

# Dynamic Factor Models

## A Very Short Introduction

Sebastian Krantz

February 24, 2023

Dynamic factor models (DFM) postulate that a small number of unobserved "factors" can be used to explain a substantial portion of the variation and dynamics in a larger number of observed time series. In addition to producing estimates of the unobserved factors, dynamic factor models have many uses in forecasting and macroeconomic monitoring. One popular application for these models is "nowcasting", in which higher-frequency data is used to produce "nowcasts" of series that are only published at a lower frequency (Chad Fulton, 2020).

Dynamic Factor Models are set up in State-Space form and can be estimated using the Kalman Filter and a number of solution algorithms, the most popular one in the economics literature being the Expectation Maximization (EM) algorithm, due to its robust numerical properties and the mixed frequency modification of Bańbura & Modugno (2014).

## 1 The Canonical (Exact) DFM

A baseline dynamic factor model can be written as follows

$$\mathbf{x}_t = \mathbf{C}_0 \mathbf{f}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \mathbf{R}) \quad (1)$$

$$\mathbf{f}_t = \sum_{j=1}^p \mathbf{A}_j \mathbf{f}_{t-j} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \mathbf{Q}_0), \quad (2)$$

where the unobserved factors  $\mathbf{f}_t$  are evolve according to a VAR(p) process and

$\mathbf{x}_t$	$n \times 1$ vector of observed series at time $t$ : $(x_{1t}, \dots, x_{nt})'$ . Some observations can be missing.
$\mathbf{f}_t$	$r \times 1$ vector of factors at time $t$ : $(f_{1t}, \dots, f_{rt})'$ .
$\mathbf{C}_0$	$n \times r$ measurement (observation) matrix.
$\mathbf{A}_j$	$r \times r$ state transition matrix at lag $j$ .
$\mathbf{Q}_0$	$r \times r$ state covariance matrix.
$\mathbf{R}$	$n \times n$ measurement (observation) covariance matrix. It is diagonal by the assumption that all covariation between the series is explained by the factors $E[\mathbf{x}_{it}   \mathbf{x}_{-i,t}, \mathbf{x}_{i,t-1}, \dots, \mathbf{f}_t, \mathbf{f}_{t-1}, \dots] = \mathbf{C} \mathbf{f}_t \forall i$ .

This model can be estimated using a classical form of the Kalman Filter and the Expectation Maximization (EM) algorithm, after transforming it to State-Space (stacked, VAR(1)) form:

$$\mathbf{x}_t = \mathbf{C} \mathbf{F}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \mathbf{R}) \quad (3)$$

$$\mathbf{F}_t = \mathbf{A} \mathbf{F}_{t-1} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \mathbf{Q}), \quad (4)$$

where  $\mathbf{x}_t$ ,  $\mathbf{e}_t$  and  $\mathbf{R}$  are as in Eq. 1, and the other matrices are

$$\mathbf{F}_t (rp \times 1) = (\mathbf{f}'_t, \mathbf{f}'_{t-1}, \dots, \mathbf{f}'_{t-p})' = (f_{1t}, \dots, f_{rt}, f_{1,t-1}, \dots, f_{r,t-1}, \dots, f_{1,t-p}, \dots, f_{r,t-p})' \quad (5)$$

$$\mathbf{C}_{(n \times rp)} = (\mathbf{C}_0, \mathbf{0}, \dots, \mathbf{0}), \text{ where } \mathbf{0} \text{ are } n \times r \text{ matrices of zeros for each factor lag} \quad (6)$$

$$\mathbf{A}_{(rp \times rp)} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{p-1} & \mathbf{0} \end{pmatrix}, \text{ where } \mathbf{0}/\mathbf{I} \text{ are } r \times r \text{ zero/identity matrices} \quad (7)$$

$$\mathbf{u}_t (rp \times 1) = (\mathbf{u}'_t, \mathbf{0}', \dots, \mathbf{0}')', \text{ with } \mathbf{0} \text{ a } r \times 1 \text{ vector of zeros} \quad (8)$$

$$\mathbf{Q}_{(rp \times rp)} = \begin{pmatrix} \mathbf{Q}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}, \text{ where } \mathbf{0} \text{ are } r \times r \text{ zero matrices.} \quad (9)$$

The estimation of this model using the EM algorithm, with initial values of the system matrices estimated through Principal Components Analysis to facilitate convergence for large  $n$ , is described in [Doz et al. \(2012\)](#). This *exact* DFM is quite restrictive as it assumes that all correlation in the data is explained by the unobserved common factors. In particular, it assumes:

1. Linearity and constant relationships (no structural breaks)
2. Idiosyncratic measurement (observation) errors ( $\mathbf{R}$  is diagonal)
3. No direct relationship between series and lagged factors (can be relaxed)
4. No relationship between lagged error terms in the either measurement or transition equation (no serial correlation)

Particularly assumption 4 is quite restrictive since it stipulates that all time dynamics in  $\mathbf{x}_t$  need to be accounted for by the factors. Within the framework of the *exact* DFM, assumption 3 can easily be relaxed to allow  $q$  lags of the factors in the measurement equation

$$\mathbf{x}_t = \sum_{i=0}^q \mathbf{C}_i \mathbf{f}_{t-i} + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \mathbf{R}). \quad (10)$$

In this case the stacked notation remains the same as long as  $q < r - 1$ , with observation matrix

$$\mathbf{C}_{(n \times rp)} = (\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_q, \mathbf{0}, \dots, \mathbf{0}) \quad (11)$$

modified to estimate the lagged loadings  $\mathbf{C}_i$ . However, because of the proliferation of parameters, this extension has not received much attention. Furthermore, in the presence of significant lagged dynamics, increasing the number of factors  $r$  is often successful in capturing these dynamics, with certain factors loading strongly on the lagged indicators and others on the contemporaneous ones.

In the current practice of estimating large DFMs of economic time series, series for a given sector often have unmodeled sector-specific dynamics. Most of the economics literature on DFMs has thus focused on relaxing restrictions about the error structure in factor models, introducing the notion of *approximate* factor models ([Stock & Watson, 2016](#)). A particular emphasis has been placed on relaxing assumption 4, i.e. allowing some of the dynamics of the time series to be unexplained by the common factors.

## 2 Approximate DFMs

The most common form of *approximate* DFM, introduced by Chamberlain & Rothschild (1983), allows for the observation errors  $\mathbf{e}_t$  to evolve according to an autoregressive AR(1) process

$$\mathbf{e}_t = \Phi \mathbf{e}_{t-1} + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}), \quad (12)$$

where  $\Phi$  is diagonal  $n \times n$  with elements  $\rho_i$ . Following Bańbura & Modugno (2014) this can be modelled as part of the state vector. In the stacked form, this model becomes

$$\mathbf{x}_t = \mathbf{C}^a \mathbf{F}_t^a \quad (13)$$

$$\mathbf{F}_t^a = \mathbf{A}^a \mathbf{F}_{t-1}^a + \mathbf{u}_t^a, \quad \mathbf{u}_t^a \sim N(\mathbf{0}, \mathbf{Q}^a), \quad (14)$$

where  $\mathbf{x}_t$  is as in Eq. 1, and the other matrices are

$$\mathbf{F}_t^a (rp+n \times 1) = (\mathbf{f}'_t, \mathbf{f}'_{t-1}, \dots, \mathbf{f}'_{t-p}, \mathbf{e}'_t)', \text{ where } \mathbf{e}_t \text{ is } n \times 1 \text{ as in Eq. 12} \quad (15)$$

$$\mathbf{C}_{(n \times rp+n)}^a = (\mathbf{C}_0, \mathbf{0}, \dots, \mathbf{0}, \mathbf{I}), \text{ where } \mathbf{0} \text{ is } n \times r \text{ and } \mathbf{I} \text{ is } n \times n \quad (16)$$

$$\mathbf{A}_{(rp+n \times rp+n)}^a = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \Phi \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p & \mathbf{0} \\ \mathbf{I}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{p-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \Phi \end{pmatrix}, \text{ where } \Phi \text{ is } n \times n \text{ as in Eq. 12} \quad (17)$$

$$\mathbf{u}_t^a (rp+n \times 1) = (\mathbf{u}'_t, \mathbf{0}', \dots, \mathbf{0}', \mathbf{v}'_t)', \text{ with } \mathbf{v}_t \text{ } n \times 1 \text{ as in Eq. 12} \quad (18)$$

$$\mathbf{Q}_{(rp+n \times rp+n)}^a = \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R} \end{pmatrix}, \text{ where } \mathbf{R} \text{ is } n \times n \text{ as in Eq. 12.} \quad (19)$$

To still be able estimate this model using a classical form of the Kalman Filter, Bańbura & Modugno (2014) introduce an error term  $\mathbf{z}_t \sim N(\mathbf{0}, \kappa \mathbf{I})$  in Eq. 13, with  $\kappa$  a very small number.

Like the *exact* DFM, this model can also easily be extended to allow for  $q$  factor lags in the observation equation, by estimating an observation matrix

$$\mathbf{C}_{(n \times rp+n)}^a = (\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_q, \mathbf{0}, \dots, \mathbf{0}, \mathbf{I}). \quad (20)$$

The same comments given above about proliferation of parameters and the possibility to include additional factors to account for lagged dynamics apply.

In general, Doz et al. (2011, 2012) show that in the presence of serial correlation, the factors are still consistently estimated as  $n, T \rightarrow \infty$  using an *exact* DFM specification. Thus modelling serial correlation is particularly important in smaller samples. Bańbura & Modugno (2014) also show that modelling the serial correlation improves forecast accuracy, but only at short horizons.

### 3 Mixed Frequency

Building on the work of [Mariano & Murasawa \(2003\)](#), [Bańbura & Modugno \(2014\)](#) and [Bok et al. \(2018\)](#) (New York FED Nowcast) popularized EM based estimation of large mixed-frequency (monthly and quarterly) DFMs as workhorse models in economic nowcasting practice. The key addition of [Mariano & Murasawa \(2003\)](#) was to model observed quarterly series by unobserved monthly counterparts, and then place appropriate restrictions on the observation matrix ( $\mathbf{C}$ ).

In particular, [Mariano & Murasawa \(2003\)](#) consider the case of an observed quarterly series  $X_t^q$  being the geometric mean of an unobserved monthly series  $\tilde{X}_t^m$  and its lags, i.e.

$$X_t^q = \tilde{X}_t^{m\frac{1}{3}} \tilde{X}_{t-1}^{m\frac{1}{3}} \tilde{X}_{t-2}^{m\frac{1}{3}}, \quad (21)$$

such that taking the natural log yields

$$\log(X_t^q) = \frac{1}{3} \log(\tilde{X}_t^m) + \frac{1}{3} \log(\tilde{X}_{t-1}^m) + \frac{1}{3} \log(\tilde{X}_{t-2}^m). \quad (22)$$

Lagging this equation 3 times and subtracting it from itself yields the (approximate) quarterly growth rate of the quarterly series

$$\log(X_t^q) - \log(X_{t-3}^q) = \frac{1}{3} [\log(\tilde{X}_t^m) - \log(\tilde{X}_{t-3}^m)] + \frac{1}{3} [\log(\tilde{X}_{t-1}^m) - \log(\tilde{X}_{t-4}^m)] + \frac{1}{3} [\log(\tilde{X}_{t-2}^m) - \log(\tilde{X}_{t-5}^m)]. \quad (23)$$

Adding and subtracting further lags and leads on the right hand side, and denoting the growth rate by lower case letters, i.e.  $x_t^q = \log(X_t^q) - \log(X_{t-3}^q)$  and  $\tilde{x}_t^m = \log(\tilde{X}_t^m) - \log(\tilde{X}_{t-1}^m)$ , yields

$$x_t^q = \frac{1}{3} [\tilde{x}_t^m + \tilde{x}_{t-1}^m + \tilde{x}_{t-2}^m] + \frac{1}{3} [\tilde{x}_{t-1}^m + \tilde{x}_{t-2}^m + \tilde{x}_{t-3}^m] + \frac{1}{3} [\tilde{x}_{t-2}^m + \tilde{x}_{t-3}^m + \tilde{x}_{t-4}^m] \quad (24)$$

$$= \frac{1}{3} \tilde{x}_t^m + \frac{2}{3} \tilde{x}_{t-1}^m + \tilde{x}_{t-2}^m + \frac{2}{3} \tilde{x}_{t-3}^m + \frac{1}{3} \tilde{x}_{t-4}^m. \quad (25)$$

This is the result of [Mariano & Murasawa \(2003\)](#). [Bańbura & Modugno \(2014\)](#) consider instead the quarterly series to be the product of the unobserved monthly series, i.e. starting from  $X_t^q = \tilde{X}_t^m \tilde{X}_{t-1}^m \tilde{X}_{t-2}^m$ , the final expression is

$$x_t^q = \tilde{x}_t^m + 2\tilde{x}_{t-1}^m + 3\tilde{x}_{t-2}^m + 2\tilde{x}_{t-3}^m + \tilde{x}_{t-4}^m. \quad (26)$$

Since  $\tilde{X}_t^m$  (and thus  $\tilde{x}_t^m$ ) is unobserved, this approach is equivalent, as only the relative weights on the lags of the unobserved series matter. We now assume that there exist  $n_Q$  quarterly variables, whose vector of unobserved monthly counterparts  $\tilde{\mathbf{x}}_t^m$  admits the same DFM representation as the observed monthly variables, i.e.  $\tilde{\mathbf{x}}_t^m = \mathbf{C}_0^q \mathbf{f}_t + \mathbf{e}_t^q$ . Inserting in a vectorized version of Eq. 26 yields

$$\mathbf{x}_t^q = \mathbf{C}_0^q \mathbf{f}_t + \mathbf{e}_t^q + 2(\mathbf{C}_0^q \mathbf{f}_{t-1} + \mathbf{e}_{t-1}^q) + 3(\mathbf{C}_0^q \mathbf{f}_{t-2} + \mathbf{e}_{t-2}^q) + 2(\mathbf{C}_0^q \mathbf{f}_{t-3} + \mathbf{e}_{t-3}^q) + \mathbf{C}_0^q \mathbf{f}_{t-4} + \mathbf{e}_{t-4}^q \quad (27)$$

$$= \mathbf{C}_0^q (\mathbf{I} \ 2\mathbf{I} \ 3\mathbf{I} \ 2\mathbf{I} \ \mathbf{I}) (\mathbf{f}_t' \ \mathbf{f}_{t-1}' \ \mathbf{f}_{t-2}' \ \mathbf{f}_{t-3}' \ \mathbf{f}_{t-4}')' + (\mathbf{I} \ 2\mathbf{I} \ 3\mathbf{I} \ 2\mathbf{I} \ \mathbf{I}) (\mathbf{e}_t^{q'} \ \mathbf{e}_{t-1}^{q'} \ \mathbf{e}_{t-2}^{q'} \ \mathbf{e}_{t-3}^{q'} \ \mathbf{e}_{t-4}^{q'})' \quad (28)$$

$$= \mathbf{C}_0^q \mathbf{J} \mathbf{F}_t^5 + \mathbf{J} \mathbf{E}_t^5 \quad (29)$$

$$= \mathbf{C}^q \mathbf{F}_t^q, \quad (30)$$

where the latter is a stacked representation similar to the *approximate* DFM case, with matrices

$$\mathbf{F}_t^q (5r+5n_Q \times 1) = (\mathbf{F}_t^{5'}, \mathbf{E}_t^{5'})' = (\mathbf{f}_t', \dots, \mathbf{f}_{t-4}', \mathbf{e}_t^{q'}, \dots, \mathbf{e}_{t-4}^{q'})', \text{ where } \mathbf{f}_t \text{ is } r \times 1 \text{ and } \mathbf{e}_t^q \text{ is } n_Q \times 1 \quad (31)$$

$$\mathbf{C}_{(n \times 5r+5n_Q)}^q = (\mathbf{C}_0^q \mathbf{J}, \mathbf{J}) = (\mathbf{C}_0^q, 2\mathbf{C}_0^q, 3\mathbf{C}_0^q, 2\mathbf{C}_0^q, \mathbf{C}_0^q, \mathbf{I}, 2\mathbf{I}, 3\mathbf{I}, 2\mathbf{I}, \mathbf{I}), \text{ where } \mathbf{I} \text{ is } n_Q \times n_Q. \quad (32)$$

In other words, mixed frequency estimation amounts to modelling the error process of unobserved variables as part of the state vector and placing restrictions on the observation matrix. Provided that the number of lags in the transition equation is  $p > 4$ , and assuming that quarterly series  $\mathbf{x}_t^q$  are observed every 3rd month and missing otherwise, the joint monthly-quarterly state space representation can be modelled as

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{C}} \tilde{\mathbf{F}}_t + \tilde{\mathbf{e}}_t, \quad \tilde{\mathbf{e}}_t \sim N(\mathbf{0}, \tilde{\mathbf{R}}) \quad (33)$$

$$\tilde{\mathbf{F}}_t = \tilde{\mathbf{A}} \tilde{\mathbf{F}}_{t-1} + \tilde{\mathbf{u}}_t, \quad \tilde{\mathbf{u}}_t \sim N(\mathbf{0}, \tilde{\mathbf{Q}}), \quad (34)$$

with system matrices

$$\tilde{\mathbf{x}}_t \text{ (} n \times 1 \text{)} = (\mathbf{x}_t^{m'}, \mathbf{x}_t^{q'})', \text{ where } \mathbf{x}_t^m \text{ is } n_M \times 1 \text{ and } \mathbf{x}_t^q \text{ is } n_Q \times 1 \quad (35)$$

$$\tilde{\mathbf{F}}_t \text{ (} rp+5n_Q \times 1 \text{)} = (\mathbf{f}_t', \dots, \mathbf{f}_{t-p}', \mathbf{e}_t^{q'}, \dots, \mathbf{e}_{t-4}^{q'})', \text{ where } \mathbf{e}_t^q \text{ is } n_Q \times 1 \text{ as in Eq. 31} \quad (36)$$

$$\tilde{\mathbf{C}}_{(n \times rp+5n_Q)} = \begin{pmatrix} \mathbf{C}_0^m & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_0^q \mathbf{J} & \mathbf{0} & \mathbf{J} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_0^m & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_0^q & 2\mathbf{C}_0^q & 3\mathbf{C}_0^q & 2\mathbf{C}_0^q & \mathbf{C}_0^q & \mathbf{0} & \mathbf{I} & 2\mathbf{I} & 3\mathbf{I} & 2\mathbf{I} & \mathbf{I} \end{pmatrix} \quad (37)$$

$$\tilde{\mathbf{e}}_t \text{ (} n \times 1 \text{)} = (\mathbf{e}_t^{m'}, \mathbf{z}_t^{q'})', \text{ where } \mathbf{e}_t^m \text{ is } n_M \times 1, \text{ and } \mathbf{z}_t^q \text{ is } n_Q \times 1 \text{ and very small} \quad (38)$$

$$\tilde{\mathbf{R}}_{(n \times n)} = \begin{pmatrix} \mathbf{R}^m & \mathbf{0} \\ \mathbf{0} & \kappa \mathbf{I} \end{pmatrix}, \text{ where } \mathbf{R}^m \text{ is } n_M \times n_M, \mathbf{I} \text{ is } n_Q \times n_Q, \text{ and } \kappa \text{ very small} \quad (39)$$

$$\tilde{\mathbf{A}}_{(rp+5n_Q \times rp+5n_Q)} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^q \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p & \mathbf{0} \\ \mathbf{I}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{p-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{I}^q \end{pmatrix}, \text{ where } \mathbf{I}^q \text{ is } 5n_Q \times 5n_Q \quad (40)$$

$$\tilde{\mathbf{u}}_t \text{ (} rp+5n_Q \times 1 \text{)} = (\mathbf{u}_t', \mathbf{0}', \dots, \mathbf{0}', \mathbf{e}_t^{q'}, \mathbf{0}', \dots, \mathbf{0}')', \text{ where } \mathbf{e}_t^q \text{ is } n_Q \times 1 \quad (41)$$

$$\tilde{\mathbf{Q}}_{(rp+5n_Q \times rp+5n_Q)} = \begin{pmatrix} \mathbf{Q}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}^q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \text{ where } \mathbf{R}^q \text{ is } n_Q \times n_Q. \quad (42)$$

This representation can easily be extended to allow additional lags of monthly variables in the observation equation, by replacing  $\mathbf{C}^m$  in Eq. 37 by a suitably altered version as in Eq. 11. It is also possible to estimate autoregressive parameters, by also including  $\mathbf{e}_t^m$  into the state vector  $\tilde{\mathbf{F}}$ , replacing  $\tilde{\mathbf{e}}_t$  by a very small amplitude process  $\tilde{\mathbf{z}}_t$  with covariance  $\kappa \mathbf{I}$ , adding autoregressive parameters  $\Phi^m$  and  $\Phi^q$  to  $\tilde{\mathbf{A}}$ , and estimating both  $\mathbf{R}^m$  and  $\mathbf{R}^q$  inside  $\tilde{\mathbf{Q}}$ . The modified EM algorithm that respects the restrictions placed on  $\tilde{\mathbf{C}}$  is detailed in Bańbura & Modugno (2014).

## References

- Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of applied econometrics*, 29(1), 133–160.
- Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10, 615–643.
- Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5), 1281–1304. Retrieved 2023-02-23, from <http://www.jstor.org/stable/1912275>
- Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1), 188–205.
- Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of economics and statistics*, 94(4), 1014–1024.
- Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of applied Econometrics*, 18(4), 427–443.
- Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415–525). Elsevier.