Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-15 (2018) 1008–1013

Identification of Generalized Dynamic Factor Models from mixed-frequency data

B. D. O. Anderson * **Alexander Braumann** ** **Manfred Deistler** ***

* *Hangzhou Dianzi University, China, Australian National University, College of Engineering and Computer Science, Australia, and Data61-CSIRO, Australia. (email: brian.anderson@anu.edu.au)*
** *Institut für Mathematische Stochastik, TU Braunschweig, (e-mail: a.braumann@tu-bs.de)*
*** *Vienna University of Technology, Institute of Statistics and Mathematical Methods in Economics (e-mail: manfred.deistler@tuwien.ac.at)*

Abstract: Modeling of high dimensional time series by linear time series models such as vector autoregressive models is often marred by the so-called "curse of dimensionality". In order to overcome this problem generalized linear dynamic factor models (GDFM's) maybe used. In high-dimensional time series the single univariate time series are often sampled at different frequencies. This is the so-called mixed-frequency situation. We consider identifiability of the underlying high-frequency GDFM (i.e. the GDFM generating the data at the highest sampling frequency occurring) in the case of mixed frequency data and we shortly describe two estimation procedures in this situation based on the EM algorithm.

Keywords: time series modeling, high-dimensional time series, factor models, mixed frequency

1. INTRODUCTION

"Traditional" modeling of multivariate time series by vector autoregressive (VAR), VARMA and linear state space (SS) models has reached a certain stage of maturity now, see e.g. Caines (1988), Hannan and Deistler (2012), Ljung (1987), Lütkepohl (2005) and Reinsel (1993). Modeling of multivariate time series is important for e.g.

- the analysis of the dynamic relations between univariate component series
- the improvement of forecasts by including additional variables.

As is well known, traditional multivariate time series analysis is plagued by the so called "curse of dimensionality". E.g. in the unrestricted VAR case, where the cross-sectional dimension (i.e. the output dimension) is N and the AR order is p, the parameter space is of dimension $N^2p + \frac{1}{2}N(N+1)$. Thus the dimension of the parameter space is proportional to N^2 whereas the amount of data $N \cdot T$ (here T is sample size) is linear in N. In a number of applications, such as macroeconomics, high density EEG's or "over-sensoring" cases in engineering applications, we have to analyse, say, 150 or 500 single time series jointly.

In such cases one has either to search for additional structure, as in dynamic principal component models, dynamic factor models or dynamic networks in order to reduce the dimension of the parameter space or one has to apply "highly regularized" estimation procedures. Here we consider "generalized linear dynamic factor models (GDFM's)" as introduced by Forni et al. (2000) and Forni and Lippi (2001), see also Stock and Watson (2002).

Such models are an outgrowth of models considered earlier in the literature. Firstly, they are an outgrowth of strictly idiosyncratic linear dynamic factor models (Geweke (1977), Scherrer and Deistler (1998)). These are models where the spectral density of the noise is assumed to be diagonal. Secondly they are also an outgrowth of a different form of model, namely generalized linear static factor models (Chamberlain (1983), Chamberlain and Rothschild (1983)), where the noise components are allowed to be weakly correlated rather than uncorrelated.

In many cases, in high dimensional time series, the single component series have been sampled at "different frequencies", e.g. quarterly and yearly in macroeconomic applications. This is the so called mixed-frequency (MF) case. In this case, identifiability and estimation requires further analysis (see e.g. Anderson et al. (2016a)).

The main contribution in this paper is to show how parameter estimation of GDFM models in the MF case can be achieved with EM methods.

2. GENERALIZED LINEAR DYNAMIC FACTOR MODELS: BACKGROUND

We consider GDFM's, where the N-dimensional vector of observations at time t, y_t^N can be represented as

$$y_t^N = \hat{y}_t^N + u_t^N \tag{1}$$

* Brian.Anderson@anu.edu.au; his work is supported by Data-61, CSIRO, and by the Australian Research Council's Discovery Project DP-160104500.

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.09.062

Here $(\hat{y}_t^N)$ is the process of latent variables[1], whose univariate component series who show strong comovement, i.e. they "move together" (for a more detailed explanation see below), and (u_t^N) is the noise process which shows weak dependency across the univariate component series. The precise meaning of these words is given by the assumptions below. The basic idea of representing (y_t^N) as in (1) (as in linear factor models in general) is to separate the comovements between the components of (y_t^N) from the individual fluctuations in the single components. Factor models - not necessarily in a time series setting - and the closely related errors-in-variables models, have a long history, dating back in the static case to the early 20th century. Determining intelligence factors from scores in intelligence tests seem to have been the first applications.

For "classical" factor models, the univariate noise components, u_{it} say, are assumed to be mutually uncorrelated. For so called generalized dynamic factor models, both for the static case (Chamberlain and Rothschild (1983)) and for the dynamic case (Forni et al. (2000)) this assumption is relaxed by postulating only weak dependence as defined below.

Throughout we assume the following (compare Deistler et al. (2010)):

$$\mathbb{E}\hat{y}_t^N = \mathbb{E}u_t^N = 0 \tag{2}$$

$$\mathbb{E}\hat{y}_t^N \left(u_s^N\right)' = 0, \quad \forall s, t \in \mathbb{Z} \tag{3}$$

and that $(\hat{y}_t^N)$ and (u_t^N) are wide sense stationary with absolutely summable covariances. Thus, using an obvious notation for the spectral densities corresponding to (1), we obtain

$$f_y^N(\lambda) = f_{\hat{y}}^N(\lambda) + f_u^N(\lambda) \tag{4}$$

The following additional (but typical) assumptions constitute the class of GDFM's considered here:

Assumption 1 There is an N_0 such that for all $N \geq N_0$, $f_{\hat{y}}^N$ is a rational spectral density with constant rank $q < N$ on $[-\pi, \pi]$.

It is really this assumption that underpins our earlier statement that there is strong comovement among the components of $\hat{y}_t^N$. In fact, there will be a white noise process of dimension q and $N \times q$ stable transfer function such that the output of the transfer function when driven by the noise would have spectrum $f_{\hat{y}}$. Under Assumption 1 $f_{\hat{y}}^N$ is singular for N large enough and its non-trivial left kernel describes the comovement between the univariate component processes.

For the asymptotic analysis considered here, not only the sample size T but also the cross sectional dimension N tends to infinity. Thus we consider a doubly–indexed stochastic process $(y_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$ where i is the cross sectional index and t denotes time.

We assume:

Assumption 2 The doubly-indexed sequence $(y_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$ corresponds to a nested sequence of models (1), in the sense that $\hat{y}_{it}$ and u_{it} do not depend on N for $i \leq N$.

[1] The term latent variable is used for variables which can not be directly observed and are of independent interest.

Assumption 3 The rank q of $f_{\hat{y}}^N$ is independent of N (for $N \geq N_0$).

Assumption 4 The state dimension, n say, of a minimal state space realization of a stable and miniphase spectral factor of $f_{\hat{y}}^N$ is independent of N (for $N \geq N_0$).

The four assumptions imply that we finally get an infinite set of measurements of a finite-dimensional system which has q inputs only. Looking ahead, we shall show how to achieve a form of averaging across the output components, as opposed to with time, to achieve goals such as identification.

Next we define weak and strong dependence as in Forni and Lippi (2001). We use e.g. $w_{u,r}^N$ to denote the r-th largest (dynamic) eigenvalue of f_u^N.

Assumption 5 (Weak dependence) $w_{u,1}^N$ is uniformly bounded in frequency λ and in N.

Assumption 6 (Strong dependence) The first q (i.e. the q largest) eigenvalues of $f_{\hat{y}}^N(\lambda)$ diverge to infinity for all frequencies λ as $N \to \infty$.

Assumptions 5 and 6, which are common in econometric literature, in effect suggest that taking all outputs together gives a system with outputs whose signal-to-noise ratio approaches infinity (blessing rather than curse of dimensionality).

There is a "theoretical" construction for finding a GDFM from observations in the Hilbert space L_2 of square integrable random functions over the underlying probability space (see Forni and Lippi (2001), see also Deistler et al. (2015)). There, based on Assumptions 5 and 6, it is shown, that the latent variables $\hat{y}_t^N$ can be obtained from the observations y_t^N by averaging out the influence of the noise u_t^N, for $N \to \infty$, by using certain dynamic linear transformations. By Assumption 1 there exists a Wold representation for $(\hat{y}_t^N)$ of the form

$$\hat{y}_t^N = \sum_{j=0}^{\infty} w_j^N \varepsilon_{t-j}, \quad w_j^N \in \mathbb{R}^{N \times q}, \tag{5}$$

where (ε_t) is a white noise innovation process with $\mathbb{E}\varepsilon_t\varepsilon_t' = 2\pi I_q$, which is independent of N (from a certain N_0 onwards). Here we will assume that $N > q$ holds. The elements of the process (ε_t) are called minimal dynamic factors. Let

$$w^N(z) = \sum_{j=0}^{\infty} w_j^N z^j \tag{6}$$

denote the transfer function corresponding to the Wold decomposition (5), where z is used to denote the complex variable as well as the backward shift on the integers.

The Smith-McMillan form of $w(z)$ (see Hannan and Deistler (2012)) we sometimes drop the superscript N for notational convenience) is given by

$$w = u\, d\, v \tag{7}$$

where u and v are unimodular polynomial matrices (i.e. have a constant nonzero determinant) and d is a $N \times q$ rational matrix with diagonal elements $\frac{n_i}{d_i}$ where d_i and n_i are coprime, monic polynomials and d_{i+1} divides d_i and n_i divides n_{i+1}. All other elements of d are zero. The matrix

d is unique for given w and the (finite) zeros of w are the zeros of the n_i and the poles of w are the zeros of the d_i. Note that, as $w(z)$ corresponds to the Wold decomposition, $w(z)$ has no poles for $|z| \leq 1$ and no zeros for $|z| < 1$. We in addition assume that $w(z)$ has also no zeros for $|z| = 1$.

As the spectral density $f_{\hat{y}}^N$ of the latent process $\hat{y}_t^N$ is rational, it can be represented by a VARMA as well as by a state space system. Due to the fact that the rank of the spectral density $f_{\hat{y}}^N$ is $q < N$, in general the corresponding VARMA or state space systems will be singular in the sense that the innovation covariance matrix is of rank q.

The transfer function w can be written as $w(z) = c^{-1}(z)d(z)$ where (c, d) are left coprime polynomial matrices, which can be shown to be unique up to unimodular left multiplication (see Hannan and Deistler (2012)). This gives a VARMA system

$$c(z)\hat{y}_t = d(z)\varepsilon_t \tag{8}$$

The conditions on the zeros and poles of the transfer $w(z)$ are equivalent to $\det(c(z)) \neq 0$, $|z| \leq 1$ and $d(z)$ has full rank q for all $|z| \leq 1$.

Alternatively $\hat{y}_t^N$ can be represented by state space system

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{9}$$

$$\hat{y}_t = Hx_t \tag{10}$$

Here x_t is the, n-dimensional say, state and $F \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{n \times q}$, $H \in \mathbb{R}^{N \times n}$ are parameter matrices; they are *not* uniquely defined by $w(z)$. Note that the SS form used here has no innovations term in (10), which allows a static factor (see below) interpretation for the state (compare Deistler et al. (2010)).

We assume that the SS system is minimal, stable and strictly miniphase. The transfer function (6) then can be written as

$$w(z) = H(I - Fz)^{-1}G = HG + \sum_{j=1}^{\infty} HF^j Gz^j. \tag{11}$$

A *static factor* process of the latent variables $(\hat{y}_t)$ is a process (z_t) of dimension less than or equal to n, with the property that there exists a constant matrix L such that

$$\hat{y}_t = Lz_t \tag{12}$$

holds. A *minimal static factor* is a static factor of minimal dimension. As easily can be seen (see Deistler et al. (2010)), the dimension, r say, of a minimal static factor is equal to the rank of the zero-lag variance matrix $\mathbb{E}\hat{y}_t\hat{y}_t'$. If we write

$$\mathbb{E}\hat{y}_t\hat{y}_t' = MM', \quad M \in \mathbb{R}^{N \times r},$$

then

$$z_t = (M'M)^{-1}M'\hat{y}_t. \tag{13}$$

A minimal static factor is unique up to premultiplication by a constant non-singular matrix.

Any minimal static factor is obtainable by a simple static linear transformation of the latent variables $\hat{y}_t$ and vice versa, i.e.

$$z_t = (M'M)^{-1}Mw(z)\varepsilon_t = k(z)\varepsilon_t \tag{14}$$

$$\hat{y}_t = Mk(z)\varepsilon_t = w(z)\varepsilon_t \tag{15}$$

for some transfer function $k(z)$ corresponding to the choice of the static factors. The transfer functions $k(z)$ and $w(z)$ respectively have no zeros and poles for all $|z| \leq 1$. It is easy to show that a minimal state space realization for (z_t) is obtained from (F, G, H) in (9)-(10) of the form (F, G, C) given by

$$C = (M'M)^{-1}M'H$$

$$H = MC$$

and vice versa; in particular the state can be chosen the same in both SS representations (see Deistler et al. (2010)).

In many cases, the dimension q of the minimal dynamic factors ε_t is strictly smaller then the dimension r of the static factors. In this case, also $k(z)$ is a tall matrix. As is stated below, in this case, we may restrict ourselves to VAR representations for (z_t).

Here a $r \times q$ transfer function $k(z)$ is called *zeroless*, if the numerator polynomials in the diagonal matrix in (7) are all equal to one.

As has been shown in Anderson and Deistler (2008), Anderson et al. (2016b), for $r > q$, zeroless transfer function are generic, i.e. the transfer functions $k(z)$ are zeroless for generic values of (F, G, C) (generic means that this holds on a set containing an open and dense subset), and zeroless transfer functions can be realized by VAR systems, e.g.

$$z_t = A_1 z_{t-1} + \cdots + A_p z_{t-p} + v_t \tag{16}$$

where (v_t) is an innovation process, with covariance matrix Σ_v, which might be singular.

3. ESTIMATION OF GDFMS FOR THE MF CASE

In this section for notational convenience we may omit the superscript N.

3.1 Mixed-frequency data

Here we partition the process y_t as follows

$$y_t = \begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix}, \tag{17}$$

where the fast, N_f-dimensional say, component y_t^f is observed for every $t \in \mathbb{Z}$ and the slow, N_s-dimensional, component y_t^s is observed for every $t \in M\mathbb{Z}$. This corresponds to an observation scheme which is suited for so-called stock variables. For the case of so-called flow variables, the observed slow variables are sums. The latter case is not considered here [2]. Here w.l.o.g we focus on the case $M = 2$.

The population second moments of (y_t) which can be directly observed are:

$$\gamma^{ff}(h) = \mathbb{E}y_{t+h}^f(y_t^f)', \quad h \in \mathbb{Z} \tag{18}$$

$$\gamma^{fs}(h) = \mathbb{E}y_{t+h}^f(y_t^s)', \quad h \in \mathbb{Z} \tag{19}$$

$$\gamma^{ss}(h) = \mathbb{E}y_{t+h}^s(y_t^s)', \quad h \in 2\mathbb{Z} \tag{20}$$

Of course only the observed outputs can be used for estimation.

[2] A flow variable is related to a period of time whereas a stock indicates a quantity of a variable at a point of time. Examples of stocks are wealth and capital stock. Examples of flows are income and investment.

The analysis of identifiability is based on those second moments which can be directly observed, where N_f and N_s tend to infinity.

3.2 Estimation with denoising via PCA

In the high frequency case, i.e. when all components of y_t are observed for every $t \in \mathbb{Z}$, the following denoising procedure has been proposed by Stock and Watson (2002): Consider the eigenvalue composition

$$T^{-1}\sum_{t=1}^{T} y_t y_t' = O_1\Lambda_1 O_1' + O_2\Lambda_2 O_2', \tag{21}$$

where Λ_1 is the diagonal $r \times r$ matrix with its diagonal elements being the r largest eigenvalues of the matrix on the left hand side, and O_1 is the $N \times r$ matrix of the corresponding eigenvectors. The matrices Λ_2 and O_2 are defined accordingly.

Then we define an estimator for the static factors,

$$\hat{z}_t = N^{-1/2} O_1' y_t, \tag{22}$$

and a corresponding estimator for the latent variables,

$$\hat{\hat{y}}_t = N^{1/2} O_1 \hat{z}_t = O_1 O_1' y_t. \tag{23}$$

This procedure is shown to provide consistent estimates of the latent variables $\hat{y}_t$ for T and N going to infinity, and in a certain sense mimics the Hilbert space construction described in Forni and Lippi (2001).

The dynamics of the static factors then are modeled by a possibly singular autoregression on the estimated static factors. Singularity occurs in the case where the dimension of the minimal static factors r is larger then the dimension of the minimal dynamic factors q. Estimation is performed by using the Yule-Walker equations (see Deistler et al. (2010)).

We extend this procedure to the mixed frequency case as follows (see Felsenstein (2014)):

(1) Denoising step: Our aim here is to estimate a minimal mixed-frequency static factor with a maximum number of fast components. We estimate $\gamma(0) = \mathbb{E} y_t y_t'$ by

$$\hat{\gamma}(0) = \frac{2}{T}\sum_{t=1}^{T/2} y_{2t} y_{2t}' = \begin{pmatrix} \hat{\gamma}^{ff}(0) & \hat{\gamma}^{fs}(0) \\ \hat{\gamma}^{sf}(0) & \hat{\gamma}^{ss}(0) \end{pmatrix}. \tag{24}$$

Clearly $\hat{\gamma}(0)$ is a consistent estimator for $\gamma(0)$ (under suitable assumptions). Nevertheless we see from (22) that we do not get a high frequency static factor of y_t in general. Thus we determine minimal static factors for y_t^f and y_t^s, say z_t^f, $t \in \mathbb{Z}$, and z_t^s, $t \in 2\mathbb{Z}$, respectively, separately by PCA. Note that $\begin{pmatrix} z_t^f \\ z_t^s \end{pmatrix}$ is not necessarily a minimal static factor for y_t, because some components in this vector may be linearly dependent.

In order to achieve the desired minimality we proceed as follows:

Note that by construction all elements of z_t^f are linearly independent. Now we apply a sequence of rank tests (see e.g. Al-Sadoon (2015)) as follows: We add the first element of z_t^s to the vector z_t^f and we test the corresponding covariance matrix for rank deficiency. If the null hypothesis is rejected the first element of z_t^s is added to z_t^f, otherwise it is deleted. In the next step this procedure is repeated for this newly obtained vector and the second element of z_t^s, and so on. In this way we obtain an estimate of the minimal mixed frequency static factor.

(2) Estimation of the dynamics of the mixed frequency static factors: Commencing from the mixed frequency estimates of the mixed frequency static factors, we estimate the parameters of a high frequency VAR system for the static factors as given in (16). As has been shown in Anderson et al. (2016a) in this situation we have generic identifiability, i.e. identifiability on a subset containing an open and dense subset of the parameter space. As far as estimation is concerned, in a certain sense a Gaussian maximum likelihood type procedure is appropriate here. We apply an EM algorithm (Dempster et al. (1977)). The basic idea of the EM algorithm is to commence from the likelihood of the so called complete data and to condition this likelihood on the actually observed data. In a second step this conditioned likelihood is optimized and then the procedure is iterated. The application of the EM algorithm for state space models has been given in Shumway and Stoffer (2016) can be described as follows:

The Gaussian log-likelihood for the complete data for the model in Equation (16) is, ignoring constants, of the form

$$L_T(\tau) = \log(\det(V_1)) + x_1^T V_1^{-1} x_1 + T\,\log(\det(\Sigma_\nu)) + \sum_{t=1}^{T}(z_t - Ax_t)^T \Sigma_\nu^{-1}(z_t - Ax_t), \tag{25}$$

where $x_t = (z_{t-1}', \ldots, z_{t-p}')'$, $A = (A_1, \ldots, A_p)$, $x_1 \sim \mathcal{N}(0, V_1)$, $\tau = \left(\mathrm{vec}\,(A)^T, \mathrm{vech}\,(\Sigma_\nu)^T, \mathrm{vech}\,(V_1)^T\right)^T$.

This is the likelihood written down as if all components of z_t were available. In the k-th step

- we condition on the components of the z_t's which are actually available. This conditioning is based either on an initial parameter $\tau^{(0)}$ or on the parameters $\tau^{(k-1)}$ from the previous step in the EM algorithm and it is performed by the Kalman filter and smoothing algorithm applied to the state x_t. This leads to an interpolation of the missing components in the static factors z_t.
- using these interpolated values, we optimize the conditioned likelihood with respect to the parameters τ and obtain the updated parameter $\tau^{(k)}$.

The overall estimation procedure seems to be novel and needs further investigation.

3.3 The EM algorithm applied to a state space version of GDFMs

An alternative (see e.g. Mariano and Murasawa (2003) and Bańbura and Modugno (2014)) is to approximate[3] the

[3] The approximation arises from the fact that the noise spectral density f_u^N is assumed to be diagonal.

GDFM as given in Equations (1), (12) and (16) by a state space model of the following form:

$$y_t = (L\ 0 \cdots\ 0) \underbrace{\begin{pmatrix} z_t \\ z_{t-1} \\ \cdots \\ z_{t-p+1} \\ \tilde{u}_t \end{pmatrix}}_{=\tilde{x}_t} + \eta_t \tag{26}$$

$$\tilde{x}_t = \begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p & 0 \\ I_n & & & & \\ & \ddots & & & \\ & & I_n & 0 & \\ 0 & \cdots & & 0 & \alpha \end{pmatrix} \tilde{x}_{t-1} + \begin{pmatrix} v_t \\ 0 \\ \vdots \\ 0 \\ \xi_t \end{pmatrix} \tag{27}$$

where η_t is white noise with covariance matrix κI_N, κ being a fixed value. The idiosyncratic variable u_t is modelled as

$$u_{it} = \eta_{it} + \tilde{u}_{it}, \tag{28}$$
$$\tilde{u}_{it} = \alpha_i \tilde{u}_{i,t-1} + \xi_{it}, \quad |\alpha_i| < 1, \tag{29}$$

where the η_{is} and $\tilde{u}_{jt}$ are uncorrelated for all $i, j \in \mathbb{N}$ and $s, t \in \mathbb{Z}$, and where $\alpha = \mathrm{diag}(\alpha_1, \ldots, \alpha_N)$,

The state space model (26)-(27) then is estimated using an EM algorithm as described above.

4. EFFECTS OF NON-IDENTIFIABILITY ON THE EM-ALGORITHM

The convergence behavior of the EM algorithm described above still needs further analysis. This is part of ongoing research of the authors of this paper. Thereby two problems are under investigation:

- The effect of (possible) non-minimality of the state space system (26),(27) on the performance of the EM algorithm.
- The effect of (possible) non-identifiability due to mixed-frequency data.

We demonstrate the second effect for an extreme case, namely for vector moving average (VMA) models, where we do not even have generic identifiability (see Deistler et al. (2016)).

We assume that the data (y_t) are generated by VMA(1) systems of the form

$$y_t = \varepsilon_t + B_1 \varepsilon_{t-1}, \tag{30}$$

where (ε_t) is a white noise process with covariance matrix $\Sigma > 0$ and $\det(I + B_1 z) \neq 0$ for all $|z| \leq 1$. Here (y_t) does not correspond to a GDFM. However it may correspond to a static factor from a GDFM.

The data can only be observed at mixed frequencies as described in Section 3.1, i.e. $y_t = \begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix}$.

Following Metaxoglou and Smith (2007) we write

$$y_t = (I + B_1 z)\varepsilon_t = (I + \tilde{B}_1 z) w_t + \eta_t,$$

where (w_t) is a white noise process with covariance matrix Σ_w and (η_t) is a white noise process with a fixed diagonal covariance matrix κI_n. Here κ is small and fixed such that $\det(I + \tilde{B}_1 z) \neq 0$, $|z| \leq 1$. The processes (w_t) and (η_t) are assumed to be independent.

The unknown parameter B_1 is estimated via an EM algorithm applied on the time variable state space model

$$y_t^{\times} = C_t \underbrace{\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix}}_{=a_t} + \eta_t \tag{31}$$

$$a_t = \begin{pmatrix} 0_{N\times N} & 0_{N\times N} \\ I_N & 0_{N\times N} \end{pmatrix} a_{t-1} + \begin{pmatrix} I_N \\ 0 \end{pmatrix} w_t \tag{32}$$

where $C_t = \begin{pmatrix} I_N & \tilde{B}_1 \end{pmatrix}$ for $t \in 2\mathbb{Z}$ and $C_t = \begin{pmatrix} I_{N_f} & \tilde{B}_{1,N_f} \end{pmatrix}$ for $t \in 2\mathbb{Z} - 1$. Here $\tilde{B}_{1,N_f}$ denotes the first N_f rows of the matrix $\tilde{B}_1$. The EM algorithm applied is an extension of the EM algorithm discussed in Metaxoglou and Smith (2007) to the mixed frequency case. Note that we can derive B_1 from $\tilde{B}_1$ via the equation $B_1 = \begin{pmatrix} I_N & \tilde{B}_1 \end{pmatrix} K$, where K is the steady state Kalman gain corresponding to the single frequency version of model (31)-(32).

The effect of non-identifiability on the behavior of the EM algorithm is demonstrated via a Monte Carlo analysis. We report results on two models,

Model 1:

$$y_t = \varepsilon_t + \underbrace{\begin{pmatrix} 0.785 & 0 \\ -0.462 & 0.806 \end{pmatrix}}_{B_1} \varepsilon_{t-1}, \tag{33}$$

Model 2:

$$y_t = \varepsilon_t + \underbrace{\begin{pmatrix} -5.545 & 0 \\ 0.070 & -0.704 \end{pmatrix}}_{B_1} \varepsilon_{t-1}, \tag{34}$$

where $\varepsilon_t \sim \mathcal{N}(0, I_2)$. For each model we generate mixed frequency data with 500 observations. In order to deal with the problem of local optima the EM algorithm is initialized with 50 different starting values. The estimate $\hat{B}_1$ is given by the result corresponding to the "best value" of the likelihood. We repeat this 100 times.

The analysis reveals that the EM algorithm is very sensitive with respect to its starting value. In particular it does not always converge to a value near the true parameter B_1, as can be seen in Figure 1. In this figure we plot for Models 1, 2 (on the x-axis) the 100 simulated estimation errors $||\hat{B}_1 - B_1||$ (on the y-axis). This result corresponds to the theoretical result, shown by Deistler et al. (2016), that in the mixed frequency case the set of observationally equivalent VMA(1) system constitutes a lower dimensional manifold. The EM algorithm seems to converge to arbitrary representatives of that manifold. Thus convergence of the EM algorithm does not imply convergence to the "true" parameter, but to the "true" equivalence class.

5. CONCLUSION

In this contribution we discuss identifiability and estimation of generalized linear dynamic factor models in the case of mixed frequency data. In particular we discuss two estimation procedures: The first procedure is based on a

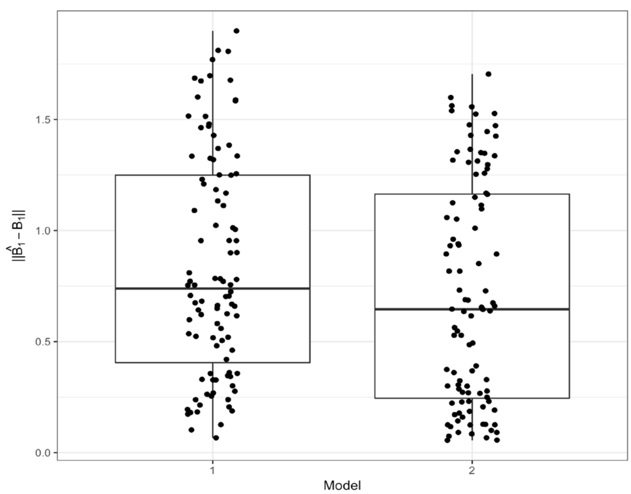

Fig. 1. The distribution of the Frobenius norm $||\hat{B}_1 - B_1||$ for Models 1 and 2 given in equations (33) and (34).

denoising step using PCA followed by applying an EM algorithm for estimating an AR model for the static factors. This procedure seems to be novel. The second procedure, which has already been proposed in the literature, uses an EM algorithm applied to a state space formulation of GDFMs.

It is shown by an example, that convergence problems for the EM algorithm may arise due to non-identifiability. Further investigations concerning convergence properties of the EM algorithm are needed.

REFERENCES

Al-Sadoon, M.M. (2015). A general theory of rank testing.

Anderson, B.D.O., Deistler, M., Felsenstein, E., Funovits, B., Koelbl, L., and Zamani, M. (2016a). Multivariate AR Systems and Mixed Frequency Data: g-Identifiability and Estimation. *Econometric Theory*, 32(4), 793–826. doi:10.1017/S0266466615000043.

Anderson, B.D.O., Deistler, M., Felsenstein, E., and Koelbl, L. (2016b). The Structure of Multivariate AR and ARMA Systems: Regular and Singular Systems; the Single and the Mixed Frequency Case. *Journal of Econometrics*, 192(2), 366–373. doi: doi:10.1016/j.jeconom.2016.02.004.

Anderson, B.D. and Deistler, M. (2008). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement, and System Integration*, 1(4), 284–292.

Bańbura, M. and 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.

Caines, P.E. (1988). *Linear Stochastic Systems.* John Wiley NYC.

Chamberlain, G. (1983). Funds, Factors and Diversification in Arbitrage Pricing Models. *Econometrica*, 51(5), 1305–1324.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure and Meanvariance Analysis on Large Asset Markets. *Econometrica*, 51(5), 1281–1304.

Deistler, M., Anderson, B.D.O., Filler, A., Zinner, C., and Chen, W. (2010). Generalized Linear Dynamic Factor Models - An Approach via Singular Autoregressions. *European Journal of Control*, 16(3), 211–224. doi: 10.3166/ejc.16.211-224.

Deistler, M., Koelbl, L., and Anderson, B.D. (2016). Non-identifiability of vma and varma systems in the mixed frequency case. *Econometrics and Statistics.*

Deistler, M., Scherrer, W., and Anderson, B.D. (2015). The structure of generalized linear dynamic factor models. In *Empirical Economic and Financial Research*, 379–400. Springer.

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

Felsenstein, E. (2014). *Regular and Singular AR and ARMA Models: The Single and The Mixed Frequency Case.* Ph.D. thesis, Vienna University of Technology.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric Theory*, 17, 1113–1141.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *Review of Economics and Statistics*, 82(4), 540–554.

Geweke, J. (1977). *Latent Variables in Socio-Economic Models, Ed.: Aigner, D. J.*, chapter 19, The Dynamic Factor Analysis of Economic Time Series Models, 365–383. North-Holland Amsterdam.

Hannan, E.J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems.* SIAM Classics in Applied Mathematics, Philadelphia.

Ljung, L. (1987). *System Identification Theory for the User.* Prentice Hall, New Jersey.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer Science & Business Media.

Mariano, R.S. and Murasawa, Y. (2003). A New Coincident Index of Business Cycles based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, 18(4), 427–443.

Metaxoglou, K. and Smith, A. (2007). Maximum likelihood estimation of varma models using a state-space em algorithm. *Journal of Time Series Analysis*, 28(5), 666–685.

Reinsel, G. (1993). *Elements of Multivariate Time Series Analysis.* Springer Series in Statistics, New York.

Scherrer, W. and Deistler, M. (1998). A Structure Theory for Linear Dynamic Errors-in-Variable Models. *SIAM Journal on Control and Optimization*, 36(6), 2148–2175.

Shumway, R.H. and Stoffer, D.S. (2016). *Time series analysis and its applications: with R examples.* Springer Science & Business Media.

Stock, J.H. and Watson, M.W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179.