Dynamic Factor Models

Estimation of VAR Systems from Mixed-Frequency Data: The Stock and the Flow Case

Lukas Koelbl [a] Alexander Braumann [a] Elisabeth Felsenstein [a] Manfred Deistler [a b]

ESTIMATION OF VAR SYSTEMS FROM MIXED-FREQUENCY DATA: THE STOCK AND THE FLOW CASE

Lukas Koelbl[a], Alexander Braumann[a], Elisabeth Felsenstein[a] and Manfred Deistler[a,b]

[a]*Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria*
[b]*Institute for Advanced Studies, Vienna, Austria*

ABSTRACT

This paper is concerned with estimation of the parameters of a high-frequency VAR model using mixed-frequency data, both for the stock and for the flow case. Extended Yule–Walker estimators and (Gaussian) maximum likelihood type estimators based on the EM algorithm are considered. Properties of these estimators are derived, partly analytically and by simulations. Finally, the loss of information due to mixed-frequency data when compared to the high-frequency situation as well as the gain of information when using mixed-frequency data relative to low-frequency data is discussed.

Keywords: Dynamic models; EM estimation method; extended Yule-Walker equations; Mixed frequency data

JEL classifications: C18; C38

Dynamic Factor Models
Advances in Econometrics, Volume 35, 43–73

ISSN: 0731-9053/doi:10.1108/S0731-905320150000035002

1. INTRODUCTION

In this paper, we consider the problem of estimating the parameters of an n-dimensional high-frequency VAR model

$$y_t = \begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \nu_t, \quad t \in \mathbb{Z} \tag{1}$$

using mixed-frequency data. We actually observe mixed-frequency data of the form

$$\begin{pmatrix} y_t^f \\ w_t \end{pmatrix} \tag{2}$$

where

$$w_t = \sum_{i=1}^{N} c_i y_{t-i+1}^s \tag{3}$$

where $c_i \in \mathbb{R}$, $1 < N \in \mathbb{N}$ and at least one $c_i \neq 0$. Here, the n_f-dimensional, say, fast component y_t^f is observed at the highest (sampling) frequency $t \in \mathbb{Z}$ and the n_s-dimensional slow component w_t is observed only for $t \in N\mathbb{Z}$, that is, for every N-th time point. In this paper, we assume that $n_f \geqslant 1$. Equation (3) represents the general case. For the case of flow data we have that $c_i = 1$, for $i = 1, \ldots, N$, whereas for the case of stock data we have that $c_1 = 1$ and $c_i = 0$, for $i = 2, \ldots, N$.

Throughout we assume the following for the high-frequency VAR model: The system parameters $A_i \in \mathbb{R}^{n \times n}$ satisfy the stability assumption

$$\det(a(z)) \neq 0 \;\; |z| \leqslant 1 \tag{4}$$

where $a(z) = I - A_1 z - \cdots - A_p z^p$ and the polynomial order p is given or specified. Here, z is used for the complex variable as well as for the backward shift on the integers $\mathbb{Z}$. We assume that (ν_t) is white noise and we only consider the stable steady-state solution $y_t = a(z)^{-1}\nu_t$. The rank q of the innovation covariance matrix $\Sigma_\nu = \mathbb{E}(\nu_t \nu_t^T)$ is given or specified, where $q \leqslant n$ holds. When the innovation matrix Σ_ν is nonsingular, the system is called regular, otherwise it is called singular. Singular autoregressive systems are important as models for latent variables and the corresponding static factors in generalized dynamic factor models (GDFMs) (see Deistler, Anderson, Filler, Zinner, & Chen, 2010; Forni, Hallin, Lippi, & Reichlin, 2000; Forni, Hallin, Lippi, & Zaffaroni, 2015; Stock & Watson, 2002). They are also important for DSGE models for the case where the number of shocks is strictly smaller than the number of outputs (see Komunjer & Ng, 2011).

The parameter space for the high-frequency models considered is given by

$$\Theta = \left\{ \left(A_1, \ldots, A_p\right) \middle| \det(a(z)) \neq 0, |z| \leqslant 1 \right\} \times \left\{ \Sigma_\nu \middle| \Sigma_\nu = \Sigma_\nu^T, \Sigma_\nu \geqslant 0, \mathrm{rk}(\Sigma_\nu) = q \right\}$$

where rk(A) denotes the rank of the matrix A. Since Σ_ν is of rank $q \leqslant n$, we can write $\Sigma_\nu = bb^T$, where b is an $(n \times q)$ matrix. Accordingly, $\nu_t = b\varepsilon_t$, where $\mathbb{E}\left(\varepsilon_t \varepsilon_t^T\right) = I_q$. For given Σ_ν, b is unique up to postmultiplication by an orthogonal matrix. For a particular unique choice of b, see Filler (2010).

Model (1) can be written in companion form as

$$\underbrace{\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}}_{x_{t+1}} = \underbrace{\begin{pmatrix} A_1 & \ldots & A_{p-1} & A_p \\ I_n & & & \\ & \ddots & & \\ & & I_n & 0 \end{pmatrix}}_{\mathcal{A}} \underbrace{\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix}}_{x_t} + \underbrace{\begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathcal{B}} \varepsilon_t \tag{5}$$

$$y_t = \underbrace{\left(A_1 \cdots A_p\right)}_{\mathcal{C}} x_t + b\varepsilon_t \tag{6}$$

The solutions of Eq. (1) and of Eqs. (5), (6) are of the form

$$y_t = a(z)^{-1}\nu_t = a(z)^{-1}b\varepsilon_t = \left(\mathcal{C}(I - \mathcal{A}z)^{-1}\mathcal{B}z + b\right)\varepsilon_t \tag{7}$$

where $k(z) = a(z)^{-1} = \sum_{j=0}^{\infty} k_j z^j$ is the transfer function from (ν_t) to (y_t) and $k(z)b = \mathcal{C}(I - \mathcal{A}z)^{-1}\mathcal{B}z + b$ is the transfer function from (ε_t) to (y_t).

Due to the mixed-frequency structure of the observed data, the population second moments

$$\gamma^{ff}(h) = \mathbb{E}\left(y^f_{t+h}\left(y^f_t\right)^T\right), \quad h \in \mathbb{Z}$$
$$\gamma^{wf}(h) = \mathbb{E}\left(w_{t+h}\left(y^f_t\right)^T\right), \quad h \in \mathbb{Z}$$
$$\gamma^{ww}(h) = \mathbb{E}\left(w_{t+h}(w_t)^T\right), \quad h \in N\mathbb{Z} \tag{8}$$

can be directly estimated. For estimation of the high-frequency parameters, identifiability is a core issue. In our context, identifiability means that the parameters of the high-frequency system can be uniquely obtained from the population second moments given in Eq. (8). As has been discussed in Anderson et al. (2012, 2015), identifiability can only be guaranteed generically. This means that we can guarantee identifiability on a set containing an open and dense subset of the parameter space. We call this g-identifiability. For the remaining part of the paper, unless the contrary is stated explicitly, we assume that the true system is in the identifiable set.

The paper is organized as follows: In Section 2, we introduce extended Yule–Walker estimators, first for the case of stock variables (see Chen & Zadrozny, 1998) and then for the general case (3), as well as a (Gaussian) maximum likelihood type estimator based on the EM algorithm. Note that these estimators do not necessarily lead to a stable system, nor do they necessarily give a positive (semi)-definite innovations covariance matrix of rank q. For these reasons, in Section 3, algorithms are discussed for transforming these estimators to a stable and positive (semi)-definite form, respectively. In Section 4, the asymptotic properties of the extended Yule–Walker estimators are derived. In Section 5, a simulation study is presented, in which we compare the extended Yule–Walker estimators with the maximum likelihood type estimator. Furthermore, the information loss due to mixed-frequency data compared to high-frequency data on the one hand and the information gain from using mixed-frequency data compared to low-frequency data on the other hand are discussed.

2. MIXED-FREQUENCY ESTIMATORS

2.1 Extended Yule–Walker Estimators: The Stock Case

In Chen and Zadrozny (1998), extended Yule–Walker (XYW) equations have been proposed for estimation of the high-frequency parameters from mixed-frequency stock data. On a population level, these XYW equations are of the form

$$\underbrace{\mathbb{E}\left[y_t\left(\left(y^f_{t-1}\right)^T,\ldots,\left(y^f_{t-np}\right)^T\right)\right]}_{Z_1} = \underbrace{\left(A_1,\ldots,A_p\right)}_{A}\underbrace{\mathbb{E}\left[\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p}\end{pmatrix}\left(\left(y^f_{t-1}\right)^T,\ldots,\left(y^f_{t-np}\right)^T\right)\right]}_{Z_0} \tag{9}$$

where $Z_0 \in \mathbb{R}^{np\times n_f np}$ and $Z_1 \in \mathbb{R}^{n\times n_f np}$. In Anderson et al. (2012, 2015), it has been shown that Z_0 has full row rank np on a generic subset of the parameter space. Thus, in this case, $\left(A_1,\ldots,A_p\right)$ are uniquely determined from $\left(A_1,\ldots,A_p\right) = Z_1 Z_0^{\dagger}$, where $Z_0^{\dagger} = Z_0^T\left(Z_0 Z_0^T\right)^{-1}$ is the Moore–Penrose pseudo-inverse of Z_0.

XYW estimators are obtained by replacing the population second moments by their sample counterparts:

$$\hat{\gamma}^{ff}(h) = \frac{1}{T}\sum_{t=1}^{T-h} y^f_{t+h}\left(y^f_t\right)^T, \quad h \geqslant 0 \tag{10}$$

$$\hat{\gamma}^{ff}(h) = \hat{\gamma}^{ff}(-h)^T \tag{11}$$

$$\hat{\gamma}^{wf}(h) = \frac{1}{T/N}\sum_{t=t_1}^{t_2} w_{Nt}\left(y^f_{Nt-h}\right)^T \tag{12}$$

where the estimator of $\gamma^{wf}(h)$ has only (approximately) $1/N$-th of the summands compared to the estimator of $\gamma^{ff}(h)$ due to the missing observations and

$$t_1 = \begin{cases} 1 & N > h \\ \left\lfloor \frac{h}{N} \right\rfloor + 1 & N \leqslant h \end{cases}, \quad t_2 = \begin{cases} \left\lfloor \frac{T}{N} \right\rfloor & h \geqslant 0 \\ \left\lfloor \frac{T+h}{N} \right\rfloor & h < 0 \end{cases}$$

For the case of stock variables considered here, we have $w_{Nt} = y^s_{Nt}$.

Let $\hat{Z}_0$ and $\hat{Z}_1$ denote the corresponding estimators of Z_0 and Z_1, respectively. If Z_0 has full row rank and if we assume that its estimator is consistent, $\hat{Z}_0$ will be also of full row rank, from a certain T_0 onwards. The XYW equations with Z_0 and Z_1 replaced by $\hat{Z}_0$ and $\hat{Z}_1$ are overdetermined in general and can be solved in different ways. We may define the XYW estimator as follows:

$$\hat{A}_{\mathrm{XYW}} = \hat{Z}_1 \hat{Z}_0^{\dagger} \tag{13}$$

where $\hat{Z}_0^{\dagger} = \hat{Z}_0^T \left(\hat{Z}_0 \hat{Z}_0^T \right)^{-1}$ is the Moore–Penrose pseudo inverse of $\hat{Z}_0$. Alternatively, the generalized method of moments (GMM) estimator (see Hansen, 1982), which is defined as

$$\hat{A}_{\mathrm{GMM}} = \underset{A \in \mathbb{R}^{n \times np}}{\arg\min}\, \mathrm{vec}\left(\hat{Z}_1 - A\hat{Z}_0\right)^T Q_T \mathrm{vec}\left(\hat{Z}_1 - A\hat{Z}_0\right) \tag{14}$$

can be used. Here, (Q_T) is a sequence of random weighting matrices which converges almost surely to a constant, symmetric, positive-definite matrix Q_0. Let vec(.) and $\otimes$ denote columnwise vectorization and the Kronecker product, respectively. In addition, we assumed that $\hat{Z}_0$ has full row rank and that Q_T is nonsingular. The solution of Eq. (14) is

$$\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right) = \underbrace{\left(\left(\hat{Z}_0 \otimes I_n\right) Q_T \left(\hat{Z}_0^T \otimes I_n \right) \right)^{-1} \left(\hat{Z}_0 \otimes I_n\right) Q_T}_{\hat{G}^{\dagger}_{Q_T}} \mathrm{vec}\left(\hat{Z}_1\right)$$

Note that if we set $Q_T = I_{n^2 p n_f}$, we obtain the XYW estimator.

It may be useful to exploit the information contained in covariances corresponding to higher-order lags in order to improve the quality of the estimator: For $k \geqslant 0$, we obtain

$$\begin{aligned} &\underbrace{\mathbb{E}\left(y_t \left(\left(y^f_{t-1}\right)^T, \ldots, \left(y^f_{t-np-k}\right)^T \right) \right)}_{Z_{1,k}} \\ &\qquad = \left(A_1, \ldots, A_p\right) \underbrace{\mathbb{E}\left(\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} \left(\left(y^f_{t-1}\right)^T, \ldots, \left(y^f_{t-np-k}\right)^T \right) \right)}_{Z_{0,k}} \end{aligned} \tag{15}$$

Note that the XYW estimators do not use the information contained in the autocovariances of the slow process, which can be directly observed.

A symmetric estimator $\hat{\Sigma}_\nu$ for Σ_ν is obtained from

$$\mathrm{vec}\left(\hat{\Sigma}_\nu\right) = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - \left(\hat{\mathcal{A}}\otimes\hat{\mathcal{A}}\right)\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)\right)^{-1}\mathrm{vec}(\hat{\gamma}(0)) \quad (16)$$

where $\mathcal{G} = (I_n, 0, \ldots, 0)$ and $\hat{\mathcal{A}}$ is the companion form corresponding to Eq. (5) with the true parameters replaced by their estimators. Here, we assumed that $\left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)\right)$ is nonsingular, an assumption which is generically fulfilled (see Anderson et al., 2015).

2.2 Extended Yule−Walker Estimators: The General Case

Let

$$z_t = \sum_{i=1}^{N} c_i y_{t-i+1} = \begin{pmatrix} \sum_{i=1}^{N} c_i y^f_{t-i+1} \\ w_t \end{pmatrix} = \begin{pmatrix} z^f_t \\ z^s_t \end{pmatrix} \quad (17)$$

and $\gamma_z(h) = \mathbb{E}\left(z_t z^T_{t-h}\right)$. For convenience, we assume that $c_N \neq 0$. In this case, we obtain another form of the XYW equations as follows:

$$\underbrace{\mathbb{E}\left[z_t\left(\left(y^f_{t-N}\right)^T, \ldots, \left(y^f_{t-np-N+1}\right)^T\right)\right]}_{Z^g_1} = \left(A_1, \ldots, A_p\right)\underbrace{\mathbb{E}\left[\begin{pmatrix} z_{t-1} \\ \vdots \\ z_{t-p} \end{pmatrix}\left(\left(y^f_{t-N}\right)^T, \ldots, \left(y^f_{t-np-N+1}\right)^T\right)\right]}_{Z^g_0} \quad (18)$$

In Koelbl (2015), it is shown that the matrix Z_0^{g} has full row rank on a generic subset of the parameter space and thus the XYW estimator in the general case is defined as

$$\hat{A}_{\mathrm{XYW}}^{\mathrm{g}} = \hat{Z}_1^{\mathrm{g}}\left(\hat{Z}_0^{\mathrm{g}}\right)^{\dagger} \tag{19}$$

where $\hat{Z}_0^{\mathrm{g}}$ and $\hat{Z}_1^{\mathrm{g}}$ are the estimators for Z_0^{g} and Z_1^{g}, respectively.

Let $m = \max(p, N)$, $\mathcal{H}_m = \left(c_1 I_n, c_2 I_n, \ldots, c_N I_n, 0_{n\times n(m-N)}\right)$, $\mathcal{G}_m = \left(I_n, 0_{n\times n(m-1)}\right)$, and $\mathcal{A}_m$ be the companion form with the system parameters

$$A_i^m = \begin{cases} A_i & i = 1, \ldots, p \\ 0 & i = p+1, \ldots, m \end{cases}$$

Then it is easy to see that

$$\mathrm{vec}\left(\gamma_z(0)\right) = \left((\mathcal{H}_m \otimes \mathcal{H}_m)\left(I_{(nm)^2} - (\mathcal{A}_m \otimes \mathcal{A}_m)\right)^{-1}\left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right)\right)\mathrm{vec}(\Sigma_\nu) \tag{20}$$

Again in Koelbl (2015), it is shown that the first term on the right-hand side of Eq. (20) is generically nonsingular. Thus, we can define the symmetric estimator for the noise parameters as

$$\mathrm{vec}\left(\hat{\Sigma}_\nu\right) = \left((\mathcal{H}_m \otimes \mathcal{H}_m)\left(I_{(nm)^2} - \left(\hat{\mathcal{A}}_m \otimes \hat{\mathcal{A}}_m\right)\right)^{-1}\left(\mathcal{G}_m^T \otimes \mathcal{G}_m^T\right)\right)^{-1}\mathrm{vec}\left(\hat{\gamma}_z(0)\right)$$

2.3 Maximum Likelihood Estimation and the EM Algorithm

First, again, we consider the case of stock variables. Throughout this section we assume, for the sake of simplicity, that $q = n$, that is, the AR process is regular. For a given N, we define the blocked observed process $\tilde{y}_t = \left(y_t^T, \left(y_{t-1}^f\right)^T, \ldots, \left(y_{t-N+1}^f\right)^T\right)^T$, $t \in N\mathbb{Z}$ (see Anderson et al. 2015). The

(Gaussian) log-likelihood (ignoring the constant) of observations times -2 is given by (see Hannan & Deistler, 2012)

$$\tilde{L}_T(\theta) = \frac{1}{T}\log\left(\det\left(\tilde{\Gamma}_T(\theta)\right)\right) + \frac{1}{T}\tilde{Y}^T_{\text{obs}}\left(\tilde{\Gamma}_T(\theta)\right)^{-1}\tilde{Y}_{\text{obs}} \tag{21}$$

where $\tilde{Y}_{\text{obs}} = \left(\tilde{y}^T_N, \tilde{y}^T_{2N}, \ldots, \tilde{y}^T_T\right)^T$ are the observed data and $\tilde{\Gamma}_T(\theta) = \mathbb{E}\left(\tilde{Y}_{\text{obs}}(\theta)\tilde{Y}^T_{\text{obs}}(\theta)\right)$. Note that in the latter expression, the random variables $\tilde{Y}_{\text{obs}}(\theta)$ are considered to be generated by an arbitrary parameter $\theta \in \Theta$. Also note that $\tilde{\Gamma}_T(\theta)$ is a square matrix with dimension $\left(n + (N-1)n_f\right)T/N$, which is in general a very large matrix in block Toeplitz form and accordingly its inversion may cause problems, even when taking advantage of its block Toeplitz structure (see Wax & Kailath, 1983).

The EM algorithm (MLE-EM) is an iterative procedure to find the minimizer of $\tilde{L}_T(\theta)$. The idea of this algorithm is to successively minimize the conditional expectation of

$$L_T(\theta) = \frac{1}{T}\log(\det(\Gamma_T(\theta))) + \frac{1}{T}Y^T(\Gamma_T(\theta))^{-1}Y \tag{22}$$

of the complete-data $Y = \left(y^T_1, \ldots, y^T_T\right)^T$ given the observed data $\tilde{Y}_{\text{obs}}$. Here, $\Gamma_T(\theta) = \mathbb{E}\left(Y(\theta)Y^T(\theta)\right)$ where now the random variables $Y(\theta)$ are generated by an arbitrary parameter value $\theta \in \Theta$. We now follow Shumway and Stoffer (1982): Observe that for the stock case the mixed-frequency data can be represented by a time variable state-space system

$$\begin{aligned} x_{t+1} &= \mathcal{A}x_t + \mathcal{B}\varepsilon_t \\ y^{\times}_t &= \mathcal{G}_t x_{t+1} \end{aligned} \tag{23}$$

where $x_t = \left(y^T_{t-1}, \ldots, y^T_{t-p}\right)^T$ and $\mathcal{G}_{Nt} = (I_n, 0, \ldots, 0)$, $\mathcal{G}_{Nt-k} = \left(I_{n_f}, 0, \ldots, 0\right)$ for $k = 1, \ldots, N-1$. The vector $y^{\times}_t$ contains only observable components.

Throughout this section, we assume that the innovations ν_t in Eq. (1) are Gaussian white noise and that $x_1 \sim \mathcal{N}_{np}(0, V_1)$ is independent from $\nu_1, \ldots, \nu_T$. The complete-data log-likelihood, when modified with respect to the initial values (and using the same symbol as in Eq. (22)), can be factorized such that

$$L_T(\tau) = \log(\det(V_1)) + x_1^T V_1^{-1} x_1 \\ + T\log(\det(\Sigma_\nu)) + \sum_{t=1}^{T} (\mathcal{G}x_{t+1} - Ax_t)^T \Sigma_\nu^{-1} (\mathcal{G}x_{t+1} - Ax_t)$$

where $\mathcal{G} = (I_n, 0, \ldots, 0)$ and $\tau = \left(\text{vec}(A)^T, \text{vech}(\Sigma_\nu)^T, \text{vech}(V_1)^T\right)^T$.

The algorithm starts with an initial value $\tau^{(0)}$, for example, where $\left(\hat{A}_1, \ldots, \hat{A}_p\right)$ and $\hat{\Sigma}_\nu$ are XYW estimates and where V_1 is the corresponding solution of the Lyapunov equations, that is, $\Gamma_p = \mathcal{A}\Gamma_p\mathcal{A}^T + \mathcal{B}\mathcal{B}^T$ where $\Gamma_p = \mathbb{E}\left(x_t x_t^T\right)$. Let $\tau^{(k)}$ be the estimate at the k-th iteration, iteration $k+1$ is as follows:

E-step. The expected complete-data log-likelihood conditional on $\tilde{Y}_{\text{obs}}$, say $Q\left(\tau|\tau^{(k)}\right) = \mathbb{E}_{\tau^{(k)}}\left(L_T(\tau)|\tilde{Y}_{\text{obs}}\right)$, where the notation $\mathbb{E}_{\tau^{(k)}}$ indicates that the conditional expectation is taken corresponding to the parameter $\tau^{(k)}$, is given by

$$Q\left(\tau|\tau^{(k)}\right) = \log(\det(V_1)) + \text{tr}\left(V_1^{-1}\left(x_{1|T}x_{1|T}^T + P_{1|T}\right)\right) + \text{T}\log(\det(\Sigma_\nu)) \\ + \text{tr}\left(\Sigma_\nu^{-1}\left(\mathcal{G}S_{11}\mathcal{G}^T + AS_{00}A^T - \mathcal{G}S_{10}A^T - AS_{10}^T\mathcal{G}^T\right)\right)$$

with

$$S_{00} = \sum_{t=1}^{T-1}\left(x_{t|T}x_{t|T}^T + P_{t|T}\right)$$
$$S_{11} = \sum_{t=2}^{T}\left(x_{t|T}x_{t|T}^T + P_{t|T}\right)$$
$$S_{10} = \sum_{t=2}^{T}\left(x_{t|T}x_{t-1|T}^T + P_{t,t-1|T}\right)$$

where

$$x_{t|T} = \mathbb{E}_{\tau^{(k)}}\left(x_t|\tilde{Y}_{\text{obs}}\right)$$
$$P_{t|T} = \mathbb{E}_{\tau^{(k)}}\left(x_t - x_{t|T}\right)\left(x_t - x_{t|T}\right)^T$$
$$P_{t,t-1|T} = \mathbb{E}_{\tau^{(k)}}\left(x_t - x_{t|T}\right)\left(x_{t-1} - x_{t-1|T}\right)^T$$

They can be calculated by means of the Kalman filter or other smoothing algorithms (see Shumway & Stoffer, 1982).

M-step. Determine $\tau^{(k+1)}$ by minimizing the expected conditional likelihood, that is, $\tau^{(k+1)} = \arg\min_\tau Q(\tau|\tau^{(k)})$. The parameter updates are given by

$$\begin{aligned} A^{(k)} &= \mathcal{G}S_{10}S_{00}^{-1} \\ \Sigma^{(k)} &= T^{-1}\left(\mathcal{G}S_{11}\mathcal{G}^T - \mathcal{G}S_{10}S_{00}^{-1}S_{10}^T\mathcal{G}^T\right) \\ V_1^{(k)} &= P_{1|T} + x_{1|T}x_{1|T}^T \end{aligned}$$

The algorithm stops if the relative decrease of the log-likelihoods of observations is below a prespecified threshold. Here, we do not discuss convergence properties of the EM algorithm.

In the general case, the state-space system has to be written as (see also Mariano & Murasawa, 2010)

$$\begin{aligned} x_{t+1}^m &= \mathcal{A}_m x_t^m + \mathcal{B}_m \varepsilon_t \\ y_t^\times &= \mathcal{G}_t^m x_{t+1}^m \end{aligned} \tag{24}$$

where $x_t^m = \left(y_{t-1}^T, \ldots, y_{t-m}^T\right)^T$, $\mathcal{B}_m = \left(b^T, 0_{n\times n(m-1)}\right)^T$, $\mathcal{G}_{Nt-k}^m = \left(I_{n_f}, 0, \ldots, 0\right)$ for $k = 1, \ldots, N-1$ and

$$\mathcal{G}_{Nt}^m = \begin{pmatrix} I_{n_f} & 0 & \cdots & 0 & 0 & 0_{n_f\times n(m-N)} \\ 0_{n_s\times n_f} & c_1 I_{n_s} & \cdots & 0 & c_N I_{n_s} & 0_{n_s\times n(m-N)} \end{pmatrix}$$

The algorithm described above for the stock case can be analogously applied to the general case with the state-space model (5), (6) modified in a straightforward way.

3. PROJECTING THE MF ESTIMATORS ON THE PARAMETER SPACE

It is well known that in the high-frequency case, the Yule–Walker estimator always leads to a stable AR polynomial, provided that $\hat{\Gamma}_p > 0$ holds, where $\hat{\Gamma}_p$ is the high-frequency estimator of $\Gamma_p > 0$ (see Deistler & Anderson 2010). Furthermore, in the high-frequency case, the estimated covariance matrix of the noise is positive definite. In general, the XYW/GMM estimators do not fulfill these desirable properties and the same is true for the estimators obtained from the EM algorithm. Indeed in many simulations such a situation occurs. Consequently, in a second step, one has to check whether the estimated parameters, say $\hat{\theta}$, lie in the parameter space Θ. If $\hat{\theta}$ is not contained in this space, the question of finding a $\hat{\theta}_P \in \Theta$, which is sufficiently close to $\hat{\theta}$, arises. In this paper, we separate this problem in two sub-problems: The first problem is to find a stable polynomial matrix close to an unstable estimator of $a(z)$. The second problem is to find a positive (semi)-definite covariance matrix of rank q, which is close to an indefinite (symmetric) estimator of Σ_ν.

3.1 Stabilization of the Estimated System Parameters

In this section we commence from an unstable estimate for the system parameters, say $\hat{A}_{\text{un}} \in \mathbb{R}^{n \times np}$, corresponding to $\hat{a}_{\text{un}}(z)$, such that there exists a $z_0 \in \mathbb{C}$, $|z_0| \leqslant 1$, and $\det(\hat{a}_{\text{un}}(z_0)) = 0$. As $S = \{(A_1, \ldots, A_p) | \det(a(z)) \neq 0, |z| \leqslant 1\}$ is an open set, there exists no best approximation of such an $\hat{A}_{\text{un}}$, for instance in Frobenius norm, by an element of S. In addition, in general, S is nonconvex. We consider the problem of finding

$$\inf_{A \in S} \|A - \hat{A}_{\text{un}}\|^2 \tag{25}$$

There exists a substantial literature dealing with finding the "nearest" stable polynomial both for the univariate (see Combettes & Trussell, 1992; Moses & Liu, 1991; Orbandexivry, Nesterov, & van Dooren, 2013; Stoica & Moses, 1992) and the multivariate case (see Balogh & Pintelon, 2008; D'haene, Pintelon, & Vandersteen, 2006). An interesting way to solve the univariate stabilization problem is proposed in Orbandexivry et al. (2013) using the so-called Dikin Ellipsoid. We will repeat the most important steps of this procedure and generalize it to the multivariate case, which can be easily done. We point out that all these methods need a stable initial value.

Problem (25) can be reformulated as in Orbandexivry et al. (2013) to

$$\inf_{A,P} \frac{1}{2} \|A - \hat{A}_{\text{un}}\|_F^2 \tag{26}$$

where minimization with respect to P runs over $P = P^T > 0$, $P - \mathcal{A}P\mathcal{A}^T > 0$, with $\mathcal{A}$ the companion form of A. For a fixed $P = P^T > 0$, we can define the set

$$S_P = \left\{A \in \mathbb{R}^{n \times np} : \; P - \mathcal{A}P\mathcal{A}^T > 0, \mathcal{A} \text{ is the companion form of } A\right\} \subset S$$

and the function $b_P(A) = -\log\left(\det\left(P - \mathcal{A}P\mathcal{A}^T\right)\right)$, which is a barrier function. It follows from theorem 5 in Orbandexivry et al. (2013) that for $A \in S$, $P = P^T > 0$ such that $P - \mathcal{A}P\mathcal{A}^T > 0$ and any $0 \leqslant \alpha < 1$, the so-called Dikin Ellipsoid $\mathcal{E}(P, A; \alpha) = A + \left\{H \in \mathbb{R}^{n \times np} : \left\langle b''_P(A)H, H\right\rangle \leqslant \alpha\right\}$ is a subset of S_P where $\langle A, B\rangle = \mathrm{tr}\left(AB^T\right)$ and $\left\langle b''_P(A)H, H\right\rangle$ is the second derivative of $b_P(A)$ in a given direction H. Now, for given A and α, the question arises which P should be chosen such that $\mathcal{E}(P, A; \alpha)$ is maximized. In Orbandexivry et al. (2013) the authors argue that a good choice, say P^*, is given by solving

$$Q^{-1} - \mathcal{A}^T Q^{-1}\mathcal{A} = npI_{np} \tag{27}$$

$$P^* - \mathcal{A}P^*\mathcal{A}^T = Q \tag{28}$$

We are now in a position to formulate a new, restricted optimization problem for a given $0 \leqslant \alpha < 1$, $A \in S$ and a corresponding P^*:

$$\min_H \tfrac{1}{2}\|A + H - \hat{A}_{\mathrm{un}}\|_F^2 \tag{29}$$

where H is such that $\left\langle b''_{P*}(A)H, H\right\rangle \leqslant \alpha$. Note that we now have a convex optimization problem. It can be shown that $\left\langle b''_{P*}(A)H, H\right\rangle \leqslant \alpha$ can be rewritten as $\mathrm{vec}(H)^T B\mathrm{vec}(H) \leqslant \alpha$, where

$$\frac{1}{2}B = \left(P^* \otimes \mathcal{G}Q^{-1}\mathcal{G}^T\right) + \left(P^*\mathcal{A}^T Q^{-1}\mathcal{A}P^* \otimes \mathcal{G}Q^{-1}\mathcal{G}^T\right)$$
$$+ \left(P^*\mathcal{A}^T Q^{-1}\mathcal{G}^T \otimes \mathcal{G}Q^{-1}\mathcal{A}P^*\right)K_{n,np} \tag{30}$$

and $K_{n,np}$ is a commutation matrix, see Magnus and Neudecker (1979). It is easy to conclude that the matrix B is symmetric positive definite and thus can be factorized as $B = UDU^T$, where D is a diagonal matrix with positive entries d_i, $i = 1, \ldots, n^2p$, and U is an orthonormal matrix. The solution of Eq. (29) can be derived (see Orbandexivry et al., 2013, p. 1199) by finding the root of the function

$$\psi(\lambda) = \sum_{i=1}^{n^2p} \frac{d_i\left(e_i^T U^T \mathrm{vec}(\hat{A}_{\mathrm{un}} - A)\right)^2}{(1+\lambda d_i)^2} - \alpha = 0 \tag{31}$$

with respect to $\lambda \in (0, \infty)$, where e_i is the i-th unit vector, and then substituting into

$$\mathrm{vec}(H) = \left(I_{n^2p} + \lambda B\right)^{-1} \mathrm{vec}\left(\hat{A}_{\mathrm{un}} - A\right) \tag{32}$$

It is worth mentioning that there exists a unique λ and therefore a unique H.

A stable initial estimator may be obtained, for example, by reflecting the unstable roots of $\hat{a}_{\mathrm{un}}(z)$ on the unit circle (see Lippi & Reichlin, 1994). The whole stabilization procedure has to be iterated.

3.2 Positive (Semi)-Definiteness of the Noise Covariance Matrix

Under assumptions that guarantee consistency of the sample second moments and the system parameters A, Eq. (16) gives a consistent estimator for Σ_ν. This estimate is symmetric but may not be positive (semi)-definite and of rank q. Consider

$$\inf_{\Sigma_{\mathrm{ps}} \in D} \|\Sigma_{\mathrm{ps}} - \hat{\Sigma}_\nu\|_F^2 \tag{33}$$

where $D = \left\{\Sigma_\nu \in \mathbb{R}^{n\times n} \mid \Sigma_\nu = \Sigma_\nu^T, \Sigma_\nu \geqslant 0, \mathrm{rk}(\Sigma_\nu) = q\right\}$. The matrix $\hat{\Sigma}_\nu$ can be represented as $\hat{\Sigma}_\nu = Q\Lambda Q^T$ where Λ is the diagonal matrix containing the eigenvalues λ_i in descending order and Q is an orthonormal matrix containing the corresponding eigenvectors. For simplicity, we assume that the q-th and the $(q+1)$-th eigenvalue are distinct.

To obtain an arbitrarily close solution of problem (33), we define $\hat{\Sigma}_{\mathrm{ps}} = Q\Lambda_+ Q^T$, where Λ_+ is a diagonal matrix with entries

$$\lambda_i^+ = \begin{cases} \max(\lambda_i, \varepsilon) & i = 1, \ldots, q \\ 0 & i = q+1, \ldots, n \end{cases}$$

for sufficiently small $\varepsilon > 0$. Note that by the so-called Wielandt–Hoffman Theorem (see Hoffman & Wielandt, 1953) $\sum_{i=1}^{n} \left(\lambda_i^A - \lambda_i^B\right)^2 \leqslant \|A - B\|_F^2$ holds for symmetric matrices $A, B \in \mathbb{R}^{n \times n}$, where λ_i^A and λ_i^B are the corresponding eigenvalues in a descending order, respectively. Thus $\hat{\Sigma}_{\text{ps}}$ gives an arbitrarily close solution of (33).

4. ASYMPTOTIC PROPERTIES OF THE XYW/GMM ESTIMATORS

In this section, we derive the asymptotic properties of the XYW estimator as well as of the generalized method of moments estimators. Whereas under suitable assumptions the Yule–Walker estimator has the same asymptotic covariance as the maximum likelihood estimator in the high-frequency case and thus is asymptotically efficient, this is not the case for the XYW/GMM estimators. The asymptotic distribution of XYW/GMM estimators is derived along the idea of first deriving the asymptotic distribution of the sample second moments of the observations, that is, deriving Bartlett's formula for the mixed-frequency case, and then, in a second step, linearizing the function attaching the parameters to the second-order moments of the observations. Throughout this section, we additionally assume that ν_t in Eq. (1) is independent identically distributed, $(\nu_t) \sim \text{IID}_n(0, \Sigma_\nu)$, and that $\eta = \mathbb{E}\left(\nu_t \nu_t^T \otimes \nu_t \nu_t^T\right)$ exists. For notational simplicity, we write (y_t) as $y_t = \sum_{j=-\infty}^{\infty} k_j \nu_{t-j}$, where $k_j = 0$ for $j < 0$. Note that (z_t) can be analogously represented as $z_t = \sum_{j=-\infty}^{\infty} \tilde{k}_j \nu_{t-j}$, where $\tilde{k}_j = \sum_{i=1}^{N} c_i k_{j-i+1}$. For convenience, we restrict ourselves mainly to the case of stock variables, that is, $w_{Nt} = y_{Nt}^s$. The general case will be discussed at the end of the section. In the following, we will use the partition $k_j = \begin{pmatrix} k_j^f \\ k_j^s \end{pmatrix}$, where k_j^f denotes the first n_f and k_j^s the last n_s rows of k_j, respectively.

Let $\kappa = \eta - \text{vec}(\Sigma_\nu)\text{vec}(\Sigma_\nu)^T - (\Sigma_\nu \otimes \Sigma_\nu) - K_{n,n}(\Sigma_\nu \otimes \Sigma_\nu)$, $\gamma^{z^f f}(h) = \mathbb{E}\left(z_t^f \left(y_t^f\right)^T\right) = \gamma^{ff}(h)$ and $\hat{\gamma}^{z^f f}(h)$ be the corresponding estimator. We denote convergence in distribution by $\xrightarrow{d}$ and convergence in probability by $\xrightarrow{p}$.

Theorem 1. Under the assumptions stated above in this section, we obtain

$$\sqrt{T}\left(\begin{pmatrix} \mathrm{vec}\left(\hat{\gamma}^{zf}(0)\right) \\ \mathrm{vec}\left(\hat{\gamma}^{wf}(0)\right) \\ \vdots \\ \mathrm{vec}\left(\hat{\gamma}^{zf}(s)\right) \\ \mathrm{vec}\left(\hat{\gamma}^{wf}(s)\right) \end{pmatrix} - \begin{pmatrix} \mathrm{vec}\left(\gamma^{zf}(0)\right) \\ \mathrm{vec}\left(\gamma^{wf}(0)\right) \\ \vdots \\ \mathrm{vec}\left(\gamma^{zf}(s)\right) \\ \mathrm{vec}\left(\gamma^{wf}(s)\right) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_h\left(0, \Sigma_\gamma\right)$$

where $h = (s+1)nn_f$ and $s \in \mathbb{N}$. Σ_γ is obtained as described in Lemmas 1−3 in the appendix.

The proof of Theorem 1 is also given in the appendix.

Remark 1. The last theorem can be extended to any lag including negative ones. Indeed, we will use lags $(1-p, \ldots, np)$ for the XYW and GMM estimator. Also note that the assumption that the innovations are i.i.d. can be relaxed, see, for example, Hall and Heyde (1980) and Francq and Zakoian (2009).

Remark 2. Note that we do not distinguish between singular and non-singular normal distributions. For a detailed discussion about singular normal distributions, see Khatri (1961), Rao (1972), and Anderson (1994).

Having obtained the asymptotic distribution of the covariance estimators, we have to linearize the mapping attaching the system parameters to the second moments of the observations. The next theorem derives the asymptotic distribution of the XYW/GMM estimators and is related to Gingras (1985). Let Θ_{XYW} be the generic set of the system and noise parameters where Z_0 has full row rank np.

Theorem 2. Let (y_t) be the output of system (1) with inputs $(\nu_t) \sim \mathrm{IID}_n(0, \Sigma_\nu)$, $\theta \in \Theta_{\mathrm{XYW}}$ and assume that $\eta = \mathbb{E}\left(\nu_t \nu_t^T \otimes \nu_t \nu_t^T\right)$ exists. Then the GMM estimator

$$\begin{aligned} \mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right) &= \left(\left(\hat{Z}_0 \otimes I_n\right) Q_T \left(\hat{Z}_0^T \otimes I_n\right)\right)^{-1} \left(\hat{Z}_0 \otimes I_n\right) Q_T \mathrm{vec}\left(\hat{Z}_1\right) \\ &= \hat{G}_{Q_T}^{\dagger} \mathrm{vec}\left(\hat{Z}_1\right) \end{aligned}$$

is asymptotically normal with zero mean and a covariance matrix given by

$$\Sigma_{\mathrm{GMM}} = \left(G_{Q_0}^{\dagger} JP\right) \Sigma_\gamma \left(G_{Q_0}^{\dagger} JP\right)^T \tag{34}$$

that is,

$$\sqrt{T}\left(\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right)-\mathrm{vec}(A)\right) \stackrel{d}{\rightarrow} \mathcal{N}_{n^2p}(0,\Sigma_{\mathrm{GMM}}) \tag{35}$$

Here, $Q_T \stackrel{p}{\rightarrow} Q_0$ where Q_0 is constant, symmetric, and positive definite and Σ_γ is the asymptotic covariance of the mixed-frequency covariances, described in Theorem 1, for the lags $(-p+1,\dots,np)$. Furthermore,

$$G^{\dagger}_{Q_0}=\left((Z_0\otimes I_n)Q_0\left(Z_0^T\otimes I_n\right)\right)^{-1}(Z_0\otimes I_n)Q_0$$

$$J=\begin{pmatrix} D & 0_{n\times n} & & 0_{n\times n} \\ 0_{n\times n} & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n\times n} \\ 0_{n\times n} & \cdots & 0_{n\times n} & D \end{pmatrix} \in \mathbb{R}^{n^2pn_f\times n(n+1)n_fp}$$

$$D=\Big(-A_p \quad 0_{n\times(n_f-1)n} \quad -A_{p-1} \quad 0_{n\times(n_f-1)n} \quad \dots \quad -A_1 0_{n\times(n_f-1)n} I_n 0_{n\times(n_f-1)n}\Big)$$

and the permutation matrix P is given as $P=\left(I_{(n+1)p}\otimes P_2\right)$, where

$$P_2=\left(I_{n_f}\otimes\begin{pmatrix} I_{n_f} \\ 0_{n_s\times n_f}\end{pmatrix}, \qquad I_{n_f}\otimes\begin{pmatrix} 0_{n_f\times n_s} \\ I_{n_s}\end{pmatrix}\right)$$

Proof. We commence with the observation that

$$\begin{aligned}
\sqrt{T}&\left(\mathrm{vec}\left(\hat{A}_{\mathrm{GMM}}\right)-\mathrm{vec}(A)\right) \\
&= \sqrt{T}\left(\hat{G}^{\dagger}_{Q_T}\mathrm{vec}\left(\hat{Z}_1\right)-\hat{G}^{\dagger}_{Q_T}\left(\hat{Z}_0^T\otimes I_n\right)\mathrm{vec}(A)\right) \\
&= \sqrt{T}\hat{G}^{\dagger}_{Q_T}\mathrm{vec}\left(\hat{Z}_1-A\hat{Z}_0-Z_1+AZ_0\right) \\
&= \sqrt{T}\hat{G}^{\dagger}_{Q_T}\underbrace{\left(I_{pnn_f}\otimes\left(I_n \quad -A\right)\right)}_{J_1}\mathrm{vec}\begin{pmatrix}\hat{Z}_1-Z_1 \\ \hat{Z}_0-Z_0\end{pmatrix} \\
&= \sqrt{T}\hat{G}^{\dagger}_{Q_T}\underbrace{J_1J_2}_{J}P\left(\begin{pmatrix}\mathrm{vec}\left(\hat{\gamma}^{z^f f}(i)\right) \\ \mathrm{vec}\left(\hat{\gamma}^{wf}(i)\right)\end{pmatrix}-\begin{pmatrix}\mathrm{vec}\left(\gamma^{z^f f}(i)\right) \\ \mathrm{vec}\left(\gamma^{wf}(i)\right)\end{pmatrix}\right)
\end{aligned}$$

for $i = -p+1, \ldots, np$, where J_2 is a reordering matrix. Since under our assumptions, the sample autocovariances are consistent estimators and $Q_T \xrightarrow{p} Q_0$, the same is true for $\hat{G}^{\dagger}_{Q_T} JP$, that is, $\hat{G}^{\dagger}_{Q_T} JP \xrightarrow{p} G^{\dagger}_{Q_0} JP$. Theorem 1 and Slutsky's Lemma then directly lead to the result of the theorem. ■

Remark 3. It is well known that under our assumptions the asymptotic covariance for the high-frequency Yule−Walker estimator is of the form $\left(\Gamma_p^{-1} \otimes \Sigma_\nu\right)$ (see Hannan, 1970; Lütkepohl, 2005) and thus, in this case, the fourth moment of the innovations does not influence the asymptotic covariance of the parameter estimates. In the case discussed here, the fourth moment of the innovations does not vanish under linearization in general.

Having obtained the expression for the asymptotic covariance, we can determine the asymptotically optimal weighting matrix for the GMM estimator.

Theorem 3. Under the assumptions of Theorem 2, the optimal asymptotic weighting matrix for the GMM estimator is

$$Q_0^* = \left(JP\Sigma_\gamma P^T J^T\right)^{-1} \tag{36}$$

and the corresponding asymptotic covariance is given by

$$\Sigma^*_{\mathrm{GMM}} = \left((Z_0 \otimes I_n) Q_0^* \left(Z_0^T \otimes I_n\right)\right)^{-1} \tag{37}$$

For the XYW estimator, where $Q_0 = I_{n^2 p n_f}$, the asymptotic covariance is given by

$$\Sigma_{\mathrm{XYW}} = \left(\left(Z_0^{\dagger}\right)^T \otimes I_n\right) JP\Sigma_\gamma P^T J^T \left(Z_0^{\dagger} \otimes I_n\right) \tag{38}$$

Proof. The proof of the theorem directly follows from theorem 3.2 in Hansen (1982). ■

Remark 4. Using the blocked process for the AR(1) case as described in Anderson et al. (2015) it is easy to derive the asymptotic covariance matrix of the maximum likelihood estimator for the mixed-frequency stock case (see Koelbl, 2015). In contrast to the high-frequency case, where under our assumptions the Yule−Walker estimator is asymptotically equivalent to the maximum likelihood estimator, the mixed-frequency

XYW estimator is, in general, not equivalent to the mixed-frequency maximum likelihood estimator (see Example 2).

Remark 5. In order to estimate the asymptotic covariance matrix of the XYW/GMM estimators as well as the asymptotically optimal weighting matrix Q_0^*, we have to estimate the fourth moment of (ν_t), unless we assume that (ν_t) has a Gaussian distribution, where the fourth moment does not occur. In Koelbl (2015), it is shown that the fourth moment can be reconstructed on a generic subset of Θ from

$$\mathrm{vec}(\kappa) = \left(\mathcal{G}_2 \left(I_{(np)^4} - \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A} \right)^{-1} \mathcal{G}_2^T \right)^{-1} \mathrm{vec}(\psi)$$

where $\mathcal{G}_2 = \mathcal{G} \otimes \mathcal{G} \otimes \mathcal{G} \otimes \mathcal{G}$ and $\psi = \mathbb{E}\left(y_t y_t^T \otimes y_t y_t^T \right) - \mathrm{vec}(\gamma(0))\mathrm{vec}(\gamma(0))^T - (\gamma(0) \otimes \gamma(0)) - K_{n,n}(\gamma(0) \otimes \gamma(0))$.

Remark 6. Until now, the asymptotic results obtained in this section are only valid for the stock case. Nevertheless, an adaption to the general case is straightforward: Using an obvious notation and following the same steps as in the proof of Theorem 2, we obtain

$$\sqrt{T}\left(\mathrm{vec}\left(\hat{A}^{\mathrm{g}}_{\mathrm{GMM}} \right) - \mathrm{vec}(A) \right) = \sqrt{T}\left(\hat{G}^{\mathrm{g}}_{Q_T} \right)^{\dagger} JP \left(\begin{pmatrix} \mathrm{vec}\left(\hat{\gamma}^{zff}(i) \right) \\ \mathrm{vec}\left(\hat{\gamma}^{wf}(i) \right) \end{pmatrix} - \begin{pmatrix} \mathrm{vec}\left(\gamma^{zff}(i) \right) \\ \mathrm{vec}\left(\gamma^{wf}(i) \right) \end{pmatrix} \right)$$

for $i = N - p, \ldots, N + np - 1$. The results concerning the asymptotic behavior of the covariance estimators, that is, Theorem 1 and Lemmas 1–3, are still valid.

5. SIMULATIONS

In our context the following issues arise: First, the comparison of the XYW estimators and the MLE-EM estimators. Second, the information loss caused by mixed-frequency data in relation to high-frequency data. Third, the information gain obtained by using mixed-frequency data when compared to low-frequency data. It should be emphasized here, that we only present preliminary results and in order to get a more complete picture further work is needed.

It is intuitively clear that the quality of the mixed-frequency estimators depends on N and on n_f, for example, because the number of summands in Eq. (12) depends on N, see Example 3. Moreover, the quality of the mixed-frequency estimators depends on the underlying parameters of the high-frequency model. In particular, if we are close to the (mixed-frequency) nonidentifiable subset in the parameter space, a large information loss due to mixed-frequency data in comparison to high-frequency data can be expected. One way to measure information loss in our context would be to compare the (asymptotic) covariances of the MLEs from mixed-frequency data (MF-MLE) with the (asymptotic) covariances of the MLEs from high-frequency data (HF-MLE). Another way would be to compare the one-step-ahead prediction error covariances. This is of particular importance for comparisons to the low-frequency case, where in general identifiability cannot be achieved. In order to demonstrate the effects of being close to the nonidentifiable subset, we consider a simple example:

Example 1. Assume that $p = 1$, $n_f = n_s = 1$, $N = 2$, $\Sigma_\nu = I_2$ and the case of stock variables. The system parameters $A_1 = \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}$ are not identifiable if and only if they satisfy the equations $a_{fs} = 0$, $a_{sf} = 0, a_{ss} \neq 0$ (see Anderson et al., 2015). For the models

$$y_t = \begin{pmatrix} 0.9 & 0 \\ a_{sf} & 0.8 \end{pmatrix} y_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}_2(0, I_2) \tag{39}$$

and $a_{sf} \in \{0, 0.01, 0.1, 0.25\}$, we obtain the following sections of the likelihoods as shown in Fig. 1 where we only vary over a_{ss}.

Table 1 reports the mean squared errors (MSE) of the XYW estimator corresponding to (13) and the MLE-EM, initialized by the XYW estimator, for the model class described above for four different values of a_{sf}. It shows that also close to the identifiability boundary problems for the estimators arise. The last two columns show the relative number of hits of the estimators for a_{ss} for the intervals $[-0.9, -0.7]$ and $[0.7, 0.9]$, respectively. Note that, in particular, for the nonidentifiable case $a_{sf} = 0$, the MLE-EM gives estimates close to the class of equivalent system parameters

$$\left\{ \begin{pmatrix} 0.9 & 0 \\ 0 & 0.8 \end{pmatrix}, \begin{pmatrix} 0.9 & 0 \\ 0 & -0.8 \end{pmatrix} \right\}$$

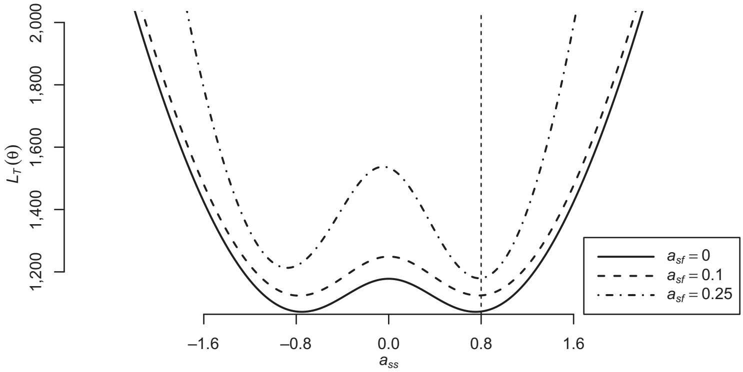

Fig. 1. Sections of the Likelihood Functions $L_T(\theta)$ for Three Different Values of a_{sf}.

Table 1. Comparison of XYW and MLE-EM Estimators in Terms of Mean-Squared Errors and Hits.

a_{sf}	Estimators	MSE $\hat{a}_{ff}$	MSE $\hat{a}_{sf}$	MSE $\hat{a}_{fs}$	MSE $\hat{a}_{ss}$	$[-0.9, -0.7]$	$[0.7, 0.9]$
0	XYW	0.131	0.086	25.335	1.698	0.04	0.05
	MLE-EM	0.004	0.007	0.001	1.160	0.43	0.51
0.01	XYW	0.124	0.085	11.135	1.589	0.05	0.06
	MLE-EM	0.001	0.006	0.001	1.100	0.42	0.55
0.1	XYW	0.002	0.016	0.016	0.061	0.00	0.26
	MLE-EM	0.001	0.001	0.001	0.001	0.00	1.00
0.25	XYW	0.001	0.030	0.003	0.028	0.00	0.39
	MLE-EM	0.001	0.003	0.001	0.004	0.00	1.00

Furthermore, in this case the matrix Z_0 is singular and in particular the solution set of Eq. (9) is given as

$$\begin{pmatrix} 0.9 & d_1 \\ 0 & d_2 \end{pmatrix}$$

where d_1 and d_2 are arbitrary. The MSEs of the XYW estimator for a_{fs} are relatively large compared to the MLE-EM in this case.

Also, on an intuitive level, the memory of the data generating process is assumedly important for the information loss discussed above. This is demonstrated in Example 2:

Example 2. Consider the following two models:
Model 1:

$$y_t = \begin{pmatrix} 0.9556 & 0.8611 \\ -0.6914 & 0.2174 \end{pmatrix} y_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}_2(0, I_2) \tag{40}$$
$$z_{0,1} = 0.7303 \pm 0.8437i$$

Model 2:

$$y_t = \begin{pmatrix} -1.2141 & 1.1514 \\ -0.9419 & 0.8101 \end{pmatrix} y_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}_2(0, I_2) \tag{41}$$
$$z_{0,1} = -2 \pm 2.4294i$$

where $z_{0,1}$ denotes the roots of the determinant of the autoregressive polynomial. The correlations of the two processes are depicted in Fig. 2 where the black bars are the unobserved autocorrelations.

In comparing these two models, Table 2 shows the MSE

$$\mathrm{MSE}\left(\hat{\theta}\right) = \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{7} \left(\theta_i - \hat{\theta}_i^j\right)^2$$

for the parameters $\theta = \left(\mathrm{vec}(A)^T, \mathrm{vech}(\Sigma_\nu)^T\right)^T$ for $T = 500$, $m = 10^3$ simulation runs, $N = 2$ and the case of stock variables. Here, HF-YW is the standard Yule–Walker estimator from high-frequency data, HF-XYW is the

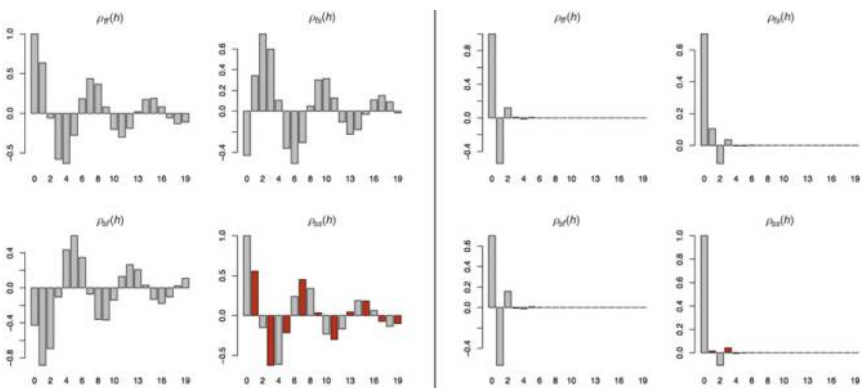

Fig. 2. Autocorrelations of Model 1 (left) and Model 2 (right).

Table 2. Absolute and Relative Mean-Squared Errors of the System and Noise Parameters.

	Estimators	Model 1		Model 2	
		Absolute	Relative	Absolute	Relative
HF	YW	0.007	1	0.009	1
	XYW	0.012	1.77	0.272	30.07
MF	MLE-EM	0.010	1.53	0.977	107.93
	XYW	0.076	11.70	2.977	329.69

Table 3. Absolute and Relative Root Mean-Squared One-Step-Ahead Forecasting Errors.

	Estimators	Model 1		Model 2	
		Absolute	Relative	Absolute	Relative
LF	YW	3.607	1	2.859	1
MF	MLE-EM	2.371	0.66	2.857	0.99
	XYW	2.388	0.66	34.149	11.94
HF	YW	1.995	0.55	1.998	0.70

estimator obtained from inserting the sample second moments obtained from high-frequency data into the XYW estimator corresponding to the pseudo-inverse, see Eq. (13). MF-MLE-EM is the estimator described in Section 2.3, initialized by the MF-XYW estimator, and MF-XYW is the XYW estimator described in Eq. (13). In addition, as a measure for the information loss, the MSE relatively to the MSE of the HF-YW estimators are presented. In particular, the relative MSE of the MF-MLE-EMs show the information loss due to mixed-frequency data. However, convergence problems due to the existence of local minima of the likelihood may arise in calculating the MF-MLE-EM. This can be mitigated by using several starting values.

In Table 3, an analogous comparison based on one-step-ahead prediction errors is given. This table relates the mixed-frequency prediction errors to the prediction errors obtained by using Yule−Walker equations in the high-frequency case as well as the prediction errors obtained by using Yule−Walker equations in the low-frequency case. In the high-frequency case, the one-step-ahead forecast of y_t, $t \in 2\mathbb{Z}$ is based on y_{t-1}, in the low-frequency case the forecast of y_t, $t \in 2\mathbb{Z}$, is based on y_{t-2} and finally in the mixed-frequency case the forecast of y_t, $t \in 2\mathbb{Z}$ is based on y^f_{t-1}, y_{t-2}.

In Table 4, the absolute and relative Frobenius norms of the asymptotic covariance matrices of the estimators of the system parameters are

Table 4. Absolute and Relative Norms of the Asymptotic Covariance Matrix of the System Parameters.

	Estimators	Model 1		Model 2	
		Absolute	Relative	Absolute	Relative
HF	YW	0.421	1	1.416	1
	XYW	0.516	1.23	13.690	9.67
	XYW $k = 1$	0.558	1.33	18.830	13.29
	GMM $k = 1$	0.514	1.22	12.674	8.95
MF	MLE	0.623	1.48	2.632	1.86
	XYW	1.504	3.57	112.907	79.72
	XYW $k = 1$	0.942	2.24	163.694	115.57
	GMM $k = 1$	0.878	2.09	85.775	60.56

presented. Note that, for instance, HF-XYW $k = 1$ corresponds to the high-frequency XYW estimator based on Eq. (15), that is, the XYW equations are extended with a further lag. We observe that the MF-XYW estimators have larger asymptotic covariances than the MF-MLE. This is the case even if the optimal weights are chosen for the MF-GMM $k = 1$ estimator.

In addition, it is clear that for the same high-frequency model increasing n_f will give better results for the mixed-frequency estimators and that the quality of the parameter estimators will decrease with increasing N. We demonstrate these effects in Example 3.

Example 3. In order to demonstrate the relations between the MF-XYW and the MF-MLE-EM estimator as well as the effects of increasing n_f and N for the stock case, we consider the following Model 3 for $T = 500$ and $m = 10^3$ simulation runs:

$$y_t = \begin{pmatrix} 0.9154 & 0.1002 & 0.2250 & -0.3594 \\ 2.7553 & 1.5950 & 3.3705 & -5.4438 \\ 0.4516 & -0.1998 & 0.8294 & -0.7917 \\ 0.7375 & 0.1185 & 0.7489 & -0.6667 \end{pmatrix} y_{t-1} + \nu_t,$$

$$\nu_t \sim \mathcal{N}_4\left(0, bb^T\right) \tag{42}$$

$$b = \begin{pmatrix} 1.1140 & 0 & 0 & 0 \\ -0.3807 & 0.6514 & 0 & 0 \\ 0.3448 & -0.3742 & 0.3103 & 0 \\ -0.1749 & -0.1389 & -0.2241 & 1.317 \end{pmatrix}$$

$$z_0 = -0.8783, \quad z_1 = -0.7983, \quad z_{2,3} = -0.1557 \pm 0.7614i$$

Table 5. Absolute and Relative Mean-Squared Errors for the Respective Parameter Estimators for Model 3.

Estimators	$N = 2, n_f = 2$		$N = 2, n_f = 3$		$N = 3, n_f = 3$		$N = 12, n_f = 3$	
	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative
MLE-EM	0.230	1	0.014	1	0.015	1	0.984	1
XYW	0.774	3.37	0.381	28.04	1.592	104.74	5.740	5.83

Table 6. Absolute and Relative Mean-Squared Errors of the System and Noise Parameters.

	Estimators	Model 1		Model 2	
		Absolute	Relative	Absolute	Relative
HF	YW	0.007	1	0.009	1
	XYW	0.020	2.86	0.545	60.56
MF	MLE-EM	0.043	6.14	0.123	13.67
	XYW	0.070	10.70	1.472	161.77

Here, the MF-MLE-EM clearly outperforms the MF-XYW estimator for all the four cases shown in Table 5. For both estimators, the quality decreases with increasing N and increases with increasing n_f.

Example 4. In this example, we consider again Model 1 and Model 2 as introduced in Example 2 for $T = 500$ and $m = 10^3$ simulation runs but now for the case of flow variables, that is, $w_{2t} = y^s_{2t} + y^s_{2t-1}$. Here, HF-XYW is the estimator obtained from inserting the sample second moments obtained from high-frequency data, that is, $y^f_t \in \mathbb{Z}$ and $w_t \in \mathbb{Z}$, into the XYW estimator corresponding to the pseudo-inverse, see Eq. (19). As can be seen in Table 6, the estimators for the flow case do not necessarily lead to better estimates compared to the stock case.

6. OUTLOOK AND CONCLUSIONS

In this paper, we discussed and analyzed estimators for the parameters of a high-frequency VAR model from mixed-frequency data where the low-frequency data are obtained from general linear aggregation schemes including, in particular, stock and flow data. We considered estimators obtained from the XYW equations with different weighting matrices as well as Gaussian maximum likelihood type estimators based on the EM algorithm. The problem of getting estimators resulting in stable systems and positive (semi)-definite covariances of prescribed rank q has been treated. Furthermore, we derived the asymptotic distribution of the XYW/GMM estimators. Finally, we presented a simulation study comparing XYW and Gaussian maximum likelihood estimators and discussed the information loss due to mixed-frequency data compared to high-frequency data and the information gain if we use mixed-frequency data rather than low-frequency data.

In particular, the dependence of the results obtained from the point in parameter space for high-frequency AR systems chosen needs further investigation.

ACKNOWLEDGMENTS

Support by the FWF (Austrian Science Fund under contract P24198/N18) is gratefully acknowledged. We thank Prof. Tommaso Proietti, Universita di Roma "Tor Vergata", Italy, and Prof. Brian D. O. Anderson, Australian National University, Australia, for helpful comments.

REFERENCES

Anderson, B. D. O., Deistler, M., Felsenstein, E., Funovits, B., Koelbl, L., & Zamani, M. (2015). Multivariate AR systems and mixed frequency data: g-Identifiability and estimation. *Econometric Theory*, doi:10.1017/S0266466615000043:1–34. Retrieved from http://dx.doi.org/10.1017/S0266466615000043

Anderson, B. D. O., Deistler, M., Felsenstein, E., Funovits, B., Zadrozny, P. A., Eichler, M., …, Zamani, M. (2012). Identifiability of regular and singular multivariate autoregressive models. *Proceedings of the 51th IEEE conference on decision and control (CDC)*, pp. 184–189.

Anderson, T. (1994). *The statistical analysis of time series*. New York, NY: Wiley.

Balogh, L., & Pintelon, R. (2008). Stable approximation of unstable transer function models. *IEEE Transactions on Instrumentations and Measurement*, *57*(12), 2720–2726.

Chen, B., & Zadrozny, P. A. (1998). An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data. *Advances in Econometrics*, *13*, 47–73.

Combettes, P. L., & Trussell, H. J. (1992). Best stable and invertible approximations for ARMA systems. *IEEE Transactions on Signal Processing*, *40*:3066–3069.

Deistler, M., Anderson, B. D. O., Filler, A., Zinner, C., & Chen, W. (2010). Generalized linear dynamic factor models — An approach via singular autoregressions. *European Journal of Control*, *16*(3), 211–224.

D'haene, T., Pintelon, R., & Vandersteen, G. (2006). An iterative method to stabilize a transfer function in the s- and z-domains. *IEEE Transactions on Instrumentations and Measurement*, *55*, 1192–1196.

Filler, A. (2010). *Generalized dynamic factor models Structure theory and estimation for single frequency and mixed frequency data*. PhD thesis, Vienna University of Technology.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, *82*(4), 540–554.

Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, *185*, 359–371.

Francq, C., & Zakoian, J.-M. (2009). *Bartlett's formula for a general class of non linear processes*.

Gingras, D. F. (1985). Asymptotic properties of high-order Yule-Walker estimates of the AR parameters of an ARMA time series. *IEEE Transactions on Acoustics Speech and Signal Processing*, *33*(4), 1095–1101.

Hall, P., & Heyde, C. (1980). *Martingal limit theory and its application*, New York, NY: Academic Press.

Hannan, E. J. (1970). *Multiple time series*, New York, NY: Wiley.

Hannan, E. J., & Deistler, M. (2012). *The statistical theory of linear systems*, Philadelphia, PA: SIAM Classics in Applied Mathematics.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Hoffman, A., & Wielandt, H. (1953). The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, *20*, 37–39.

Khatri, C. G. (1961). Some results for the singular normal multivariate regression models. *The Indian Journal of Statistics*, *30*, 267–280.

Koelbl, L. (2015). *VAR systems: g-Identifiability and asymptotic properties of parameter estimates for the mixed-frequency case*. PhD thesis, Vienna University of Technology.

Komunjer, I., & Ng, S. (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, *79*(6), 1995–2032.

Lippi, M., & Reichlin, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics*, *63*, 307–325.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*, Berlin: Springer.

Magnus, J. R., & Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Annals of Statistics*, *7*(2), 381–394.

Mariano, R. S., & Murasawa, Y. (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics*, *72*(1), 27–46.

Moses, L. R., & Liu, D. (1991). Determining the closest stable polynomial to an unstable one. *IEEE Transactions on Signal Processing*, *39*(4), 901–906.

Niebuhr, T., & Kreiss, J.-P. (2013). Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies. *International Statistical Review*, *82*(1), 123–140.

Orbandexivry, F.-X., Nesterov, Y., & van Dooren, P. (2013). Nearest stable system using successive convex approximations. *Automatica*, *49*, 1195–1203.

Rao, C. R. (1972). *Linear statistical inference and its applications*, New York, NY: Wiley.

Shumway, R., & Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, *3*(4), 253–264.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Associations*, *97*(460), 1167–1179.

Stoica, P., & Moses, L. R. (1992). On the unit circle problem: The Schur-Cohn procedure revisited. *Signal Processing*, *26*, 95–118.

Su, N., & Lund, R. (2011). Multivariate versions of Bartlett's formula. *Journal of Multivariate Analysis*, *105*, 18–31.

Wax, M., & Kailath, T. (1983). Efficient inversion of Toeplitz-Block Toeplitz matrix. *IEEE Transactions on Acoustics Speech and Signal Processing*, *31*, 1218–1221.

APPENDIX

The next lemma, see Su and Lund (2011), gives a multivariate version of Bartlett's formula.

Lemma 1. Under the assumptions of Section 4, we obtain

$$\lim_{T\to\infty} T\mathrm{Cov}\left(\mathrm{vec}\left(\hat{\gamma}^{z^f f}(p)\right), \mathrm{vec}\left(\hat{\gamma}^{z^f f}(q)\right)\right) = S_{p,q} + R_{p,q}$$

for $p, q \in \mathbb{Z}$, where

$$R_{p,q} = \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(k+q-p) \otimes \gamma^{z^f z^f}(k)\right) + K_{n_f,n_f}\left(\gamma^{z^f f}(k+q) \otimes \gamma^{f z^f}(k-p)\right)$$
$$S_{p,q} = \sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \left(k^f_{k-p} \otimes \tilde{k}^f_k\right) \kappa \left(k^f_{r+k-q} \otimes \tilde{k}^f_{r+k}\right)^T$$

and K_{n_f,n_f}, $K_{n,n}$ are commutation matrices.

Note that in the Gaussian case $\kappa = 0$ and thus $S_{p,q}$ is zero. Using the idea of the proof of Su and Lund (2011) and taking into account that $\hat{\gamma}^{wf}(q)$ has only approximately T/N summands, we obtain:

Lemma 2. Under the assumptions of Section 4, we obtain

$$\lim_{T\to\infty} T\mathrm{Cov}\left(\mathrm{vec}\left(\hat{\gamma}^{z^f f}(p)\right), \mathrm{vec}\left(\hat{\gamma}^{wf}(q)\right)\right) = \overline{S}_{p,q} + \overline{R}_{p,q}$$

for $p, q \in \mathbb{Z}$, where

$$\overline{R}_{p,q} = \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(k+q-p) \otimes \gamma^{z^f w}(k)\right) + K_{n_f,n_f}\left(\gamma^{z^f f}(k+q) \otimes \gamma^{fw}(k-p)\right)$$

$$\overline{S}_{p,q} = \sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \left(k^f_{k-p} \otimes \tilde{k}^f_k\right) \kappa \left(k^f_{r+k-q} \otimes \tilde{k}^s_{r+k}\right)^T$$

Replacing $\hat{\gamma}^{wf}(q)$ in Lemma 2 by the "high-frequency autocovariance estimator", that is, by $(1/T)\sum_{t=1}^{T} w_t\left(y_{t-q}^{f}\right)^T$, the result is still valid. The following result generalizes the result given in Niebuhr and Kreiss (2013) to the multivariate case.

Lemma 3. Under the assumptions of Section 4, we obtain

$$\lim_{T\to\infty} T\mathrm{Cov}\left(\mathrm{vec}\left(\hat{\gamma}^{wf}(p)\right), \mathrm{vec}\left(\hat{\gamma}^{wf}(q)\right)\right) = N\left(\tilde{S}_{p,q} + \tilde{R}_{p,q}\right)$$

for $p, q \in \mathbb{Z}$, where

$$\tilde{R}_{p,q} = \sum_{k=-\infty}^{\infty} \left(\gamma^{ff}(Nk+q-p)\otimes\gamma^{ww}(kN)\right) + K_{n_f,n_s}\left(\gamma^{wf}(Nk+q)\otimes\gamma^{fw}(Nk-p)\right)$$
$$\tilde{S}_{p,q} = \sum_{k=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} \left(k_{k-p}^{f}\otimes\tilde{k}_{k}^{s}\right)\kappa\left(k_{Nr+k-q}^{f}\otimes\tilde{k}_{Nr+k}^{s}\right)^T$$

Note that for $N = 1$ we still obtain Bartlett's formula for the high-frequency case. The three lemmas above are needed for the following proof of Theorem 1.

Proof of Theorem 1. W.l.o.g. let us assume that T is a multiple of N. We will prove this theorem for $\hat{\hat{\gamma}}^{zf}(i) = (1/T)\sum_{t=1}^{T} z_t^f\left(y_{t-i}^{f}\right)^T$ and $\hat{\hat{\gamma}}^{wf}(i) = (N/T)\sum_{t=1}^{T/N} w_{Nt}\left(y_{Nt-i}^{f}\right)^T$ instead of $\hat{\gamma}^{zf}(i)$ and $\hat{\gamma}^{wf}(i)$, respectively, since it can be shown that this change does not influence the asymptotic properties, see Hannan (1970). In order to apply theorem 14 in Hannan (1970, p. 228), we define a particular blocked process $u_t = \left(y_{Nt}^T \; y_{Nt-1}^T \; \cdots \; y_{Nt-(N-1)}^T\right)^T$ which satisfies the assumptions of this theorem. Now applying theorem 14 in Hannan (1970) leads to

$$\sqrt{\frac{T}{N}}\left(\mathrm{vec}\left(\hat{\gamma}_u(i)\right) - \mathrm{vec}\left(\gamma_u(i)\right)\right)_{i=0,\ldots,s} \xrightarrow{d} N_{nN(s+1)}(0, \Sigma_u)$$

where $\gamma_u(i)$ is the population autocovariance of the process (u_t) for lag i and $\hat{\gamma}_u(i)$ is its sample counterpart with T/N summands. In a last step we have to find a transformation matrix, say H, which transforms $\hat{\gamma}_u$ to the desired autocovariances. To obtain this transformation we define $H_1 = \left((1/N) I_{n_f^2} \;\; (1/N) I_{n_f^2} \;\; \cdots \;\; (1/N) I_{n_f^2} \right)$ and observe that for $j = 0, \ldots, s$ $\text{vec}\left(\hat{\hat{\gamma}}^{wf}(j)\right) = (N/T) \sum_{t=1}^{T/N} \text{vec}\left(w_{Nt} \left(y^f_{Nt-j} \right)^T \right) = S_j^{wf} \left(\text{vec}\left(\hat{\gamma}_u(i) \right) \right)_{i=0,\ldots,s}$ and

$$\text{vec}\left(\hat{\hat{\gamma}}^{zff}(j)\right) = H_1 \begin{pmatrix} \frac{N}{T} \sum_{t=1}^{T/N} \text{vec}\left(z^f_{Nt} \left(y^f_{Nt-j} \right)^T \right) \\ \frac{N}{T} \sum_{t=1}^{T/N} \text{vec}\left(z^f_{Nt-1} \left(y^f_{Nt-1-j} \right)^T \right) \\ \vdots \\ \frac{N}{T} \sum_{t=1}^{T/N} \text{vec}\left(z^f_{Nt-(N-1)} \left(y^f_{Nt-(N-1)-j} \right)^T \right) \end{pmatrix}$$
$$= H_1 S_j^{zff} \left(\text{vec}\left(\hat{\gamma}_u(i) \right) \right)_{i=0,\ldots,s}$$

where S_j^{zff} and S_j^{wf} are selector matrices for lag j. Finally, we can construct our particular transformation matrix $H = \left(\left(H_1 S_0^{zff} \right)^T, \left(S_0^{wf} \right)^T, \ldots, \left(H_1 S_s^{zff} \right)^T, \left(S_s^{wf} \right)^T \right)^T$ and obtain the desired result

$$\sqrt{\frac{T}{N}} \left(\begin{pmatrix} \text{vec}\left(\hat{\gamma}^{zff}(i) \right) \\ \text{vec}\left(\hat{\gamma}^{wf}(i) \right) \end{pmatrix} - \begin{pmatrix} \text{vec}\left(\gamma^{zff}(i) \right) \\ \text{vec}\left(\gamma^{wf}(i) \right) \end{pmatrix} \right)_{i=0,\ldots,s}$$
$$= \sqrt{\frac{T}{N}} H \left(\text{vec}(\hat{\gamma}_u(i)) - \text{vec}\left(\gamma_u(i) \right) \right)_{i=0,\ldots,s} \xrightarrow{d} \mathcal{N}_h\left(0, \Sigma_\gamma \right)$$

The asymptotic covariance $\Sigma_\gamma = H \Sigma_u H^T$ can be derived using Lemmas 1–3. ■