(1)	Uncorr	AR(1)	Uncorr	AR(1)			
	Best ex-post parameterisation						AR	Mean	BR
r,p	2,2	4,2	3,2	5,2	5,2	5,2			
$Q(-1)M1$	0.27	0.25	0.27	0.27	0.27	0.28	0.33	0.32	0.26
$Q(-1)M2$	0.25	0.24	0.23	0.24	0.25	0.25	0.32	0.32	0.24
$Q(-1)M3$	0.24	0.24	0.24	0.24	0.24	0.25	0.32	0.32	0.21
$Q(0)M1$	0.22	0.22	0.22	0.23	0.21	0.23	0.32	0.32	0.21
$Q(0)M2$	0.21	0.22	0.21	0.22	0.22	0.23	0.27	0.31	0.22
$Q(0)M3$	0.20	0.19	0.20	0.18	0.25	0.23	0.27	0.31	0.21
$Q(-1)M1$	0.19	0.17	0.18	0.18	0.21	0.20	0.27	0.31	0.18
Average	0.23	0.22	0.22	0.22	0.24	0.24	0.30	0.31	0.22
	Forecast combination over parameterisations								
$Q(-1)M1$	0.27	0.27	0.27	0.27	0.28	0.29			
$Q(-1)M2$	0.24	0.24	0.24	0.24	0.27	0.26			
$Q(-1)M3$	0.24	0.24	0.25	0.24	0.26	0.26			
$Q(0)M1$	0.23	0.22	0.23	0.23	0.24	0.24			
$Q(0)M2$	0.21	0.22	0.22	0.22	0.24	0.23			
$Q(0)M3$	0.19	0.19	0.19	0.19	0.25	0.24			
$Q(-1)M1$	0.18	0.18	0.18	0.18	0.21	0.20			
Average	0.22	0.22	0.23	0.23	0.25	0.25			

Notes: Table reports Root Mean Squared Forecast Errors (RMSFEs) for different data set compositions. *Small*, *Medium* and *Large* refer to data sets with 14, 46 and 101 variables. The models are estimated by EM algorithm under the assumption of serially uncorrelated (*Uncorr*) or *AR(1)* idiosyncratic component. The upper panel presents the results for the best ex-post parameterisation (in terms of number of factors r and number of their lags in the VAR p). The lower panel gives the RMSFEs for forecast combinations with equal weights across parameterisations with $r = 1, \ldots, 5$ and $p = 1, 2$. $Q(-1)$, $Q(0)$ and $Q(+1)$ refer to forecasts based on the information in preceding, current and following quarter, respectively, and *M1*, *M2* and *M3* to the months within a quarter, *Average* refers to an average RMSFE over the 7 forecast horizons. Benchmarks are the univariate autoregressive model with the number of lags chosen by AIC (*AR*) and the sample *Mean*. In addition, RMSFE for the factor model of Bańbura and Rünstler (2010) are reported (*BR*).

As for different data compositions, the results for specifications *Small* and *Medium* are comparable. In other words, in order to obtain accurate forecasts of GDP, the information on the total economy seems sufficient. This is in line with the results in e.g. Bańbura, Giannone, and Reichlin (2010a) who use US data set and a different methodology. The forecasts from *Large* specification are a bit less accurate. This may point out to difficulties in extracting relevant signal in the presence of indicators of different "quality", as pointed out by e.g. Boivin and Ng (2006).

Concerning the comparison with the model of Bańbura and Rünstler (2010), it performs on average equally well as the *Small* and *Medium* specifications. Bańbura and Rünstler (2010) use a data set that contains, apart from GDP, 76 monthly indicators that are available over the whole estimation period and apply estimation technique based on principal components and Kalman filter. Similar performance of their model suggests, on one hand, that our methodology can reliably extract relevant signal from data sets containing short history and low frequency series, such as Purchasing Managers' Surveys or national accounts and labour market data. On the other hand, it seems that there is no additional information in these series with respect to the data set used by Bańbura and Rünstler (2010). However, including the series in the data set might still be of interest, e.g. for the sake of interpretation of the *news* that their releases carry (see also the next section) or to obtain forecasts or interpolations of various quarterly variables from a single model.

As for the comparison between implementations with serially uncorrelated or AR(1) idiosyncratic component, the results are not clear-cut. For most of the considered parameterisations, modelling serial correlation seems to help for shorter forecast horizons (results for parameterisation not shown in Table 5 are available upon request). For longer horizons, there is no clear ranking between the two implementations. In addition, there is no difference in performance of the corresponding forecast combinations. Therefore, we conclude that what regards GDP, the advantage of modelling the idiosyncratic serial correlation is not obvious. Accounting for serially correlated idiosyncratic component could be more important for monthly variables. This issue is left for future research. Another issue worth exploring is that the optimal parameterisations with serially correlated idiosyncratic component seem to include more common factors than their "uncorrelated" counterparts.

As a final observation, let us point out that forecast combinations over all parameterisations perform equally well or only slightly worse than the best ex-post specification. Hence, averaging over specifications could be a valid strategy in case no single parameterisation performs best for all the horizons or when the best specification is very sensitive to the choice of the evaluation sample. In particular, there have been large differences between the forecasts from various parameterisations in the period of recent recession with different parameterisations performing best at different points in time. In such periods, averaging over parameterisations could be a good strategy.

4.4 News in data releases and forecast revisions

In the following exercise, we produce a sequence of GDP forecasts for the fourth quarter of 2008, at each update we extract the *news* components from various data groups and illustrate how they revise the forecast.

As in the previous section, the sequence of forecasts for the reference quarter is based on "expanding" information sets. The first forecast is performed on the basis of information set available in mid-July 2008 (in the terminology of the previous section this would correspond to the forecast from the first month of preceding quarter). Subsequently, we revise this forecast once a month incorporating new figures, which would have become available in real-time. In this, we follow the stylised release calendar used in the out-of-sample forecast evaluation in the previous section. The last update is based on the data of mid-January 2009 (forecast from the first month of the following quarter) and the actual GDP for the fourth quarter was released in February (flash estimate).

At each update we break down the forecast revision into the contributions of the *news* from the respective predictors using formula (17). In other words, the difference between two consecutive forecasts is the sum of the contributions of the *news* from all the variables plus the effects of model re-estimation. As decomposition (17) holds provided that the expectations are conditional on the same parameter values, the fact that the parameters are re-estimated with each forecast update has to be taken into account separately.

Figure 1 shows the evolution of the forecast as new information arrives, the actual value of the GDP for the fourth quarter and the decomposition of the revisions, obtained with the *Small* and *Medium* specifications (for the best parameterisations ex-post). For the sake of readability the series are grouped into following groups: real variables (*Real News*), European Commission and Purchasing Managers' Surveys (*Surv News*), financial series (*Fin News*) and US data.[28] Category *Re-est* reflects the effects of parameter re-estimation.

---

[28]See the Appendix for the list of series in each group. *Fin* contains also commodity prices. The contribution of a group of series is the sum of the contributions of the series within this group.

Figure 1: Contribution of news to forecast revisions for 2008 Q4: Small and Medium model

S m a l l

M e d i u m
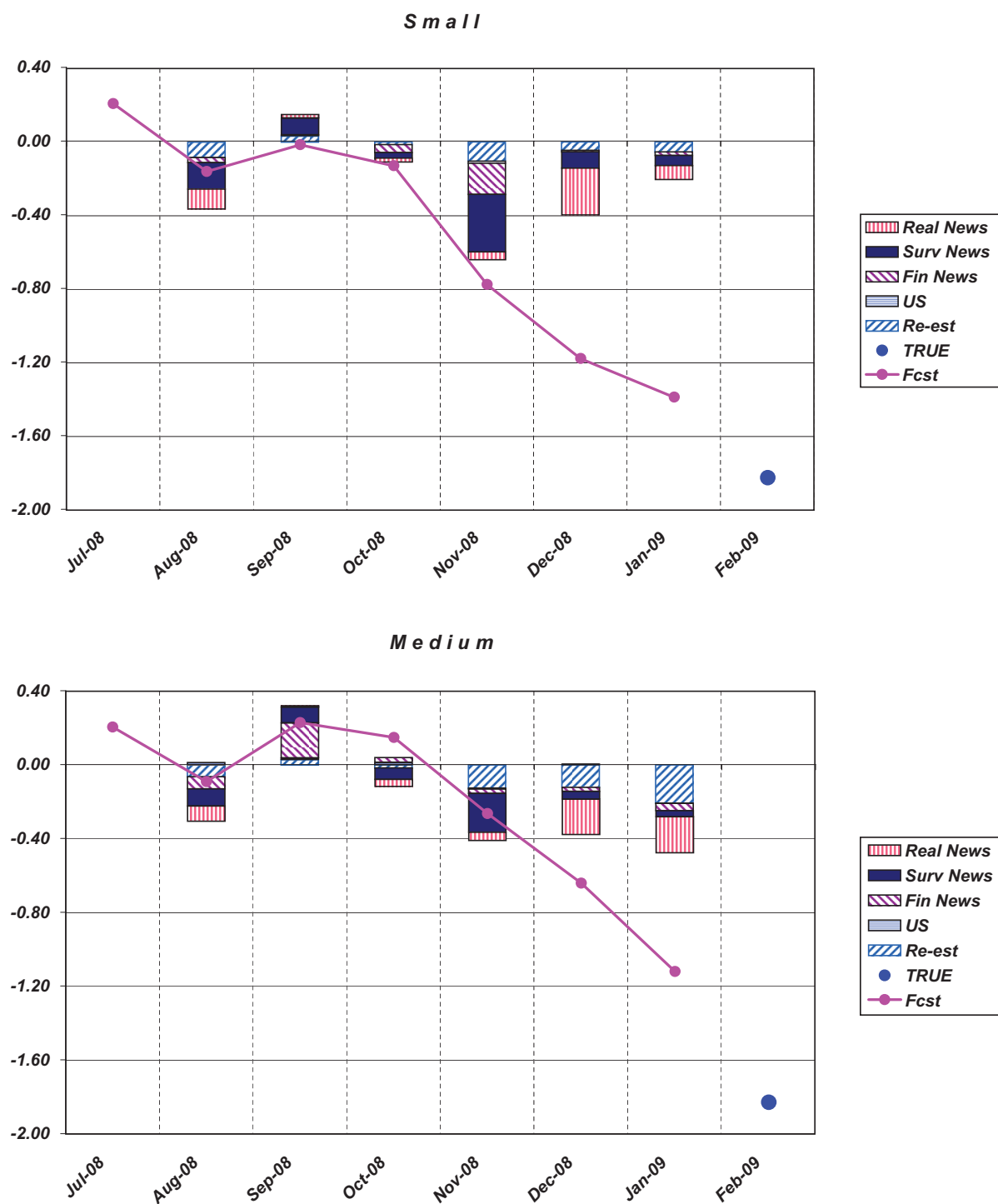
We can observe that forecasts and *news* follow qualitatively similar patterns for both data set compositions. The first forecast is relatively close to the historical average and remains at a relatively high level throughout the third quarter, compared to the actual outcome. This is in line with the results in Giannone, Reichlin, and Small (2008), who compare the accuracy of judgmental and factor model based forecasts and show that they have hard time beating naïve models, such as unconditional mean, for horizons beyond the current quarter. When looking at the contributions from different groups of data, we see that for longer forecast horizons the biggest news impact comes from the surveys. The news from real data becomes important only later in the forecast cycle, when the released numbers refer to the target quarter. This confirms the results of Giannone, Reichlin, and Small (2008) and Bańbura and Rünstler (2010) on the important role of soft data for the GDP projections when the hard data for the relevant periods are not yet available. The impact of news from US and financial data is rather limited. Finally, the effects of the re-estimation are rather large and are most likely due to extreme values that were observed in this period (many of the series, including GDP growth, attained their historical lows, several standard deviations away from their historical averages).

When looking at quantitative results, we can see that there are some differences between the specifications in how the information from new releases is incorporated. Both forecasts start from a similar level but *Small* specification seems to "head" faster towards the true outcome, in particular due to different contribution from the *news* in the financial group (the composition of this group in both specifications is different).

4.5   Backdating

As discussed in Section 2.2, a useful feature of our framework is that Kalman smoother can be applied to obtain the estimates of any missing observations in the data set which can be used e.g. in order to backdate a short history series or to interpolate a low frequency variable.

In this section we illustrate this by applying the methodology to backdating of GDP. For this purpose, we modify the data sets described above by discarding all the observations on GDP prior to March 2001. Further, we estimate the parameters of the models and obtain the estimates of the missing values of GDP from the Kalman smoother.

Figure 2 plots the back estimates of the GDP based on the three considered data set compositions and the actual quarterly growth rate of GDP. We use the best ex-post parameterisations under the assumption of serially correlated idiosyncratic component.

As we can see from Figure 2, independently of the data set used, the back estimates seem to capture well the movements of the GDP, giving reasonable estimates of the past values of the series and the different specifications yield comparable results.

Figure 2: Back estimates of GDP
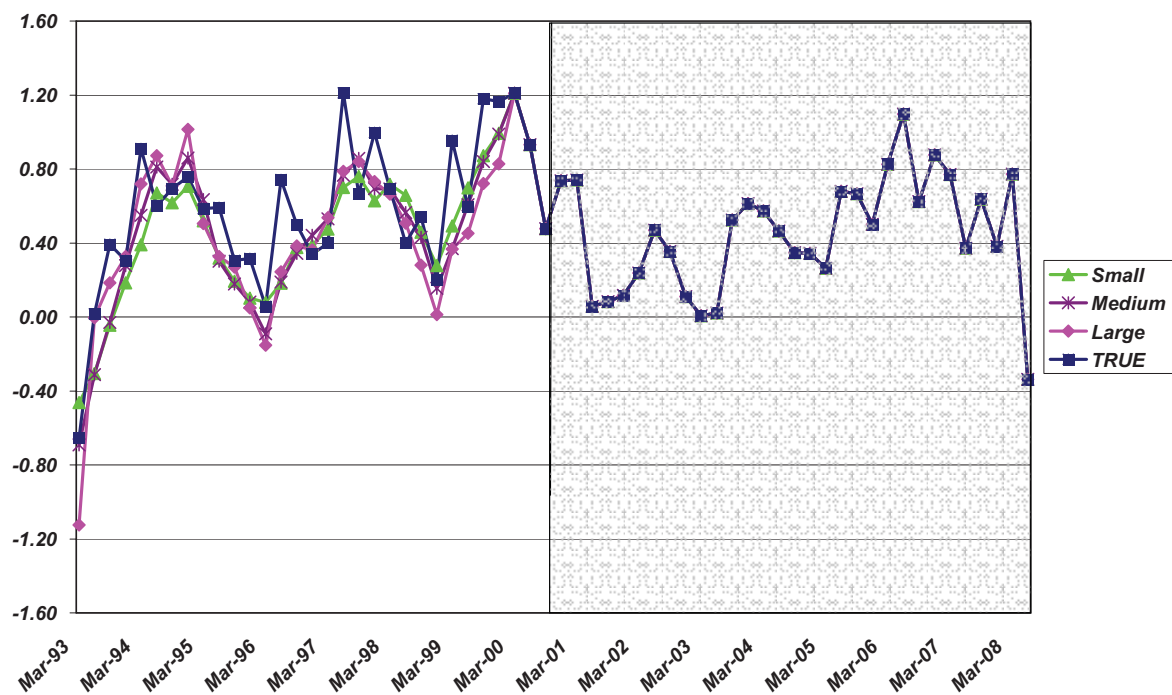
5 Conclusions

This paper proposes a methodology for the estimation of dynamic factor model in the presence of arbitrary pattern of missing data. We show how the steps of the EM algorithm of Watson and Engle (1983) should be modified in the case of missing data. We also propose how to model the dynamics of the idiosyncratic component.

We evaluate the methodology on both simulated and euro area data. Monte Carlo evidence indicates that it performs well, also in case of relatively large fractions of missing observations. We compare our approach to alternative EM algorithms proposed by Rubin and Thayer (1982) and Stock and Watson (2002b). The latter two approaches do not model the dynamics of the latent factors and as a consequence perform worse when such dynamics is strong. The advantage of our methodology is particularly evident in cases of large fraction of missing data and in small samples. The simulations also suggest that accounting for dynamics is important in real-time forecasting/nowcasting applications in which there is a large fraction of missing data at the end of the sample (see also Doz, Giannone, and Reichlin, 2007).

In the empirical part, we apply the methodology to nowcasting and backdating of the euro area GDP on the basis of data sets containing monthly and quarterly series. Thanks to the flexibility of the framework in dealing with missing data, short history and lower frequency (quarterly) variables can be easily incorporated (e.g. Purchasing Managers' Surveys, GDP components or labour statistics). We consider different sizes of cross-section corresponding to different levels of sectoral disaggregation (*Small*, *Medium* and *Large*, including 14, 46 and 101 variables respectively). *Large* specification performs a bit worse than the other two, which could be due to difficulties in extracting relevant signal in the presence of indicators of different "quality", as pointed out by e.g. Boivin and Ng (2006). As for *Small* and *Medium* specifications, they perform comparably, suggesting that, while potentially useful for interpretation, sectoral information is not necessarily needed for an accurate GDP forecast (*Small* specification contains series measuring only total economy concepts). Both specifications perform similarly to the factor model of Bańbura and Rünstler (2010) who adopt the methodology of Giannone, Reichlin, and Small (2008). This shows that, on one hand our approach works well for data sets containing short history and low frequency data such as mentioned above; on the other hand, however, incorporating such data does not lead to improvements in forecast accuracy in case of euro area GDP. The latter observation might, however, not hold for other economies, for which the pool of high frequency and long history information could be more modest. In addition, including the series in the data set might still be of interest, e.g. for the sake of interpretation of the *news* that their releases carry or to obtain forecasts of various quarterly variables from a single model.

Concerning the role of idiosyncratic dynamics, we do not find consistent improvements in the accuracy of GDP forecasts, when taking it explicitly into account. There might be more sizable improvements in the case of monthly variables, which we do not forecast here, see e.g. Stock and Watson (2002b). It is a possible extension of the current application left for future research.

Finally, another methodological contribution of our paper is that we show that a forecast revision which arises as a consequence of a release of new data is a weighted sum of model based *news* from this release. We show how to derive the *news* and the associated weights within our framework. We illustrate how this can be used in nowcasting applications to understand and interpret the contributions of various data releases to forecast updates.

References

ALTISSIMO, F., R. CRISTADORO, M. FORNI, M. LIPPI, AND G. VERONESE (2006): "New EuroCOIN: Tracking Economic Growth in Real Time," CEPR Discussion Papers 5633.

ANGELINI, E., G. CAMBA-MÉNDEZ, D. GIANNONE, G. RÜNSTLER, AND L. REICHLIN (2008): "Short-term forecasts of euro area GDP growth," Working Paper Series 949, European Central Bank.

ANGELINI, E., J. HENRY, AND M. MARCELLINO (2006): "Interpolation and backdating with a large information set," *Journal of Economic Dynamics and Control*, 30(12), 2693–2724.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135–171.

BAI, J., AND S. NG (2007): "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60.

BAŃBURA, M., D. GIANNONE, AND L. REICHLIN (2010a): "Large Bayesian VARs," *Journal of Applied Econometrics*, 25(1), 71–92.

——— (2010b): "Nowcasting," Manuscript.

BAŃBURA, M., AND G. RÜNSTLER (2010): "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*, forthcoming.

BELVISO, F., AND F. MILANI (2006): "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy," *The B.E. Journal of Macroeconomics*, 0(3).

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Measuring Monetary Policy: A Factor Augmented Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.

BERNANKE, B. S., AND J. BOIVIN (2003): "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50(3), 525–546.

BOIVIN, J., AND S. NG (2006): "Are more data always better for factor analysis?," *Journal of Econometrics*, 132(1), 169–194.

BORK, L. (2009): "Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM Algorithm Approach," CREATES Research Papers 2009-11, School of Economics and Management, University of Aarhus.

BORK, L., H. DEWACHTER, AND R. HOUSSA (2009): "Identification of Macroeconomic Factors in Large Panels," CREATES Research Papers 2009-43, School of Economics and Management, University of Aarhus.

BROCKWELL, P., AND R. DAVIS (1991): *Time Series: Theory and Methods*. Springer-Verlag, 2nd edn.

CAMACHO, M., AND G. PEREZ-QUIROS (2008): "Introducing the EURO-STING: Short Term INdicator of Euro Area Growth," Banco de España Working Papers 0807, Banco de España.

CAMBA-MENDEZ, G., G. KAPETANIOS, R. SMITH, AND M. WEALE (2001): "An automatic leading indicator of economic activity: forecasting GDP growth for European countries," *Econometrics Journal*, 4, S56–S90.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51(5), 1281–304.

CHOW, G. C., AND A. LIN (1971): "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series," *The Review of Economics and Statistics*, 53, 372375.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance Measurement with Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

——— (1988): "Risk and Return in an Equilibrium APT: Application to a New Test Methodology," *Journal of Financial Economics*, 21, 255–289.

——— (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood Estimation From Incomplete Data," *Journal of the Royal Statistical Society*, 14, 1–38.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2006): "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," Working Paper Series 674, European Central Bank.

——— (2007): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," CEPR Discussion Papers 6043, C.E.P.R. Discussion Papers.

DURBIN, J., AND S. J. KOOPMAN (2001): *Time Series Analysis by State Space Methods.* Oxford University Press.

ECB (2008): "Short-term forecasts of economic activity in the euro area," in *Monthly Bulletin, April*, pp. 69–74. European Central Bank.

ENGLE, R. F., AND M. W. WATSON (1981): "A one-factor multivariate time series model of metropolitan wage rates," *Journal of the American Statistical Association*, 76, 774–781.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The Generalized Dynamic Factor Model: identification and estimation," *Review of Economics and Statistics*, 82, 540–554.

——— (2003): "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?," *Journal of Monetary Economics*, 50, 1243–55.

——— (2005): "The Generalized Dynamic Factor Model: one-sided estimation and forecasting," *Journal of the American Statistical Association*, 100, 830–840.

FORNI, M., AND L. REICHLIN (1996): "Dynamic Common Factors in Large Cross-Sections," *Empirical Economics*, 21(1), 27–42.

——— (1998): "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," *Review of Economic Studies*, 65(3), 453–73.

GEWEKE, J. F. (1977): "The Dynamic Factor Analysis of Economic Time Series Models," in *Latent Variables in Socioeconomic Models*, ed. by D. Aigner, and A. Goldberger, pp. 365–383. North-Holland.

GEWEKE, J. F., AND K. J. SINGLETON (1980): "Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series.," *International Economic Review*, 22, 37–54.

GIANNONE, D., L. REICHLIN, AND L. SALA (2004): "Monetary Policy in Real Time," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, pp. 161–200. MIT Press.

GIANNONE, D., L. REICHLIN, AND S. SIMONELLI (2009): "Nowcasting Euro Area Economic Activity in Real-Time: The Role of Confidence Indicator," ECARES Working Papers 2009_021, Université Libre de Bruxelles.

GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55(4), 665–676.

HARVEY, A. (1989): *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

HEATON, C., AND V. SOLO (2004): "Identification of causal factor models of stationary time series," *Econometrics Journal*, 7(2), 618–627.

JUNGBACKER, B., S. KOOPMAN, AND M. VAN DER WEL (2009): "Dynamic Factor Analysis in The Presence of Missing Data," Tinbergen Institute Discussion Papers 09-010/4, Tinbergen Institute.

JUNGBACKER, B., AND S. J. KOOPMAN (2008): "Likelihood-based Analysis for Dynamic Factor Models," Tinbergen Institute Discussion Papers 08-007/4, Tinbergen Institute.

KOSE, M. A., C. OTROK, AND C. H. WHITEMAN (2003): "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, 93, 1216–1239.

MARCELLINO, M., J. H. STOCK, AND M. W. WATSON (2003): "Macroeconomic forecasting in the Euro area: Country specific versus area-wide information," *European Economic Review*, 47(1), 1–18.

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

MCLACHLAN, G. J., AND T. KRISHNAN (1996): *The EM Algorithm and Extensions.* John Wiley and Sons.

MODUGNO, M., AND K. NIKOLAOU (2009): "The forecasting power of international yield curve linkages," Working Paper Series 1044, European Central Bank.

PROIETTI, T. (2008): "Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints: Nowcasting Monthly GDP and its Main Components," MPRA Paper 6860, University Library of Munich, Germany.

QUAH, D., AND T. J. SARGENT (1992): "A Dynamic Index Model for Large Cross-Section," in *Business Cycle*, ed. by J. Stock, and M. Watson, pp. 161–200. Univeristy of Chicago Press.

REIS, R., AND M. W. WATSON (2007): "Relative Goods' Prices and Pure Inflation," NBER Working Paper 13615.

RUBIN, D. B., AND D. T. THAYER (1982): "EM Algorithms for ML Factor Analysis," *Psychometrica*, 47, 69–76.

RÜNSTLER, G., AND F. SÉDILLOT (2003): "Short-term estimates of euro area real GDP by means of monthly data," Working Paper Series 276, European Central Bank.

SARGENT, T. J., AND C. SIMS (1977): "Business Cycle Modelling without Pretending to have to much a-priori Economic Theory," in *New Methods in Business Cycle Research*, ed. by C. Sims. Federal Reserve Bank of Minneapolis.

SCHUMACHER, C., AND J. BREITUNG (2008): "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly Data," *International Journal of Forecasting*, 24, 386–398.

SHUMWAY, R., AND D. 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," in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard, and S. Fischer, pp. 351–393. MIT Press.

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

STOCK, J. H., AND M. W. WATSON (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 and Economics Statistics*, 20, 147–162.

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.

A Data Description

	Group	Series	Small	Medium	Large	BR	log	diff	M1	M2	M3
1	Real	IP total	x		x	x	x	x	2	2	2
2	Real	IP tot excl constr		x	x	x	x	x	2	2	2
3	Real	IP tot excl constr&energy			x	x	x	x	2	2	2
4	Real	IP constr			x	x	x	x	2	2	2
5	Real	IP intermediate goods		x	x	x	x	x	2	2	2
6	Real	IP capital		x	x	x	x	x	2	2	2
7	Real	IP durable consumer goods		x	x	x	x	x	2	2	2
8	Real	IP non-durable consumer goods		x	x	x	x	x	2	2	2
9	Real	IP MIG energy		x	x	x	x	x	2	2	2
10	Real	IP excl NACE Rev.1 Section E			x	x	x	x	2	2	2
11	Real	IP manufacturing			x	x	x	x	2	2	2
12	Real	IP manufacture of basic metals			x	x	x	x	2	2	2
13	Real	IP manufacture of chemicals and chemical products			x	x	x	x	2	2	2
14	Real	IP manufacture of electrical machinery			x	x	x	x	2	2	2
15	Real	IP manufacture of machinery and equipment			x	x	x	x	2	2	2
16	Real	IP manufacture of pulp, paper and paper products			x	x	x	x	2	2	2
17	Real	IP manufacture of rubber and plastic products			x	x	x	x	2	2	2
18	Real	New passenger cars	x		x	x	x	x	1	1	1
19	Real	New orders	x	x	x	x	x	x	3	3	3
20	Real	Retail turnover (deflated)	x	x	x	x	x	x	2	2	2
21	Survey	ECS Economic Sentiment indicator		x	x			x	1	1	1
22	Survey	ECS Industrial confidence indicator		x	x	x		x	1	1	1
23	Survey	ECS Industry: assessment of order-book levels		x	x	x		x	1	1	1
24	Survey	ECS Industry: assessment of stocks of finished products			x	x		x	1	1	1
25	Survey	ECS Industry: production expectations		x	x	x		x	1	1	1
26	Survey	ECS Industry: production trend observed in recent months			x	x		x	1	1	1
27	Survey	ECS Industry: assessment of export order-book levels			x	x		x	1	1	1
28	Survey	ECS Industry: employment expectations		x	x	x		x	1	1	1
29	Survey	ECS Consumer Confidence Indicator		x	x	x		x	1	1	1
30	Survey	ECS Consumer: general economic situation over next 12 months			x	x		x	1	1	1
31	Survey	ECS Consumer: unemployment expectations			x	x		x	1	1	1
32	Survey	ECS Consumer: eneral economic situation over last 12 months			x	x		x	1	1	1
33	Survey	ECS Construction Confidence Indicator		x	x	x		x	1	1	1
34	Survey	ECS Construction: order books			x	x		x	1	1	1
35	Survey	ECS Construction: empl exp			x	x		x	1	1	1
36	Survey	ECS Construction: prod recent			x	x		x	1	1	1
37	Survey	ECS Retail Confidence Indicator		x	x	x		x	1	1	1
38	Survey	ECS Retail: present business situation			x	x		x	1	1	1
39	Survey	ECS Retail: assessment of stocks			x	x		x	1	1	1
40	Survey	ECS Retail: expected business situation			x	x		x	1	1	1
41	Survey	ECS Retail: employment expectations			x	x		x	1	1	1
42	Survey	ECS Service Confidence Indicator		x	x	x		x	1	1	1
43	Survey	ECS Service: evolution of employment expected in the months ahead		x	x			x	1	1	1
44	Survey	PMS Composite: output			x			x	1	1	1
45	Survey	PMS Composite: employment			x			x	1	1	1
46	Survey	PMS Manufacturing: purchasing manager index	x	x	x			x	1	1	1
47	Survey	PMS Manufacturing: employment			x			x	1	1	1
48	Survey	PMS Manufacturing: output			x			x	1	1	1
49	Survey	PMS Manufacturing: productivity			x			x	1	1	1
50	Survey	PMS Services: output		x	x			x	1	1	1
51	Survey	PMS Services: employment			x			x	1	1	1
52	Survey	PMS Services: new business			x			x	1	1	1

		Series											
53	Survey	PMS Services: productivity					x				1	1	1
54	Real	Unemployment rate	x	x			x				2	2	2
55	Real	Index of Employment, Total Industry		x			x				4	5	3
56	Real	Index of Employment, Total Industry (excluding construction)				x	x				4	5	3
57	Real	Index of Employment, Construction				x	x				4	5	3
58	Real	Index of Employment, Manufacturing				x	x				4	5	3
59	Real	Extra EA trade, Export value	x	x		x	x				3	3	3
60	Real	Intra EA trade, Export value				x	x				3	3	3
61	Real	Extra EA trade, Import value		x		x	x				3	3	3
62	Real	Intra EA trade, Import value				x	x				3	3	3
63	US	US IP		x		x	x				2	2	2
64	US	US Unemployment rate				x	x				1	1	1
65	US	US Employment			x	x	x				1	1	1
66	US	US retail sales			x	x	x				1	1	1
67	US	US IP manf expectations		x		x	x				1	1	1
68	US	US consumer expectations		x		x	x				1	1	1
69	US	US 3-month interest rate				x	x				1	1	1
70	US	US 10-year interest rate				x	x				1	1	1
71	Financial	M3		x	x	x	x				2	2	2
72	Financial	Loans		x	x	x	x				2	2	2
73	Financial	Interest rate, 10 year		x	x	x	x				1	1	1
74	Financial	Interest rate, 3-month		x	x	x	x				1	1	1
75	Financial	Interest rate, 1-year				x	x				1	1	1
76	Financial	Interest rate, 2-year				x	x				1	1	1
77	Financial	Interest rate, 5-year				x	x				1	1	1
78	Financial	Nominal effective exch. rate		x		x	x				1	1	1
79	Financial	Real effective exch. rate CPI deflated				x	x				1	1	1
80	Financial	Real effective exch. rate producer prices deflated				x	x				1	1	1
81	Financial	Exch. rate EUR/USD		x		x	x				1	1	1
82	Financial	Exch. rate EUR/GBP				x	x				1	1	1
83	Financial	Exch. rate EUR/YEN				x	x				1	1	1
84	Financial	Dow Jones Euro Stoxx 50 Price Index	x	x	x	x	x				1	1	1
85	Financial	Dow Jones Euro Stoxx Price Index			x	x	x				1	1	1
86	Financial	S&P 500 composite price index	x	x	x	x	x				1	1	1
87	Financial	US, STOCK-EXCH. PRICES, DOW JONES, INDUSTRIAL AVERAGE			x	x	x				1	1	1
88	Financial	World market prices of raw materials in Euro. Index Total, excluding Energy		x	x	x	x				1	1	1
89	Financial	World market prices of raw materials, crude oil, USD		x	x	x	x				1	1	1
90	Financial	Gold price, USD/fine ounce			x	x	x				1	1	1
91	Financial	Oil price, brent crude, 1 month forward	x	x	x	x	x				1	1	1
92	Financial	World market prices of raw materials, Index total, Euro			x	x	x				1	1	1
93	Real	GDP	x	x	x	x	x				4	2	3
94	Real	Private Consumption	x	x		x	x				4	5	3
95	Real	Investment		x		x	x				4	5	3
96	Real	Export		x		x	x				4	5	3
97	Real	Import		x		x	x				4	5	3
98	Real	Employment	x	x		x	x				4	5	3
99	Real	Productivity		x		x	x				4	5	3
100	Survey	ECS Capacity utilisation	x	x			x				1	-1	0
101	US	US GDP	x	x	x	x	x				4	2	3

Notes: Columns under *Data set composition* indicate which series were included in each of the specifications. Columns under *Transform* specify whether logarithm and/or differencing was applied to the initial series. The last three columns provide the number of missing observations at the end of the sample caused by the publication delays in each month of a quarter. Negative numbers for capacity utilisation reflect the fact that the figure on the reference quarter is released before the end of this quarter (in its second month). ECS and PMS refer to European Commission and Purchasing Managers Surveys, respectively.

B Derivation of the EM iterations

Let us first sketch the derivation of formulas (5)-(8). They are obtained under the assumption of no serial correlation in the idiosyncratic component and $p = 1$ ($\theta = \{\Lambda, A = A_1, R, Q\}$). Under these assumptions the joint log-likelihood (for the observations and the latent factors) is given by:

$$
\begin{aligned}
l(Y, F; \theta) &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} f_0' \Sigma^{-1} f_0 - \frac{T}{2} \log |Q| - \frac{1}{2} \sum_{t=1}^{T} (f_t - A f_{t-1})' Q^{-1} (f_t - A f_{t-1}) \\
&\quad - \frac{T}{2} \log |R| - \frac{1}{2} \sum_{t=1}^{T} (y_t - \Lambda f_t)' R^{-1} (y_t - \Lambda f_t) \\
&= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} f_0' \Sigma^{-1} f_0 - \frac{T}{2} \log |Q| - \frac{1}{2} \mathrm{tr} \left[Q^{-1} \sum_{t=1}^{T} (f_t - A f_{t-1})(f_t - A f_{t-1})' \right] \\
&\quad - \frac{T}{2} \log |R| - \frac{1}{2} \mathrm{tr} \left[R^{-1} \sum_{t=1}^{T} (y_t - \Lambda f_t)(y_t - \Lambda f_t)' \right].
\end{aligned}
$$

In order to obtain the expressions for $\Lambda(j+1)$ and $A(j+1)$, we need to differentiate $L(\theta, \theta(j)) = \mathbb{E}_{\theta(j)} [l(Y, F; \theta) | \Omega_T]$ with respect to Λ and A respectively. For example, for the latter we get

$$
\begin{aligned}
\frac{\partial \mathbb{E}_{\theta(j)} [l(Y, F; \theta) | \Omega_T]}{\partial A} &= -\frac{1}{2} \frac{\partial \mathrm{tr} \left\{ Q^{-1} \sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [(f_t - A f_{t-1})(f_t - A f_{t-1})' | \Omega_T] \right\}}{\partial A} \\
&= -Q^{-1} \sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [f_t f_{t-1}' | \Omega_T] + Q^{-1} A \sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [f_{t-1} f_{t-1}' | \Omega_T],
\end{aligned}
$$

and consequently

$$
A(j+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [f_t f_{t-1}' | \Omega_T] \right) \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [f_{t-1} f_{t-1}' | \Omega_T] \right)^{-1},
$$

as provided in the main text. In an analogous manner formula (5) for $\Lambda(j+1)$ can be derived. The expressions (7) and (8) for $R(j+1)$ and $Q(j+1)$ are obtained by differentiating $L(\breve{\theta}, \theta(j))$ with respect to R and Q respectively, where $\breve{\theta} = \{\Lambda(j+1), A(j+1), R, Q\}$, see also the comment in footnote 13.

Let us now develop the formulas for $\Lambda(j+1)$ and $R(j+1)$ in the case that y_t contains missing values and (9) no longer holds. Let us differentiate $\mathbb{E}_{\theta(j)} [l(Y, F; \theta) | \Omega_T]$ with respect to Λ:

$$
\frac{\partial \mathbb{E}_{\theta(j)} [l(Y, F; \theta) | \Omega_T]}{\partial \Lambda} = -\frac{1}{2} \frac{\partial \mathrm{tr} \left\{ R^{-1} \sum_{t=1}^{T} \mathbb{E}_{\theta(j)} [(y_t - \Lambda f_t)(y_t - \Lambda f_t)' | \Omega_T] \right\}}{\partial \Lambda} \tag{20}
$$

and let us have a closer look at $\mathbb{E} [(y_t - \Lambda f_t)(y_t - \Lambda f_t)' | \Omega_T]$ (to simplify the notation we skip the subscript $\theta(j)$).

Let

$$
y_t = W_t y_t + (I - W_t) y_t = y_t^{(1)} + y_t^{(2)},
$$

where W_t is a diagonal matrix with ones corresponding to the non-missing entries in y_t and 0 otherwise. ($y_t^{(1)}$ contains the non-missing observations at time t with 0 in place of the missing ones.)

We have:

$$
\begin{aligned}
(y_t - \Lambda f_t)(y_t - \Lambda f_t)' &= \Big(W_t(y_t - \Lambda f_t) + (I - W_t)(y_t - \Lambda f_t) \Big) \Big(W_t(y_t - \Lambda f_t) + (I - W_t)(y_t - \Lambda f_t) \Big)' \\
&= W_t(y_t - \Lambda f_t)(y_t - \Lambda f_t)' W_t + (I - W_t)(y_t - \Lambda f_t)(y_t - \Lambda f_t)'(I - W_t) \\
&\quad + W_t(y_t - \Lambda f_t)(y_t - \Lambda f_t)'(I - W_t) + (I - W_t)(y_t - \Lambda f_t)(y_t - \Lambda f_t)' W_t.
\end{aligned}
$$

By the law of iterated expectations:

$$\mathbb{E}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|\Omega_T\big] = \mathbb{E}\Big[\mathbb{E}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|F, \Omega_T\big]|\Omega_T\Big].$$

As

$$\begin{aligned}
\mathbb{E}\big[W_t(y_t - \Lambda f_t)(y_t - \Lambda f_t)'(I - W_t)|F, \Omega_T\big] &= 0, \\
\mathbb{E}\big[(I - W_t)(y_t - \Lambda f_t)(y_t - \Lambda f_t)'(I - W_t)|F, \Omega_T\big] &= (I - W_t)R(j)(I - W_t)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\big[W_t(y_t - \Lambda f_t)(y_t - \Lambda f_t)'W_t|\Omega_T\big] \\
= W_t y_t y_t' W_t - W_t y_t \mathbb{E}\big[f_t'|\Omega_T\big]\Lambda' W_t - W_t \Lambda \mathbb{E}\big[f_t|\Omega_T\big]y_t' W_t + W_t \Lambda \mathbb{E}\big[f_t f_t'|\Omega_T\big]\Lambda' W_t \\
= y_t^{(1)} y_t^{(1)'} - y_t^{(1)} \mathbb{E}\big[f_t'|\Omega_T\big]\Lambda' W_t - W_t \Lambda \mathbb{E}\big[f_t|\Omega_T\big]y_t^{(1)'} + W_t \Lambda \mathbb{E}\big[f_t f_t'|\Omega_T\big]\Lambda' W_t, \quad (21)
\end{aligned}$$

we get:

$$\begin{aligned}
\mathbb{E}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|\Omega_T\big] = \\
y_t^{(1)} y_t^{(1)'} - y_t^{(1)} \mathbb{E}\big[f_t'|\Omega_T\big]\Lambda' W_t - W_t \Lambda \mathbb{E}\big[f_t|\Omega_T\big]y_t^{(1)'} + W_t \Lambda \mathbb{E}\big[f_t f_t'|\Omega_T\big]\Lambda' W_t + (I - W_t)R(j)(I - W_t). \quad (22)
\end{aligned}$$

Inserting (22) into (20) yields:

$$\begin{aligned}
\frac{\partial \mathrm{tr}\Big\{R^{-1}\mathbb{E}_{\theta(j)}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|\Omega_T\big]\Big\}}{\partial \Lambda} &= -2W_t R^{-1} y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big] + 2W_t R^{-1} W_t \Lambda \mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big] \\
&= -2R^{-1} y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big] + 2R^{-1} W_t \Lambda \mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big]. \quad (23)
\end{aligned}$$

From

$$\sum_{t=1}^{T} \frac{\partial \mathrm{tr}\Big\{R^{-1}\mathbb{E}_{\theta(j)}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|\Omega_T\big]\Big\}}{\partial \Lambda}\Bigg|_{\Lambda = \Lambda(j+1)} = 0$$

follows

$$\sum_{t=1}^{T} y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big] = \sum_{t=1}^{T} W_t \Lambda(j+1)\mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big].$$

Equivalently (as $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$) we have

$$\mathrm{vec}\left(\sum_{t=1}^{T} y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big]\right) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big] \otimes W_t\right)\mathrm{vec}\big(\Lambda(j+1)\big),$$

hence

$$\mathrm{vec}\big(\Lambda(j+1)\big) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big] \otimes W_t\right)^{-1}\mathrm{vec}\left(\sum_{t=1}^{T} y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big]\right),$$

as given by formula (11). In the similar fashion we obtain

$$\begin{aligned}
R(j+1) &= \mathrm{diag}\Bigg(\frac{1}{T}\sum_{t=1}^{T}\Big(y_t^{(1)} y_t^{(1)'} - y_t^{(1)}\mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big]\Lambda(j+1)' W_t - W_t \Lambda(j+1)\mathbb{E}_{\theta(j)}\big[f_t|\Omega_T\big]y_t^{(1)'} \\
&\quad + W_t \Lambda(j+1)\mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big]\Lambda(j+1)' W_t + (I - W_t)R(j)(I - W_t)\Big)\Bigg).
\end{aligned}$$

Let us now consider the case of $p > 1$. We can write the log-likelihood:

$$
\begin{aligned}
l(Y, F; \theta) = & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \bar{f}_0' \Sigma^{-1} \bar{f}_0 - \frac{T}{2} \log |Q| - \frac{1}{2} \mathrm{tr} \left[Q^{-1} \sum_{t=1}^{T} (f_t - A\bar{f}_{t-1})(f_t - A\bar{f}_{t-1})' \right] \\
& - \frac{T}{2} \log |R| - \frac{1}{2} \mathrm{tr} \left[R^{-1} \sum_{t=1}^{T} (y_t - \Lambda f_t)(y_t - \Lambda f_t)' \right],
\end{aligned}
$$

where $\bar{f}_{t-1} = [f_{t-1}', \dots, f_{t-p}']'$.

Consequently (6) and (8) should be modified as:

$$
A(j+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)} \left[f_t \bar{f}_{t-1}' | \Omega_T \right] \right) \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)} \left[\bar{f}_{t-1} \bar{f}_{t-1}' | \Omega_T \right] \right)^{-1}
$$

and

$$
Q(j+1) = \frac{1}{T} \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)} \left[f_t f_t' | \Omega_T \right] - A(j+1) \sum_{t=1}^{T} \mathbb{E}_{\theta(j)} \left[\bar{f}_{t-1} f_t' | \Omega_T \right] \right).
$$

The conditional moments of the factors $\mathbb{E}_{\theta(j)} \left[f_t \bar{f}_{t-1}' | \Omega_T \right]$, $\mathbb{E}_{\theta(j)} \left[\bar{f}_{t-1} \bar{f}_{t-1}' | \Omega_T \right]$, $\mathbb{E}_{\theta(j)} \left[f_t f_t' | \Omega_T \right]$ can be obtained by running the Kalman filter on the following state space form:

$$
Y_t = \begin{bmatrix} \Lambda & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, R),
$$

$$
\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-p} \end{bmatrix} + u_t \qquad u_t \sim \mathcal{N} \left(0, \begin{bmatrix} Q & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \right).
$$

Finally let us consider the case given by (15) with the idiosyncratic component following an AR(1) process. In that case $\tilde{\theta} = \{\tilde{\Lambda}, \tilde{A}, \tilde{Q}\}$, $R \equiv \mathrm{diag}(\kappa)$ and the likelihood is given by:

$$
\begin{aligned}
l(Y, F; \tilde{\theta}) = & -\frac{1}{2} \log |\tilde{\Sigma}| - \frac{1}{2} \tilde{f}_0' \tilde{\Sigma}^{-1} \tilde{f}_0 - \frac{T}{2} \log |\tilde{Q}| - \frac{1}{2} \mathrm{tr} \left[\tilde{Q}^{-1} \sum_{t=1}^{T} (\tilde{f}_t - \tilde{A}\tilde{f}_{t-1})(\tilde{f}_t - \tilde{A}\tilde{f}_{t-1})' \right] \\
& - \frac{T}{2} \log |\tilde{R}| - \frac{1}{2} \mathrm{tr} \left[\tilde{R}^{-1} \sum_{t=1}^{T} (y_t - \tilde{\Lambda}\tilde{f}_t)(y_t - \tilde{\Lambda}\tilde{f}_t)' \right].
\end{aligned}
$$

Consequently, (21) needs to be replaced by

$$
\begin{aligned}
\mathbb{E}\left[W_t (y_t - \tilde{\Lambda}\tilde{f}_t)(y_t - \tilde{\Lambda}\tilde{f}_t)' W_t | \Omega_T \right] = & \mathbb{E}\left[W_t (y_t - \Lambda f_t - \tilde{\epsilon}_t)(y_t - \Lambda f_t - \tilde{\epsilon}_t)' W_t | \Omega_T \right] \\
= & y_t^{(1)} y_t^{(1)'} - y_t^{(1)} \mathbb{E}\left[f_t' | \Omega_T \right] \Lambda' W_t - y_t^{(1)} \mathbb{E}\left[\tilde{\epsilon}_t' | \Omega_T \right] W_t - W_t \Lambda \mathbb{E}\left[f_t | \Omega_T \right] y_t^{(1)'} + W_t \Lambda \mathbb{E}\left[f_t f_t' | \Omega_T \right] \Lambda' W_t \\
& + W_t \Lambda \mathbb{E}\left[f_t \tilde{\epsilon}_t' | \Omega_T \right] W_t - W_t \mathbb{E}\left[\tilde{\epsilon}_t | \Omega_T \right] y_t^{(1)'} + W_t \mathbb{E}\left[\tilde{\epsilon}_t f_t' | \Omega_T \right] \Lambda' W_t + W_t \mathbb{E}\left[\tilde{\epsilon}_t \tilde{\epsilon}_t' | \Omega_T \right] W_t
\end{aligned}
$$

and (23) by

$$
\begin{aligned}
\frac{\partial \mathrm{tr} \left\{ \tilde{R}^{-1} \mathbb{E}\left[(y_t - \Lambda f_t - \tilde{\epsilon}_t)(y_t - \Lambda f_t - \tilde{\epsilon}_t)' | \Omega_T \right] \right\}}{\partial \Lambda} = & -2\tilde{R}^{-1} y_t^{(1)} \mathbb{E}\left[f_t' | \Omega_T \right] - 2\tilde{R}^{-1} W_t \mathbb{E}\left[\tilde{\epsilon}_t f_t' | \Omega_T \right] \\
& + 2\tilde{R}^{-1} W_t \Lambda \mathbb{E}\left[f_t f_t' | \Omega_T \right].
\end{aligned}
$$

Hence

$$\text{vec}\big(\Lambda(j+1)\big) = \left(\sum_{t=1}^{T} \mathbb{E}_{\tilde{\theta}(j)}\big[f_t f_t'|\Omega_T\big] \otimes W_t\right)^{-1} \text{vec}\left(\sum_{t=1}^{T} W_t y_t \mathbb{E}_{\tilde{\theta}(j)}\big[f_t'|\Omega_T\big] + W_t \mathbb{E}_{\tilde{\theta}(j)}\big[\tilde{\epsilon}_t f_t'|\Omega_T\big]\right).$$

The expressions for the estimates of α_i and σ_i^2 follow in an analogous manner to those for A and Q.

C   Case of a static factor model

In the special case that $A_1 = \cdots = A_p = 0$ in (2) the model reduces to a static factor model (f_t are i.i.d.). Under the identifying assumption that $Q = I$ the joint log-likelihood can be written as:

$$l(Y, F; \theta) = -\frac{1}{2}\text{tr}\left[\sum_{t=1}^{T} f_t f_t'\right] - \frac{T}{2}\log|R| - \frac{1}{2}\text{tr}\left[R^{-1}\sum_{t=1}^{T}(y_t - \Lambda f_t)(y_t - \Lambda f_t)'\right], \tag{24}$$

where $\theta = \{\Lambda, R\}$. In a similar fashion as above, maximisation of the expected joint log-likelihood gives for the $(j+1)$-iteration

$$\Lambda(j+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[y_t f_t'|\Omega_T\big]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big]\right)^{-1},$$

$$R(j+1) = \text{diag}\left(\frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[y_t y_t'|\Omega_T\big] - \Lambda(j+1)\sum_{t=1}^{T} \mathbb{E}_{\theta(j)}\big[f_t y_t'|\Omega_T\big]\right)\right).$$

For the case of non-missing data the EM steps for the static model have been derived in Rubin and Thayer (1982). In this case we have $\mathbb{E}_{\theta(j)}\big[y_t y_t'|\Omega_T\big] = y_t y_t'$, $\mathbb{E}_{\theta(j)}\big[y_t f_t'|\Omega_T\big] = y_t \mathbb{E}_{\theta(j)}\big[f_t'|\Omega_T\big]$ and the conditional moments of the factors are given by:

$$\mathbb{E}_{\theta(j)}\big[f_t|\Omega_T\big] = \Lambda(j)'(R(j) + \Lambda(j)\Lambda(j)')^{-1}y_t = \delta(j)y_t,$$

$$\mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big] = I - \Lambda(j)'(R(j) + \Lambda(j)\Lambda(j)')^{-1}\Lambda(j) + \delta(j)y_t y_t'\delta(j)'. \tag{25}$$

In the case of missing data the reasoning from the previous section applies and the same formulas (11) and (12) for $\Lambda(j+1)$ and $R(j+1)$, respectively, can be used. $\mathbb{E}_{\theta(j)}\big[f_t|\Omega_T\big]$ and $\mathbb{E}_{\theta(j)}\big[f_t f_t'|\Omega_T\big]$ can be calculated using (25) after the rows in $\Lambda(j)$ corresponding to the missing data in y_t (and the corresponding rows and columns in $R(j)$) have been removed.

Note that this approach is different from the method proposed by Stock and Watson (2002b). The latter is a popular method based on EM to calculate principal components from a panel with missing data, see e.g. Schumacher and Breitung (2008). In fact, Stock and Watson (2002b) estimate iteratively F and Λ by minimising in step $j+1$:

$$\{F(j+1), \Lambda(j+1)\} = \arg\min_{F,\Lambda}\left\{\text{tr}\left[\sum_{t=1}^{T} \mathbb{E}_{F(j),\Lambda(j)}\big[(y_t - \Lambda f_t)(y_t - \Lambda f_t)'|\Omega_T\big]\right]\right\}.$$

This objective function is proportional to the expected log-likelihood in the case of fixed factors and homoscedastic idiosyncratic component, cf. formula (24).

D   Computation of the *news*

As in Section 2.3, let Ω_v and Ω_{v+1} be two consecutive vintages of data and let I_{v+1} be the *news* content of Ω_{v+1} orthogonal to Ω_v. We have

$$\mathbb{E}\big[y_{k,t_k}|I_{v+1}\big] = \mathbb{E}\big[y_{k,t_k}I_{v+1}'\big]\mathbb{E}\big[I_{v+1}I_{v+1}'\big]^{-1}I_{v+1}, \tag{26}$$

where

$$
\mathbb{E}\left[y_{k,t_k}I'_{v+1}\right] = \left[\begin{array}{c} \mathbb{E}\left[y_{k,t_k}\left(y_{i_1,t_1} - \mathbb{E}\left[y_{i_1,t_1}|\Omega_v\right]\right)\right] \\ \mathbb{E}\left[y_{k,t_k}\left(y_{i_2,t_2} - \mathbb{E}\left[y_{i_2,t_2}|\Omega_v\right]\right)\right] \\ \vdots \\ \mathbb{E}\left[y_{k,t_k}\left(y_{i_J,t_{J_{v+1}}} - \mathbb{E}\left[y_{i_J,t_{J_{v+1}}}|\Omega_v\right]\right)\right] \end{array}\right]'
$$

and

$$
\mathbb{E}\left[I_{v+1}I'_{v+1}\right] = \left[\begin{array}{c} \mathbb{E}\left[\left(y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right]\right)\left(y_{i_l,t_l} - \mathbb{E}\left[y_{i_l,t_l}|\Omega_v\right]\right)\right] \end{array}\right]_{\{j=1,\ldots,J_{v+1};l=1,\ldots,J_{v+1}\}}.
$$

Expressions (16) and (26) can be derived using the properties of conditional expectation as a projection operator under the assumption of Gaussian data (see e.g. Brockwell and Davis, 1991, Chapter 2).

In order to obtain (26) we need to calculate $\mathbb{E}\left[y_{k,t_k}\left(y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right]\right)\right]$ and $\mathbb{E}\left[\left(y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right]\right)\left(y_{i_l,t_l} - \mathbb{E}\left[y_{i_l,t_l}|\Omega_v\right]\right)\right]$.

Given the model (1)-(2) and (4) we can write

$$
\begin{aligned}
y_{k,t_k} &= \Lambda_{k\cdot}f_{t_k} + \epsilon_{k,t_k} \qquad \text{and} \\
I_{j,v+1} &= y_{i_j,t_j} - \mathbb{E}\left[y_{i_j,t_j}|\Omega_v\right] = \Lambda_{i_j\cdot}\left(f_{t_j} - \mathbb{E}\left[f_{t_j}|\Omega_v\right]\right) + \epsilon_{i_j,t_j}.
\end{aligned}
$$

Let us denote $\mathbb{E}\left[x_t|\Omega_v\right]$ as $x_{t|\Omega_v}$, we have:

$$
\begin{aligned}
\mathbb{E}\left[y_{k,t_k}\left(y_{i_j,t_j} - y_{i_j,t_j|\Omega_v}\right)\right] &= \Lambda_{k\cdot}\mathbb{E}\left[f_{t_k}\left(f_{t_j} - f_{t_j|\Omega_v}\right)'\right]\Lambda'_{i_j\cdot} + \mathbb{E}\left[\epsilon_{k,t_k}\left(f_{t_j} - f_{t_j|\Omega_v}\right)'\right]\Lambda'_{i_j\cdot} \\
&= \Lambda_{k\cdot}\mathbb{E}\left[\left(f_{t_k} - f_{t_k|\Omega_v}\right)\left(f_{t_j} - f_{t_j|\Omega_v}\right)'\right]\Lambda'_{i_j\cdot} \\
&\quad + \Lambda_{k\cdot}\mathbb{E}\left[f_{t_k|\Omega_v}\left(f_{t_j} - f_{t_j|\Omega_v}\right)'\right]\Lambda'_{i_j\cdot} \\
&= \Lambda_{k\cdot}\mathbb{E}\left[\left(f_{t_k} - f_{t_k|\Omega_v}\right)\left(f_{t_j} - f_{t_j|\Omega_v}\right)'\right]\Lambda'_{i_j\cdot}.
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\left(y_{i_j,t_j} - y_{i_j,t_j|\Omega_v}\right)\left(y_{i_j,t_l} - y_{i_l,t_l|\Omega_v}\right)'\right] & \\
&= \Lambda_{i_j\cdot}\mathbb{E}\left[\left(f_{t_j} - f_{t_j|\Omega_v}\right)\left(f_{t_l} - f_{t_l|\Omega_v}\right)'\right]\Lambda'_{i_l\cdot} + \mathbb{E}\left[\epsilon_{i_j,t_j}\epsilon_{i_l,t_l}\right].
\end{aligned}
$$

The last term is equal to the jth element of the diagonal of R in case $j = l$ and 0 otherwise. In the case that $t_j = t_i$ the expectation $\mathbb{E}\left[\left(f_{t_j} - f_{t_j|\Omega_v}\right)\left(f_{t_i} - f_{t_i|\Omega_v}\right)'\right]$ is readily available from the Kalman smoother output. To obtain the expectations for $t_j \neq t_l$ one can augment the vector of states by appropriate number of their lags.

E *News* vs. contributions

We will show on an example why the *news* rather than the contribution analysis as proposed in Bańbura and Rünstler (2010) (see also ECB, 2008, Chart 3) is a suitable tool for interpreting the sources of forecast revisions.

As shown in Bańbura and Rünstler (2010), the forecast of variable k at time t can be written as the sum of contributions from all the variables in the data set:

$$
y_{k,t|\Omega_v} = \sum_{i=1}^{n} C_{i,v}^{k,t},
$$

where

$$C_{i,v}^{k,t} = \sum_{s:y_{i,s}\in\Omega_v} \omega_{i,v}^{k,t}(s)y_{i,s}$$

denotes the contribution of variable i to the forecast of variable k at time t given the data set Ω_v.

Let us now assume that we forecast $y_{k,t}$ using two blocks of variables: y^1 and y^2. The forecast of $y_{k,t}$ given the data vintage Ω_v can then be written (with a slight abuse of notation) as the sum of contributions from the two blocks:

$$y_{k,t|\Omega_v} = C_{1,v}^{k,t} + C_{2,v}^{k,t}.$$

The forecast revision, i.e. the difference between the forecasts based on two consecutive vintages Ω_v and Ω_{v+1}, can be expressed in terms of changes in the contributions from the two blocks:

$$y_{k,t|\Omega_{v+1}} = y_{k,t|\Omega_v} + \Delta C_{1,v+1,v}^{k,t} + \Delta C_{2,v+1,v}^{k,t},$$

where $\Delta C_{i,v+1,v}^{k,t}$ denotes a change in the contributions from variable/block i between the vintages v and $v+1$.

To see why this representation is not so convenient for understanding the sources of forecast revisions let us assume, for simplicity, that the first block contains only one variable $y^1 = \{y_{1,s}, \ s = 1, 2, \ldots\}$ and that the difference between vintages Ω_v and Ω_{v+1} is the release of $y_{1,t}$. Expressing the forecast revision in terms of the *news*, from (17) we get:

$$y_{k,t|\Omega_{v+1}} = y_{k,t|\Omega_v} + b_{v+1,1}\underbrace{\left(y_{1,t} - y_{1,t|\Omega_v}\right)}_{news}.$$

Further, given that the forecast $y_{1,t|\Omega_v}$ can be as well expressed as the sum of the contributions from the two blocks, $C_{1,v}^{1,t}$ and $C_{2,v}^{1,t}$, we have:

$$y_{k,t|\Omega_{v+1}} = y_{k,t|\Omega_v} + b_{v+1,1}\underbrace{\left(y_{1,t} - C_{1,v}^{1,t} - C_{2,v}^{1,t}\right)}_{news} = y_{k,t|\Omega_v} + \underbrace{b_{v+1,1}(y_{1,t} - C_{1,v}^{1,t})}_{\Delta C_{1,v+1,v}^{k,t}} \underbrace{- b_{v+1,1}\cdot C_{2,v}^{1,t}}_{\Delta C_{2,v+1,v}^{k,t}}.$$

Therefore, while the release expanded only the information in block one, it led to a change in the contributions of both blocks. Moreover, if $C_{1,v}^{1,t} > y_{1,t}$ it could happen that even if $b_{v+1,1} > 0$, "positive news" in y^1 (i.e. $y_{1,t} > y_{1,t|\Omega_v}$, which is possible if $C_{2,v}^{1,t} < 0$) leads to a drop in the contributions for this variable. Therefore not much can be inferred from the sign of the change in contributions what regards "the message" from a new data release.

Let us note however that if for the forecast $y_{1,t|\Omega_v}$ only the information from block one were used, we would have $b_{v+1,1}(y_{1,t} - y_{1,t|\Omega_v}) = b_{v+1,1}(y_{1,t} - C_{1,v}^{1,t}) = \Delta C_{1,v+1,v}^{k,t}$ and we would have that the changes in contributions are equivalent to the contributions from the *news*. This is the case for e.g. bridge equation models (see e.g Rünstler and Sédillot, 2003, for the implementation of bridge equations to forecast euro area GDP).

F   State space representations in the empirical applications

We provide the details of the state space representations in the "mixed frequency" empirical applications in Section 4. Let y_t^M and $\bar{y}_t^Q$ denote the $n_M \times 1$ and $n_Q \times 1$ vectors of monthly and quarterly data, respectively. The latter have been constructed as described in Section 4.2. Further, let Λ_M and Λ_Q denote the corresponding factor loading for the monthly data, y_t^M, and the unobserved monthly growth rates of

the quarterly data, y_t^Q, respectively. We first consider the case with no serial correlation in the idiosyncratic component. Combining (1), (2), (4) and (19) with $p = 1$ results in the following state space representation:

$$
\begin{bmatrix} y_t^M \\ \bar{y}_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda_M & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & I_{n_Q} & 2I_{n_Q} & 3I_{n_Q} & 2I_{n_Q} & I_{n_Q} \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{bmatrix} + \begin{bmatrix} \varepsilon_t^M \\ \xi_t^Q \end{bmatrix}
$$

$$
\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \\ \varepsilon_{t-5}^Q \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ \varepsilon_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.
$$

Let us now consider the case with the idiosyncratic component modelled as AR(1). Let $\alpha_M = \text{diag}(\alpha_{M,1}, \ldots, \alpha$ and $\alpha_Q = \text{diag}(\alpha_{Q,1}, \ldots, \alpha_{Q,n_Q})$ collect the AR(1) coefficients of the idiosyncratic component of monthly and quarterly data. We have:

$$
\begin{bmatrix} y_t^M \\ \bar{y}_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda_M & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & 0 & I_{n_Q} & 2I_{n_Q} & 3I_{n_Q} & 2I_{n_Q} & I_{n_Q} \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^M \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{bmatrix} + \begin{bmatrix} \xi_t^M \\ \xi_t^Q \end{bmatrix}
$$

$$
\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^M \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_M & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha_Q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_{n_Q} & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \varepsilon_{t-1}^M \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \\ \varepsilon_{t-5}^Q \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ e_t^M \\ e_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.
$$

ξ_t^M and ξ_t^Q have fixed and small variance κ as discussed in Section 2.1.