Econometric Theory, **0**, 2015, 1–34.
doi:10.1017/S0266466615000043

MULTIVARIATE AR SYSTEMS AND MIXED FREQUENCY DATA: G-IDENTIFIABILITY AND ESTIMATION

Brian D.O. Anderson
Australian National University and National ICT Australia Ltd.

Manfred Deistler
Vienna University of Technology and
Institute for Advanced Studies

Elisabeth Felsenstein
Vienna University of Technology

Bernd Funovits
University of Vienna and Vienna University of Technology

Lukas Koelbl
Vienna University of Technology

Mohsen Zamani
Australian National University

This paper is concerned with the problem of identifiability of the parameters of a high frequency multivariate autoregressive model from mixed frequency time series data. We demonstrate identifiability for generic parameter values using the population second moments of the observations. In addition we display a constructive algorithm for the parameter values and establish the continuity of the mapping attaching the high frequency parameters to these population second moments. These structural results are obtained using two alternative tools viz. extended Yule Walker equations and blocking of the output process. The cases of stock and flow variables, as well as of general linear transformations of high frequency data, are treated. Finally, we briefly discuss how our constructive identifiability results can be used for parameter estimation based on the sample second moments.

The authors want to thank Hans Havlicek, Vienna University of Technology, and Benedikt Pötscher, University of Vienna, for valuable suggestions concerning the proof of Theorem 2. In addition we thank Peter Phillips, several anonymous referees and handling co-editors for valuable comments. *Query 1*
Support by the FWF (Austrian Science Fund under contracts P20833/N18 and P24198/N18), the ARC (Australian Research Council under Discovery Project Grant DP1092571) and NICTA is gratefully acknowledged. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the ARC through the ICT Centre of Excellence program. Address correspondence to Manfred Deistler is with the Institute of Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria, and Institute for Advanced Studies, Vienna, Austria; e-mail: deistler@tuwien.ac.at.

1. INTRODUCTION

In many applications involving multivariate time series data, we encounter mixed frequency (MF) data, i.e., data where the univariate component time series are available at different sampling frequencies. This is particularly true for economic applications, where for instance unemployment data may be available monthly, whereas GNP data are available quarterly. To give a more general example, usually, financial data are observed more frequently than real data. Clearly too, the higher the dimensionality of the time series in an applications context, the more likely are mixed frequency data to occur. Given the frequent occurrence of MF data and the desire to use as much as possible or indeed all the information contained in these data, it is not surprising that a number of different approaches to MF problems such as parameter estimation have been developed (see e.g. Harvey and Pierse, 1984; Nijman, 1985; Kohn and Ansley, 1986; Zadrozny, 1990b; Bernanke, Gertler, Watson, Sims, and Friedman, 1997; Chen and Zadrozny, 1998; Marcellino, 1998; Mariano and Murasawa, 2003; Aruoba, Diebold, and Scotti, 2007; Wohlrabe, 2008; Ghysels and Wright, 2009; Marcellino and Schumacher, 2010; Ghysels, 2012). These approaches differ as far as their final aims, the model classes considered and the estimation procedures developed are concerned.

Most of the traditional procedures for model parameter estimation in multivariate time series assume that the individual components of the time series are available at a single frequency. This motivates one of the two basic approaches to treat MF data: The first approach is to transform the data to a single frequency before parameter estimation. In contrast, the second approach is to directly use MF data for estimation, which requires some form of new or less familiar methodology. As far as the first approach is concerned, a simple method to deal with MF data is to transform all data to the lowest sampling frequency, for instance by taking all observations at the lowest frequency, and then to estimate a model. Of course, this is straightforward but leads to loss of information. Another possibility to apply estimation methods for single frequency data is to (directly) interpolate the low frequency data, so that after such a transformation high frequency "data" are available for parameter estimation. However, exactly how one should interpolate is not clear, the more so since such interpolation is required before the model parameters have been estimated.

As far as the second basic approach, namely direct estimation from mixed frequency data, is concerned, again several different procedures have been developed: One approach is to estimate a continuous-time model from discrete-time data (see e.g. Phillips, 1973, 1976; Hansen and Sargent, 1983; Bergstrom, 1988; Zadrozny, 1988). Of course, once a continuous-time model is available, discrete-time data can be generated at arbitrary sampling frequencies, see e.g. Chambers and Thornton (2012). However, there are costs associated with this approach. Firstly, in a stationary context, the spectral density of a single frequency discrete-time process is obtained by the so-called aliasing formula from the continuous-time spectral density (see e.g., Hannan, 1970) and, in a "nonparametric" context, the function attaching the discrete-time spectrum to the continuous-time spectrum is not injective. Hence restrictions to the continuous-time system have to be imposed in order to guarantee uniqueness from discrete-time data as has been pointed out already in the famous Nyquist–Shannon sampling theorem. Secondly, there is no easy relation between *model classes* of related discrete-time and continuous-time systems. This observation is sometimes called the embedding problem, see Brockwell (1995). For instance, some discrete-time AR(1) systems correspond to continuous-time AR(1) systems, other to continuous-time ARMA(2,1) systems.

The most frequent approaches to direct model estimation from MF data are MIDAS (mixed-data sampling) regression, where we have a mixed frequency model (see e.g. Ghysels, Santa-Clara, and Valkanov, 2006; Ghysels, Sinko, and Valkanov, 2007). Another approach is to use Kalman filtering to generate missing observations (see e.g. Zadrozny, 1990a; Mariano and Murasawa, 2003). Of course, the use of an optimal Kalman filter for reconstructing missing observations requires knowledge of the underlying parameters; so there is some circularity in the logic of this approach, but this can be dealt with through an iterative loop.

In this paper we consider the following setting: We assume a high frequency, multivariate autoregressive model generating the outputs. Some outputs are not observed at the highest frequency but instead at an integer submultiple of the highest frequency or, more generally, a linear transformation of these outputs is available, thus leading to MF data. Our main focus is on the discussion of the case of stock variables. No a priori restrictions except for those concerning the order of the system, stability, and the rank of the innovation variance are imposed. The final aim of the analysis considered here is to directly estimate the parameters of the underlying high frequency model from the available MF data. Clearly, once these parameters are available, then also all variances and covariances of the full high frequency output process are available and, based on those second moments, linear least squares approximation problems like interpolation, nowcasting, and forecasting can be solved. The analysis given here is focused on identifiability of the parameters of the underlying high frequency model from the second moments of the observations. The main result given in the paper is that generically identifiability can be ensured using these second moments. We will use the term g-identifiability for generic identifiability. Our results are "constructive" in the sense that we present "realization algorithms", i.e., algorithms which give the parameters from the population second moments of the observations. In addition it is shown that the problem is well-posed in the sense that the parameters depend continuously on these second moments. Thus, when these population second moments are replaced by their sample counterparts, this leads to consistent estimation procedures. We consider that identifiability and well-posedness are fundamental for understanding certain estimation problems, as has been argued in detail in Hannan and Deistler (2012). A more abstract analysis is presented in Deistler and Seifert (1978).

The techniques we use in the paper draw heavily from Chen and Zadrozny (1998) and Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, and Zamani (2012) and the approach given for blocking is related to Ghysels (2012). The paper is organized as follows: In Section 2 the setting and the problem are described. Section 3 deals with the special case of AR(1) models for the high frequency process where, for simplicity, we only treat the bivariate case. This case is simpler than the general case, since here missing observations do not change the autoregressive structure and an explicit description of the (generic) set of all identifiable parameters can be derived. Section 4 deals with general AR(p) systems. In this setting we have not been able to give the explicit description mentioned above. However, as a "second best" result we do show g-identifiability, which we consider is still a strong result. Actually we obtained an even stronger result namely that identifiability holds on the complement of a proper algebraic variety. Whereas in most parts of the paper only the case of stock variables has been considered, the genericity results are extended to general (linear) aggregation schemes, covering flow and stock variables, in Section 5. Section 6 is concerned with obtaining additional identifiability results by prescribing column degrees, rather than merely the maximum lag, for the autoregression. In Sections 4–6 the results are based on extended Yule Walker equations. In Section 7 we introduce a realization algorithm based on blocking the observations in order to obtain g-identifiability. Blocking of observations has been considered in Filler (2010), Ghysels (2012), Chen, Anderson, Deistler, and Filler (2012), and Zamani (2014). Section 8 provides an outlook and a conclusion. The Appendix consists of the proofs of Theorems 1, 2, 3, 5, 6, 7, and 8.

2. HIGH FREQUENCY AR SYSTEMS AND MIXED FREQUENCY DATA

We consider the case where the high frequency system, i.e., the system generating all outputs at the highest frequency, is a vector autoregression of order p. Let

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \nu_t,\ t \in \mathbb{Z}, \tag{2.1}$$

where y_t is n-dimensional, $A_i \in \mathbb{R}^{n \times n}$ and $(\nu_t \mid t \in \mathbb{Z})$ is n-dimensional white noise. Throughout we assume that the high frequency system (2.1) is stable, i.e., that

$$\det(a(z)) \neq 0\ |z| \leq 1, \tag{2.2}$$

where $a(z) = I - A_1 z - \cdots - A_p z^p$. Here z is used for the complex variable as well as for the backward shift on the integers $\mathbb{Z}$.

We consider the case where the innovation variance $\Sigma_\nu = \mathbb{E}\left(\nu_t \nu_t^T\right)$ is nonsingular as well as the case where this variance is singular. Then the corresponding autoregressive systems are called regular and singular, respectively. Singular autoregressive systems are important as models for latent variables and the corresponding static factors in generalized dynamic factor models (GDFMs) (see Forni, Hallin, Lippi, and Reichlin, 2000; Stock and Watson, 2002; Doz, Giannone, and Reichlin, 2011). The innovations of such autoregressive systems are dynamic factors. Singular AR systems occur if the dimension of the minimal static factors is larger than the dimension of the minimal dynamic factors, see Deistler, Anderson, Filler, Zinner, and Chen (2010). The case where singularity is caused by definitional equations is discussed in Phillips (1974).

We assume the lag order p and the rank q of Σ_ν with $1 \leq q \leq n$ to be known. The parameter space for the high frequency models considered is:

$$\Theta = \left\{ \left(A_1, \ldots, A_p\right) \mid \det\left(a(z)\right) \neq 0,\ |z| \leq 1 \right\} \\ \times \left\{ \Sigma_\nu \mid \Sigma_\nu = \Sigma_\nu^T,\ \Sigma_\nu \geq 0, rk\left(\Sigma_\nu\right) = q \right\}.$$

Note that the corresponding set of system parameters

$$\mathcal{S} = \left\{ \left(A_1, \ldots, A_p\right) \mid \det\left(a(z)\right) \neq 0,\ |z| \leq 1 \right\}$$

is open in $\mathbb{R}^{n \times np}$. Also note that we have no additional a priori restrictions, unless the contrary is stated explicitly.

If Σ_ν is of rank $q \leq n$, we can write Σ_ν as $\Sigma_\nu = bb^T$ 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, Prop. 3.1.2).

System (2.1) can be written in companion form as

$$\underbrace{\begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}}_{=x_{t+1}} = \underbrace{\begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_n & & & \\ & \ddots & & \\ & & I_n & 0 \end{pmatrix}}_{=\mathcal{A}} \underbrace{\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}}_{=x_t} + \underbrace{\begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=\mathcal{B}} \varepsilon_t, \tag{2.3}$$

$$y_t = \underbrace{\left(A_1 \cdots A_p\right)}_{\mathcal{C}} x_t + b\varepsilon_t. \tag{2.4}$$

We restrict ourselves to the steady state and thus causal and stationary solution. The solution of (2.1) or (2.3), (2.4) is of the form

$$y_t = a(z)^{-1} b\varepsilon_t = \left(\mathcal{C}\left(I - \mathcal{A}z\right)^{-1} \mathcal{B}z + b\right)\varepsilon_t, \tag{2.5}$$

where $k(z) = a(z)^{-1}b = \mathcal{C}\left(I - \mathcal{A}z\right)^{-1}\mathcal{B}z + b$ is the transfer function from (ε_t) to (y_t). The Lyapunov equation, where $\Gamma_p = \mathbb{E}\left(x_t x_t^T\right)$, for the system (2.3) is

$$\Gamma_p - \mathcal{A}\Gamma_p\mathcal{A}^T = \mathcal{B}\mathcal{B}^T. \tag{2.6}$$

Under the stability assumption (2.2) the system (2.3), (2.4) has a unique stationary solution, for given $(\mathcal{A}, \mathcal{B})$, and the variance of x_t can be expressed as

$$\Gamma_p = \sum_{j=0}^{\infty} \mathcal{A}^j \mathcal{B}\mathcal{B}^T \left(\mathcal{A}^T\right)^j. \tag{2.7}$$

Note that this is true also for the singular case, because we have restricted ourselves to stable systems and stationary solutions.

Of course the system (2.3), (2.4) is a state space system with x_t being a state. For results concerning state space systems and basic system theory needed for this paper see Kailath (1980, Chap. 6), and Hannan and Deistler (2012, Chap. 1 and 2). A state space system is called *controllable* if

$$\operatorname{rk}\left(\mathcal{B}, \mathcal{A}\mathcal{B}, \ldots, \mathcal{A}^{np-1}\mathcal{B}\right) = np,$$

see Hannan and Deistler (2012, Chap. 2), where this is called reachable. As can be easily seen from (2.7), using the Cayley–Hamilton theorem, the system (2.3), (2.4) is controllable if and only if Γ_p is nonsingular. A system is called *observable* if

$$\operatorname{rk}\begin{pmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{np-1} \end{pmatrix} = np.$$

Thus the system (2.3), (2.4) is observable if and only if A_p is nonsingular.

A system is both controllable and observable if and only if it is minimal, i.e., the state dimension np is minimal among all systems with given transfer function $k(z)$.

A subset of the parameter space Θ is called *generic* if it contains a subset that is open and dense in Θ. Clearly, the state space system (2.3), (2.4) corresponding to a regular AR system is always controllable. For the singular AR case, as is well known, see e.g., Lee and Markus (1967), the corresponding state space systems are generically controllable. It is immediate that systems (2.3), (2.4) are generically observable. Thus systems of the form (2.3), (2.4) are generically minimal.

Except for Section 5, we consider the case where there are stock variables only. Let

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

where the n_f-dimensional, say, fast (or to be more precise high frequency) component y_t^f is observed at the highest (sampling) frequency $t \in \mathbb{Z}$ and the n_s-dimensional slow (or to be more precise low frequency) component y_t^s is observed only for $t \in N\mathbb{Z}$ (N being an integer $N > 1$), i.e., for every N-th time point. Throughout we assume $n_f \geq 1$.

The matrices

$$A_i = \begin{pmatrix} a_{ff}(i) & a_{fs}(i) \\ a_{sf(i)} & a_{ss}(i) \end{pmatrix}, \; i = 1, \ldots, p, \; \Sigma_\nu = \begin{pmatrix} \sigma_{ff} & \sigma_{fs} \\ \sigma_{sf} & \sigma_{ss} \end{pmatrix}$$

are partitioned accordingly.

The central problem considered in this paper is *identifiability*, i.e., whether, for given Θ, the parameters A_i and Σ_ν of the high frequency system are uniquely determined by the population second moments of the observations. In the case of stock variables these population second moments are of the form

$$\begin{aligned}\gamma^{ff}(h) &= \mathbb{E}\left(y^f_{t+h}\left(y^f_t\right)^T\right),\ h \in \mathbb{Z},\\ \gamma^{sf}(h) &= \mathbb{E}\left(y^s_{t+h}\left(y^f_t\right)^T\right),\ h \in \mathbb{Z},\\ \gamma^{ss}(h) &= \mathbb{E}\left(y^s_{t+h}\left(y^s_t\right)^T\right),\ h \in N\mathbb{Z}. \end{aligned} \tag{2.8}$$

The observed outputs can be represented by the so-called blocked process

$$\tilde{y}_t = \begin{pmatrix} y_t \\ y^f_{t-1} \\ \vdots \\ y^f_{t-N+1} \end{pmatrix},\ t \in N\mathbb{Z},$$

whose covariances are exactly the second moments described above.

Note that if identifiability holds and if in addition there is an algorithm for obtaining the parameters of the high frequency system from the population second moments of the observations, we can reconstruct the missing moments $\gamma^{ss}(h) = \mathbb{E}\left(y^s_{t+h}(y^s_t)^T\right),\ h \in N\mathbb{Z} - j,\ j \in \{1, \ldots, N-1\}$ and thus all $\gamma(h) = \mathbb{E}\left(y_{t+h} y_t^T\right),\ h \in \mathbb{Z}$. Then linear least squares methods for forecasting, nowcasting, and interpolation of nonobserved output variables can be applied. If such an algorithm defines a continuous function and thus the problem is well-posed, the algorithm may be applied to sample second moments, in order to yield consistent estimators of the system and noise parameters of the underlying high frequency system and thus of missing second moments. In other words, identifiability and well-posedness are important in obtaining consistent estimators of the system and noise parameters, A_i and Σ_ν respectively.

Note however that there are problems which do not require identifiability of all high frequency parameters. As has been shown in Bai, Ghysels, and Wright (2013), this may be important for forecasting and as has been shown in Ghysels, Hill, and Motegi (2014), also analysis of Granger causality does not require identifiability of all high frequency parameters.

3. IDENTIFIABILITY FOR AR(1) SYSTEMS

In this section we consider the special case of AR(1) systems. In addition, for simplicity, we restrict ourselves to the case $N = 2$, $n_f = n_s = 1$. Furthermore, we assume throughout this section that the AR(1) system is regular. This is done for two reasons. First, it gives an example illustrating the problem. Second, as will be shown below, this analysis yields special results; in particular the subset, Θ_I say, of Θ, where identifiability is obtained (without imposing additional restrictions), can be described explicitly. We show that the complement of this set is a so-called semi-algebraic set, see Bochnak, Coste, and Roy (1998, p. 24, Def. 2.1.4), i.e., a set of (multivariate) polynomial zeros where in addition inequalities are imposed, and so we conclude that for generic parameter values identifiability is obtained. This case has been described in detail in Anderson et al. (2012).

We first consider the case where in addition Σ_ν is diagonal. Using a self-evident notation, we can write

$$\begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} = \underbrace{\begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}}_{\mathcal{A}} \begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix}. \tag{3.1}$$

Now, the one-step-ahead predictor for y_{t-1}^f, $t-1$ odd, based on observed outputs is obtained from the following equation:

$$y_{t-1}^f = a_{ff} y_{t-2}^f + a_{fs} y_{t-2}^s + \nu_{t-1}^f,$$

and the two-step-ahead predictor of y_t, t even, is obtained from

$$y_t = \mathcal{A}^2 y_{t-2} + \mathcal{A}\nu_{t-1} + \nu_t.$$

Combining both equations gives a three-dimensional system on $2\mathbb{Z}$:

$$\underbrace{\begin{pmatrix} y_t^f \\ y_t^s \\ y_{t-1}^f \end{pmatrix}}_{\tilde{y}_t} = \underbrace{\begin{pmatrix} \mathcal{A}^2 & 0 \\ a_{ff} \; a_{fs} & 0 \end{pmatrix}}_{\tilde{\mathcal{A}}} \begin{pmatrix} y_{t-2}^f \\ y_{t-2}^s \\ y_{t-3}^f \end{pmatrix} + \underbrace{\begin{pmatrix} \mathcal{A}\nu_{t-1} + \nu_t \\ \nu_{t-1}^f \end{pmatrix}}_{\tilde{\nu}_t}. \tag{3.2}$$

Note that (3.2) is an AR(1) system on $2\mathbb{Z}$ whose outputs $\tilde{y}_t$ are the observed variables and thus may serve as a model for the MF data.

As the components of $\tilde{y}_{t-2}$ are linearly independent by the regularity assumption, $\tilde{\mathcal{A}}$ and $\Sigma_{\tilde{\nu}} = \mathbb{E}\left(\tilde{\nu}_t \tilde{\nu}_t^T\right)$ are uniquely determined from the second moments of $(\tilde{y}_t \mid t \in 2\mathbb{Z})$. However, not all entries in $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$ are free, since

$$\tilde{\mathcal{A}} = \begin{pmatrix} a_{ff}^2 + a_{fs}a_{sf} & a_{ff}a_{fs} + a_{fs}a_{ss} & 0 \\ a_{sf}a_{ff} + a_{ss}a_{sf} & a_{sf}a_{fs} + a_{ss}^2 & 0 \\ a_{ff} & a_{fs} & 0 \end{pmatrix}, \tag{3.3}$$

$$\Sigma_{\tilde{\nu}} = \begin{pmatrix} \sigma_{ff} & 0 & 0 \\ 0 & \sigma_{ss} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{ff} & 0 \\ 0 & \sigma_{ss} \end{pmatrix} \begin{pmatrix} a_{ff} & a_{sf} & 1 \\ a_{fs} & a_{ss} & 0 \end{pmatrix} \tag{3.4}$$

must hold.

Here the high frequency system has 6 free parameters, whereas a general AR(1) system for $n = 3$ has 15 free parameters. In order to analyze identifiability we solve (3.3), (3.4) for given $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$ for the high frequency parameters $\mathcal{A}$ and Σ_ν. We see that if a_{fs} and a_{sf} are both zero, then only a_{ss}^2 is unique, otherwise $\mathcal{A}$ and Σ_ν are unique and thus we have nonidentifiability if and only if $a_{fs} = a_{sf} = 0$ and $a_{ss} \neq 0$ hold.

It is interesting to note that we have identifiability whenever the two component processes $\left(y_t^f | t \in \mathbb{Z}\right)$ and $\left(y_t^s | t \in \mathbb{Z}\right)$ are not orthogonal.

If we drop the assumption $\sigma_{sf} = 0$ we obtain:

THEOREM 1. *Assume that $p = 1$, $n_f = n_s = 1$, $\Sigma_\nu > 0$ and $N = 2$. The system and noise parameters $\begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}$, σ_{ff}, σ_{sf} and σ_{ss} are not identifiable if and only if they satisfy the equations*

$$\begin{aligned} a_{fs} &= 0, \\ a_{sf} + \frac{\sigma_{sf}}{\sigma_{ff}}(a_{ss} - a_{ff}) &= 0, \\ a_{ss} &\neq 0. \end{aligned} \tag{3.5}$$

The complement of the set of solutions of (3.5) contains an open and dense subset of Θ.

The proof is given in the appendix.

Remark 1. An interesting interpretation of Theorem 1 is the following. The parameters of the underlying high frequency model cannot be obtained if and only if there is a static linear transformation such that the transformed model has a diagonal innovation variance and the transformed system matrix is diagonal with nonzero $(2, 2)$ entry. Note that such a transformation must be of the form $T = \begin{pmatrix} 1 & 0 \\ -\sigma_{sf}\sigma_{ff}^{-1} & 1 \end{pmatrix}$ and that for given T the conditions of nonidentifiability arising are exactly the same as in (3.5). Thus identifiability for systems with nondiagonal innovation variance can be traced back to identifiability for systems with diagonal innovation variance.

Note that equations (3.3), (3.4) may also be used for identifiability analysis for $n = q > 2$ though dealing with the various special cases is more intricate since scalars are replaced by matrices. We repeat that the advantage of the analysis given above is that the subset of identifiable parameters is explicitly given, and that the genericity property stands out clearly.

On the other hand, the analysis above cannot be extended to the case $p > 1$ since the blocked process $(\tilde{y}_t)$ is no longer AR but ARMA in general. This can be seen from a two dimensional AR(2) example of the form

$$y_t = A_2 y_{t-2} + \nu_t \quad t \in \mathbb{Z},$$

where A_2 is nondiagonal. Note that the fact that AR systems are not closed under marginalization is well known, see e.g. Amemiya and Wu (1972) and Tiao (1972).

4. G-IDENTIFIABILITY RESULTS USING EXTENDED YULE WALKER EQUATIONS

As has been mentioned above, the analysis given in the previous section cannot be generalized to the case of order $p > 1$. The reason is that, in general, the blocked process $(\tilde{y}_t \mid t \in 2\mathbb{Z})$ will not be AR, or to put it in more general terms, AR processes are not closed under marginalization.

In the analysis given below we will not be able to give an explicit characterization (i.e., explicit restrictions on the system and noise parameters) of the subset Θ_I of all identifiable systems of Θ. As a "second best" result, we will show that Θ_I is generic which means that identifiability holds in typical cases. To repeat, a subset of the parameter space Θ is called *generic* if it contains a subset that is open and dense in Θ. We use the term g-identifiability if identifiability can be ensured on a generic subset of the parameter space.

In this paper we present two approaches to obtain a result on generic identifiability, which are both "constructive", i.e., based on algorithms for actually calculating the (unique) parameters $\theta \in \Theta$ from the second moments (2.8) of the observations. The first approach is based on extended Yule Walker equations (see Chen and Zadrozny, 1998) and is dealt with in this section. The alternative approach based on the blocking concept is presented in Section 7.

4.1. Derivation of the Extended Yule Walker Equations for Mixed Frequency Data

By postmultiplying equation (2.1) by $y_{t-j}^T,\ j > 0$ and forming expectations, we obtain Yule Walker equations. The problem with these equations is that matrices on both the left and right hand side contain unobserved second moments. In order to overcome this problem, we postmultiply equation (2.1) by $\left(y_{t-j}^f\right)^T,\ j > 0$ and form expectations. Thereby we obtain extended Yule Walker equations (XYW, see Chen and Zadrozny, 1998) as

$$\mathbb{E}\left(y_t\left(\left(y_{t-1}^f\right)^T,\left(y_{t-2}^f\right)^T,\dots\right)\right) = (A_1,\dots,A_p)\mathbb{E}\left(\begin{pmatrix} y_{t-1}\\ \vdots \\ y_{t-p}\end{pmatrix}\left(\left(y_{t-1}^f\right)^T,\left(y_{t-2}^f\right)^T,\dots\right)\right). \quad \textbf{(4.1)}$$

Let

$$K := \mathbb{E}\left(x_t\left(y_{t-1}^f\right)^T\right) = \mathbb{E}\left(\begin{pmatrix} y_{t-1}\\ \vdots \\ y_{t-p}\end{pmatrix}\left(y_{t-1}^f\right)^T\right)$$

$$= \begin{pmatrix} \gamma^{ff}(0) \\ \gamma^{sf}(0) \\ \vdots \\ \gamma^{ff}(-p+1) \\ \gamma^{sf}(-p+1) \end{pmatrix} = \Gamma_p \begin{pmatrix} I_{n_f} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad \textbf{(4.2)}$$

From equation (2.3), i.e., $x_{t+1} = \mathcal{A}x_t + \mathcal{B}\varepsilon_t$, we have that $x_t = \sum_{i=0}^{\infty} \mathcal{A}^i \mathcal{B}\varepsilon_{t-i-1}$ and $x_{t+s} = \mathcal{A}^s x_t + \sum_{i=0}^{s-1} \mathcal{A}^i \mathcal{B}\varepsilon_{t+s-i-1}$. The block columns of the second matrix on the right hand side of (4.1) are of the form $\mathbb{E}\Big(x_t\,(y_{t-j-1}^f)^T\Big) = \mathbb{E}\Big(x_{t+j}\,(y_{t-1}^f)^T\Big) = \mathbb{E}\Big((\mathcal{A}^j x_t + \sum_{i=0}^{j-1} \mathcal{A}^i \mathcal{B}\varepsilon_{t+j-i-1})(y_{t-1}^f)^T\Big) = \mathcal{A}^j \mathbb{E}\Big(x_t\,(y_{t-1}^f)^T\Big) = \mathcal{A}^j K,\ j \geq 0.$

Thus the rightmost matrix in the extended Yule Walker equations (4.1) can be written as $(K, \mathcal{A}K, \mathcal{A}^2K, \dots)$. From the Cayley-Hamilton theorem and since $\mathcal{A} \in \mathbb{R}^{np\times np}$, we see that the second matrix on the right hand side of (4.1) has full row rank if and only if the matrix consisting of the first np blocks has full row rank. In this way we have obtained our XYW which are of the form

$$\mathbb{E}\left[y_t\left(\left(y_{t-1}^f\right)^T, \dots, \left(y_{t-np}^f\right)^T\right)\right] = (A_1, \dots, A_p)\underbrace{\mathbb{E}\left[\begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p}\end{pmatrix}\left(\left(y_{t-1}^f\right)^T, \dots, \left(y_{t-np}^f\right)^T\right)\right]}_{=:Z}. \tag{4.3}$$

The crucial point is that the matrix Z can be written as

$$Z = \left(K, \mathcal{A}K, \mathcal{A}^2K, \dots, \mathcal{A}^{np-1}K\right), \tag{4.4}$$

and therefore has the structure of a controllability matrix.

Clearly, the system parameters $\left(A_1, \dots, A_p\right)$ of (2.1) are identifiable if Z has full row rank np, or equivalently, and in the language of linear system theory, the pair $(\mathcal{A}, K)$ is controllable. Note, however, that contrary to usual controllability matrices, here K depends on $\mathcal{A}$, which makes the task of verifying generic controllability more demanding.

Remark 2. Consider again the two-dimensional AR(1) process of Section 3. In this case the matrix Z in equation (4.3) is rank deficient for $a_{ss} = 0,\ a_{sf} = 0$ even if $a_{fs} \neq 0$ which shows that the condition $rk(Z) = np$ is not necessary for identifiability. For the example discussed in the previous section, where for the case $\sigma_{sf} = a_{fs} = a_{sf} = 0$ the classes of observationally equivalent parameters consist of two points (corresponding to the two choices of the square root of a_{ss}^2), the solution set of the XYW is a nontrivial affine subset. This shows that the XYW do not use the full information contained in the second moments of the observations.

Remark 3. The advantage of commencing from XYW is that they immediately give linear and consistent estimators.

4.2. g-Identifiability of System and Noise Parameters

The parameter space in this section is the set Θ of all $\left(\left(A_1, \dots, A_p\right), \Sigma_\nu\right)$ where $\left(A_1, \dots, A_p\right) \in \mathcal{S}$ and Σ_ν has rank q. In particular, we assume that there are no cross restrictions between system and noise parameters. We analyze identifiability of system parameters first.

The next theorem, which is a central result of this paper, shows that the matrix Z in equation (4.3) is generically of full row rank and thus we have generic identifiability for $\left(A_1,\ldots,A_p\right)$. Note that this holds both for regular and singular AR systems, for all sampling frequency ratios N, and all $n_f \geq 1$.

THEOREM 2. *The matrix Z in the extended Yule Walker equations (4.3) has full row rank np on a generic subset of the parameter space Θ, and thus the system parameters are g-identifiable.*

A proof of this theorem has been given in Anderson et al. (2012) and a more elegant proof is presented in the appendix.

Let vec denote columnwise vectorization and let us define $\mathcal{G} = (I_n, 0, \ldots, 0)$.

THEOREM 3. *The noise parameters Σ_ν are g-identifiable in Θ from*

$$\mathrm{vec}\,\Sigma_\nu = \left((\mathcal{G}\otimes\mathcal{G})\left(I_{(np)^2} - (\mathcal{A}\otimes\mathcal{A})\right)^{-1}\left(\mathcal{G}^T\otimes\mathcal{G}^T\right)\right)^{-1}\mathrm{vec}\gamma\,(0). \tag{4.5}$$

Again, the proof is given in the appendix.

Remark 4. From Theorems 2 and 3 we see that the system and noise parameters are g-identifiable, i.e., identifiable on the intersection of the set described in the proofs of Theorems 2 and 3. Note that the results shown in the proofs of the theorems are stronger than mere genericity results, because the set where Z has not full row rank np is a proper algebraic set (see Wonham, 1985, p. 28) and the same statement holds for the case of noise parameters.

Note that the property that Z has full row rank np depends on $\left(A_1,\ldots,A_p\right)$ as well as on Σ_ν whereas the uniqueness of Σ_ν obtained via (4.5) depends only on $\left(A_1,\ldots,A_p\right)$. The first assertion is easy to see by considering the AR(1) case of the previous section: Consider the special cases $a_{ff} \neq a_{ss}$, both nonzero, $a_{fs} = a_{sf} = 0$. If Σ_ν is diagonal Z has rank 1 otherwise Z has rank 2.

Remark 5. We have not been able to give an explicit description of those elements in Θ which are not identifiable or those parameters where Z is not of full row rank np.

Remark 6. Although we only have considered the case of two different sampling frequencies, an extension of the results given here to three or more sampling frequencies is straightforward.

If the system (2.3), (2.4) is not controllable, i.e., if Γ_p is singular, then clearly we have nonidentifiability even for high frequency data, as the Yule Walker equations then have no unique solution. Note however, that, as will be shown in Section 6, in such a situation identifiability might be obtained by suitably prescribing the column degrees in $a(z)$.

5. FLOW VARIABLES AND MORE GENERAL AGGREGATION SCHEMES

In the previous sections only stock variables have been considered. Here we deal with the case where the process $\left(y_t^s \mid t \in \mathbb{Z}\right)$ consists of flow variables or variables aggregated by more general schemes. For flow variables the aggregation to the corresponding observed process, $(w_t \mid t \in N\mathbb{Z})$ say, is of the form

$$w_t = y_t^s + y_{t-1}^s + \cdots + y_{t-N+1}^s = \left(1 + z + \cdots + z^{N-1}\right) y_t^s,\ t \in N\mathbb{Z}. \tag{5.1}$$

Remember that z denotes the backward shift on $\mathbb{Z}$.

Note that the second moments required in the extended Yule Walker equations are the autocovariances $\mathbb{E}\left(y_{t+h}^f \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$ and the cross covariances $\mathbb{E}\left(y_{t+h}^s \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$. We now show how these cross covariances can be retrieved from the cross covariances $\mathbb{E}\left(w_{t+h} \left(y_t^f\right)^T\right)$, $h \in \mathbb{Z}$ of the observations.

To show this, assume for the moment that w_t is available for all $t \in \mathbb{Z}$ and that the inverse of the linear transformation (5.1) exists for all $t \in \mathbb{Z}$, i.e.,

$$y_t^s = \underset{M \to \infty}{\text{l.i.m.}} \sum_{j=0}^{M} h_j^{(M)} w_{t-j}, \quad h_j^{(M)} \in \mathbb{R}^{n_s \times n_s},\ t \in \mathbb{Z}, \tag{5.2}$$

where l.i.m. denotes the limit in mean square. Then

$$\gamma^{sf}(h) = \mathbb{E}\left(y_{t+h}^s \left(y_t^f\right)^T\right) = \lim_{M \to \infty} \sum_{j=0}^{M} h_j^{(M)} \underbrace{\mathbb{E}\left(w_{t+h-j} \left(y_t^f\right)^T\right)}_{\gamma^{wy^f}(h-j)}. \tag{5.3}$$

Note that for our purposes the inverse of the linear transformation (5.2) only has to exist for the special input (w_t). In order to show the existence of the inverse transformation (5.2), it is more convenient to use the frequency domain rather than the time domain, see Rozanov (1967) and Hannan (1970). Let

$$f_{y^s y^s}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma^{ss}(h) e^{-i\lambda h}$$

and

$$f_{ww}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \mathbb{E}\left(w_{t+h} (w_t)^T\right) e^{-i\lambda h}$$

denote the spectral density of $(y_t^s \mid t \in \mathbb{Z})$ and $(w_t \mid t \in \mathbb{Z})$, respectively. As is well known, then the spectral density $f_{ww}(\lambda)$ of $(w_t \mid t \in \mathbb{Z})$ satisfies

$$f_{ww}(\lambda) = \left(1 + e^{-i\lambda} + \cdots + e^{-i(N-1)\lambda}\right) I_{n_s} f_{y^s y^s}(\lambda) I_{n_s} \left(1 + e^{i\lambda} + \cdots + e^{i(N-1)\lambda}\right)$$

and thus

$$\int \left(1+e^{-i\lambda}+\cdots+e^{-i(N-1)\lambda}\right)^{-1} I_{n_s} f_{ww}(\lambda) I_{n_s} \left(1+e^{i\lambda}+\cdots+e^{i(N-1)\lambda}\right)^{-1} d\lambda$$
$$= \int f_{y^s y^s}(\lambda) d\lambda < \infty.$$

Therefore each row of $\left(1+e^{-i\lambda}+\cdots+e^{-i(N-1)\lambda}\right)^{-1} I_{n_s}$ is an element of the frequency domain $\mathcal{L}_2(f_{ww} d\lambda)$ of f_{ww} and by the isomorphism between the frequency and the time domain the inverse transformation (5.2) is well defined. From (5.3) we then obtain

$$\begin{aligned} f_{y^s y^f}(\lambda) &= (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma^{sf}(h) e^{-i\lambda h} \\ &= \left(1+e^{-i\lambda}+\cdots+e^{-i(N-1)\lambda}\right)^{-1} I_{n_s} f_{wy^f}(\lambda) \end{aligned} \quad \textbf{(5.4)}$$

and thus $\gamma^{sf}(h)$, $h \in \mathbb{Z}$. In this way, we get all covariances in the extended Yule Walker equations. Note that whereas for the case of stock variables these covariances are the covariances of the observations, in the case considered here, they have to be reconstructed as described above.

A completely analogous derivation holds if we replace (5.1) by the more general aggregation scheme

$$w_t = k_0 y_t^s + k_1 y_{t-1}^s + \cdots + k_{N-1} y_{t-N+1}^s,\ k_0 \text{ nonsingular.} \quad \textbf{(5.5)}$$

Thus, taking into account that generically Z has row rank equal to np, we obtain the following result:

THEOREM 4. *Given the aggregation scheme (5.5) for the slow variables ($w_t \mid t \in N\mathbb{Z}$), the system and noise parameters of the high frequency system (2.1) are g-identifiable from $\gamma^{ff}(h)$ and $\gamma_{wy^f}(h)$, $h \in \mathbb{Z}$.*

Note that if we set $k_0 = I$ and $k_j = 0$, $j = 1, \ldots, N-1$, we have the case of stock variables. Thus Theorems 2 and 3 are special cases of Theorem 4. As is immediately seen, Theorem 4 also covers the case where the slow variables are formed by a mixture of stock and flow variables.

6. G-IDENTIFIABILITY FOR PRESCRIBED COLUMN DEGREES

In this section we are interested in identifiability of mixed frequency AR systems for the case that the column degrees of $a(z)$ rather than the degree of $a(z)$ are prescribed. Let $p_1, \ldots, p_n$ denote these prescribed column degrees. Let $\Theta_{(p_1,\ldots,p_n)}$ denote the subspace of Θ where the highest degree of the respective i-th column of $a(z)$ is bounded by p_i and let $(\bar{A}_1, \ldots, \bar{A}_p)$ denote all columns of $(A_1, \ldots, A_p)$ which are not prescribed to be zero.

Prescribing column degrees may be essential in obtaining identifiability for at least two reasons. First, consider the high frequency case only. Note that for the regular AR case, Γ_p is always nonsingular and thus no identifiability issues occur. In the case of singular AR systems, however, Γ_p might be singular and thus even from high frequency data we might not have identifiability. If this occurs, identifiability (from high frequency data) can be obtained by suitably prescribing column degrees, see Deistler, Filler, and Funovits (2011). Second, let us consider the mixed frequency case. For singular AR systems where in addition Γ_p is singular, of course, Z cannot be of full row rank. Appropriately prescribing column degrees leads to a modification of the matrix Z, as described below, which generically entails identifiability of the parameters of the high frequency system. As has been mentioned above, singular AR systems serve as models for latent variables or for static factors in dynamic factor models, see Deistler et al. (2010) and they occur in the case of definitional equations, see Phillips (1974).

In the high frequency case, if Γ_p is singular, identifiability can be obtained by selecting a basis for the row space of Γ_p consisting of the first basis rows. Let $S\Gamma_p$ denote the matrix formed by these basis rows where S is a $\sum_{i=1}^{n} p_i \times np$ selector matrix, for more details see Deistler et al. (2011). Thus, in the high frequency case, by prescribing appropriate column degrees we can always obtain identifiability. Clearly, for Γ_p singular, the system (2.3), (2.4) is not controllable and thus not minimal. A controllable system, however, is obtained as follows: We shall first consider the case $p_i > 0$, $i = 1, \ldots, n$.

Let us define

$$\bar{x}_{t+1} = \bar{\mathcal{A}}\bar{x}_t + \bar{\mathcal{B}}\varepsilon_t, \tag{6.1}$$

where $\bar{x}_t = Sx_t$, $\bar{\mathcal{A}} = S\mathcal{A}S^T$, $\bar{\mathcal{B}} = S\mathcal{B}$. Equation (6.1) has been called the quasi-companion form in Deistler et al. (2011).

It immediately follows that

$$y_t = (\bar{A}_1, \ldots, \bar{A}_p)\bar{x}_t + b\varepsilon_t. \tag{6.2}$$

Note that in the case of nonzero column degrees only "structural" zeros and ones, i.e., elements of $\mathcal{A}$ and $\mathcal{B}$ independent of the parameters $(A_1, \ldots, A_p)$ and Σ_ν, are deleted in $\mathcal{A}$ and $\mathcal{B}$. Thus, in this case of prescribed column degrees, identifiability of $(A_1, \ldots, A_p)$ and Σ_ν is equivalent to the identifiability of $(\bar{A}_1, \ldots, \bar{A}_p) = (A_1, \ldots, A_p)S^T$ and Σ_ν.

Completely analogously to Section 4.1 we can derive the modified extended Yule Walker equations for our reduced equation (6.2)

$$\begin{aligned}\mathbb{E}\left(y_t\left((y^f_{t-1})^T, \ldots, (y^f_{t-np+s})^T\right)\right) &= (\bar{A}_1, \ldots, \bar{A}_p)\mathbb{E}\left(\bar{x}_t\left((y^f_{t-1})^T, \ldots, (y^f_{t-np+s})^T\right)\right) \\ &= (\bar{A}_1, \ldots, \bar{A}_p)\underbrace{\left(\bar{K}, \bar{\mathcal{A}}\bar{K}, \bar{\mathcal{A}}^2\bar{K}, \ldots, \bar{\mathcal{A}}^{np-s-1}\bar{K}\right)}_{=:\bar{Z}},\end{aligned} \tag{6.3}$$

where s is the number of prescribed zero columns in $(A_1, \ldots, A_p)$ and $\bar{\mathcal{A}} \in \mathbb{R}^{(np-s)\times(np-s)}$ and where $\bar{\Gamma}_p = \mathbb{E}(\bar{x}_t \bar{x}_t^T)$ and $\bar{K} = \bar{\Gamma}_p \begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix}$. Now obviously the parameter matrices $(\bar{A}_1, \ldots, \bar{A}_p)$ are identifiable if the matrix $\bar{Z}$ has full row rank. In an analogous way as in the two preceding sections we obtain:

THEOREM 5. *For prescribed nonzero column degrees $p_1, \ldots, p_n$ the system parameters $(\bar{A}_1, \ldots, \bar{A}_p)$ and the noise parameters Σ_ν are g-identifiable with respect to the parameter space $\Theta_{(p_1 \ldots p_n)}$. Moreover this statement remains true for more general aggregation schemes (5.5).*

The proof of this theorem is in Appendix 8.

We now consider the case where there is at least one i such that $p_i = 0$. In this case we define two subprocesses of (y_t): Let $y_t^r = S_1 y_t$ contain all components $y_t^{(i)}$ of y_t with $p_i > 0$, and let $y_t^z = S_2 y_t$ contain all components $y_t^{(i)}$ with $p_i = 0$. It is easy to see that y_t^r is again an AR process. We obtain the following theorem:

THEOREM 6. *For prescribed column degrees $p_1, \ldots, p_n$ the system parameters $(\bar{A}_1, \ldots, \bar{A}_p)$ and the noise parameters Σ_ν are g-identifiable with respect to the parameter space $\Theta_{(p_1 \ldots p_n)}$ if $\left(y_t^r\right)$ contains at least one fast component. Moreover this statement remains true for more general aggregation schemes (5.5).*

The proof of this theorem can be found in Appendix 8.

7. AN ALTERNATIVE APPROACH: G-IDENTIFIABILITY RESULTS VIA BLOCKING

In this section we describe an alternative approach for obtaining g-identifiability results, namely blocking. We present this approach for three reasons: First, it provides additional insights into the structure of the problem; second, all second moments of the observations are used (note that in the XYW equations the available autocovariances of the slow process have not been used), and third, it leads to an alternative estimation procedure.

Blocking has been used in signal processing for a number of purposes, see Bittanti, Colaneri, and De Nicolao (1988). For blocking in case of mixed frequency data see Filler (2010), Ghysels (2012), Chen et al. (2012), and Zamani (2014). Here we draw to a great extent from fundamental results in system theory and again the reader not familiar with these results is referred to Kailath (1980) and Hannan and Deistler (2012).

In this section we restrict ourselves to the case that $\Gamma_p > 0$, A_p is nonsingular, and to the case of stock variables. The idea of blocking has been developed already in Section 3 for the AR(1) case. For convenience of the presentation, let us assume throughout in this section that $N = 2$ holds: The results presented here for $N = 2$ can also be obtained for general N where in Theorem 7 the additional assumption that $\mathcal{A}$ is diagonalizable has to be introduced. Note that the set of θ such that $\mathcal{A}$ is diagonalizable is generic, see e.g., Felsenstein (2014, Lemma 3.4.2).

We commence from the high frequency process $(y_t \mid t \in \mathbb{Z})$. In blocked form this process can be written as $(Y_t \mid t \in 2\mathbb{Z})$, where $Y_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$. Let $\mathcal{L}_2$ denote the Hilbert space of all scalar square integrable random variables over an underlying probability space $(\Omega, \mathcal{A}, \mathcal{P})$. By the results obtained by Akaike (1974) and Kalman (1965) (see also Hannan and Deistler, 2012, Chap. 2) we have the following state construction: We project all one-dimensional "future" random variables $y_{t+h}^{(i)}$, $h > 0$ on the Hilbert space spanned by all past variables $y_{t-j}^{(i)}$, $j \geq 0$ - here $y_t^{(i)}$ is the i-th component of y_t. The Hilbert space spanned by these projections, called the state space, is finite dimensional if and only if the spectral density of (y_t) is rational, and in this case every basis of the state space forms a minimal state of a stable and miniphase state space system with (y_t) as outputs and the innovations of (y_t) as inputs. The dimension of the state space is the so-called McMillan degree of the transfer function corresponding to such a stable and miniphase system. Note that miniphase means that there is a causal inverse to the system and thus the transfer function of a minimal, stable, and miniphase system has no zeros inside the unit circle. Such a transfer function corresponds to the Wold decomposition.

Note that the state spaces of $(y_t \mid t \in \mathbb{Z})$ for even t and $(Y_t \mid t \in 2\mathbb{Z})$ are the same and both processes are AR processes with corresponding minimal state dimension smaller than or equal to np and equal to np if and only if A_p is nonsingular and $\Gamma_p > 0$. In particular, we have from (2.3) the state space representation, for $p > 1$ (for $p = 1$ see Section 3),

$$x_{t+1} = \underbrace{\mathcal{A}^2}_{A_b} x_{t-1} + \underbrace{(\mathcal{B}, \mathcal{A}\mathcal{B})}_{B_b} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}, \tag{7.1}$$

$$Y_t = \begin{pmatrix} I_n & 0 & 0 & \cdots & 0 \\ 0 & I_n & 0 & \cdots & 0 \end{pmatrix} \mathcal{A}^2 x_{t-1} + \begin{pmatrix} b & A_1 b \\ 0 & b \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} \tag{7.2}$$

for $t \in 2\mathbb{Z}$, where, if for notational simplicity we consider the case of p even, the minimal state is given as

$$x_{t+1} = \begin{pmatrix} Y_t \\ \vdots \\ Y_{t-p+2} \end{pmatrix}.$$

Now, for the mixed frequency case, for stock variables and $N = 2$, we use the blocked observed process as $(\tilde{y}_t \mid t \in 2\mathbb{Z})$, $\tilde{y}_t = \begin{pmatrix} y_t \\ y_{t-1}^f \end{pmatrix}$ exactly as in Section 3. Also note that the second moments of $(\tilde{y}_t \mid t \in 2\mathbb{Z})$ are precisely the second moments (2.8) of the observations.

From (7.1), (7.2) we obtain the following state space representation for $(\tilde{y}_t \,|\, t \in 2\mathbb{Z})$

$$x_{t+1} = A_b x_{t-1} + B_b \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}, \tag{7.3}$$

$$\tilde{y}_t = \underbrace{\begin{pmatrix} I_n & 0 & 0 & \cdots & 0 \\ 0 & \left(I_{n_f}, 0\right) & 0 & \cdots & 0 \end{pmatrix} \mathcal{A}^2}_{C_b} x_{t-1} + \underbrace{\begin{pmatrix} b & A_1 b \\ 0 & \left(I_{n_f}, 0\right) b \end{pmatrix}}_{D_b} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}. \tag{7.4}$$

Whereas in (7.1), (7.2), as easily can be seen, the driving noise $\left(\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} \,|\, t \in 2\mathbb{Z}\right)$ are innovations for $(Y_t \,|\, t \in 2\mathbb{Z})$, the $\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}$ are not innovations for $\tilde{y}_t$, at least in the regular case, as is already clear from considering the dimensions of the respective vectors.

The spectral density of $(\tilde{y}_t \mid t \in 2\mathbb{Z})$, using (7.3), (7.4), can be represented as

$$f_{\tilde{y}}\left(z^2\right) = \left(C_b \left(I_{np}\left(z^2\right)^{-1} - A_b\right)^{-1} B_b + D_b\right) \times \left(B_b^T \left(I_{np} z^2 - A_b^T\right)^{-1} C_b^T + D_b^T\right), \tag{7.5}$$

where the spectral factor on the right side of (7.5) is “fat” and not miniphase. On the other hand, the spectral density of $(\tilde{y}_t \mid t \in 2\mathbb{Z})$ can be written as

$$f_{\tilde{y}}\left(z^2\right) = k\left(z^2\right) k^T\left(z^{-2}\right), \tag{7.6}$$

where $k\left(z^2\right)$ is a stable and miniphase spectral factor (not to be confused with $k(z)$ below (2.5)). Thus for a suitable quadruple $\left(\bar{A}_b, \bar{B}_b, \bar{C}_b, \bar{D}_b\right)$ denoting a minimal, stable and miniphase state space system, we have

$$k\left(z^2\right) = \left(\bar{C}_b \left(I_{np}\left(z^2\right)^{-1} - \bar{A}_b\right)^{-1} \bar{B}_b + \bar{D}_b\right). \tag{7.7}$$

The question of the relation of the state dimensions of minimal, stable and miniphase state space systems for $(\tilde{y}_t \,|\, t \in 2\mathbb{Z})$ and $(Y_t \,|\, t \in 2\mathbb{Z})$ arises. The next theorem states that despite the fact that the unobserved outputs have been omitted in $\tilde{y}_t$, generically the McMillan degrees of the transfer function corresponding to (7.1), (7.2) and in $k\left(z^2\right)$ in (7.7) are the same.

THEOREM 7. *For* $((A_1, \ldots, A_p), \Sigma_\nu) \in \Theta$, *if* A_p *is nonsingular,* $\Gamma_p > 0$, *and if for eigenvalues of* $\mathcal{A}$ *such that* $\lambda_i \neq \lambda_j$ *it follows that* $\lambda_i^2 \neq \lambda_j^2$ *holds, the McMillan degree of a causal and miniphase spectral factor* $k(z^2)$ *of* $f_{\tilde{y}}(z^2)$ *is equal to* np.

The proof of this theorem is given in the appendix. This proof is based on the fact that for a process the rank of the Hankel matrix of the covariances $\mathcal{H}^\gamma$ is the same as the rank of the Hankel matrix of a miniphase transfer function $\mathcal{H}^k$, which is the McMillan degree.

Remark 7. It is easy to show that the assumptions of Theorem 7 define a generic subset of the parameter space Θ.

The next theorem shows the relation between $(\bar{A}_b, \bar{C}_b)$ and (A_b, C_b):

THEOREM 8. *Under the assumptions of Theorem 7 the matrices* (A_b, C_b) *and* $(\bar{A}_b, \bar{C}_b)$ *in (7.5) and (7.7) respectively are the same up to basis change, i.e.,*

$$\bar{A}_b = T^{-1} A_b T, \tag{7.8}$$

$$\bar{C}_b = C_b T \tag{7.9}$$

for a suitably chosen nonsingular $np \times np$ *matrix* T.

Again, the proof of this theorem is given in the appendix.

Remark 8. Note that Theorem 7 is essential for (7.8), (7.9) because it ensures that $\bar{A}_b$ and A_b are of the same dimension. Also note that the result of Theorem 8 holds despite the fact that the states in (7.3), (7.4) and in the minimal, stable and miniphase system corresponding to (7.7) are not the same, even up to basis change.

These considerations lead us to the following procedure set out below in the next paragraphs, subject to the following assumptions. In addition to the assumptions of Theorem 7 we assume (a) that the pair $\left(\begin{pmatrix} I_{n_f} & 0 & \cdots & 0 \end{pmatrix}, \mathcal{A}\right)$ is observable, which is generic, see Anderson et al. (2012), and (b) that the eigenvalues of $\mathcal{A}$ are distinct, which also is generic, as is easy to see, since the eigenvalues are the inverse of the zeros of $\det a(z)$.

The matrices $(\bar{A}_b, \bar{C}_b)$ can be obtained from the Hankel matrix of covariances $\mathcal{H}^\gamma$ as shown in the proof of Theorem 8 in the Appendix, see equations (G.1) and (G.2). Thus for a given S the matrices $(\bar{A}_b, \bar{C}_b)$ depend continuously on the covariances of the observations.

We are left with the task to find the root of the matrix $\bar{A}_b$, say $\bar{\mathcal{A}}$, and to find the transformation T corresponding to the basis change to yield $\mathcal{A} = T\bar{\mathcal{A}}T^{-1}$ where $\mathcal{A}$ has the companion structure (2.3). By our assumptions, the eigendecomposition of $\mathcal{A}$ is $\mathcal{A} = Q\Lambda Q^{-1}$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{np})$ contains the eigenvalues, and $Q = (q_1, .., q_{np})$ where q_i are the eigenvectors. Now it immediately follows that $\bar{\mathcal{A}} = T^{-1}Q\Lambda Q^{-1}T$ and $\bar{A}_b = T^{-1}Q\Lambda^2 Q^{-1}T$, which is an eigendecomposition of $\bar{A}_b$. Note that under our assumptions $T^{-1}Q$ and Λ^2 can be determined from $\bar{A}_b$. Now we have

$$\bar{C}_b T^{-1} Q = \begin{pmatrix} \begin{pmatrix} I_n & 0 & \cdots & 0 \end{pmatrix} \mathcal{A}^2 Q \\ \begin{pmatrix} (I_{n_f}, 0) & 0 & \cdots & 0 \end{pmatrix} \mathcal{A} Q \end{pmatrix}, \tag{7.10}$$

where we used $(0\ I_n \cdots 0)\mathcal{A}^2 = (I_n\ 0 \cdots 0)\mathcal{A}$. Since Q is a matrix of eigenvectors for both $\mathcal{A}$ and $\mathcal{A}^2$, we can determine the eigenvalues λ_i of $\mathcal{A}$ by considering a submatrix of (7.10):

$$\begin{pmatrix} I_{n_f}\ 0 \cdots 0 \end{pmatrix} \mathcal{A}^2 q_i = \begin{pmatrix} I_{n_f}\ 0 \cdots 0 \end{pmatrix} \lambda_i^2 q_i, \tag{7.11}$$

$$\begin{pmatrix} I_{n_f}\ 0 \cdots 0 \end{pmatrix} \mathcal{A} q_i = \begin{pmatrix} I_{n_f}\ 0 \cdots 0 \end{pmatrix} \lambda_i q_i. \tag{7.12}$$

Generic observability of $\left(\left(I_{n_f}\ 0 \cdots 0 \right), \mathcal{A} \right)$ guarantees that the first n_f components of q_i are not all zero (by the so-called PBH test, see Kailath, 1980, p. 135) and thus we obtain λ_i. $\bar{\mathcal{A}}$ can be determined as follows:

$$\bar{\mathcal{A}} = T^{-1} Q \Lambda Q^{-1} T = T^{-1} \mathcal{A} T. \tag{7.13}$$

Now we partition T in p block-rows T_i with dimension $n \times np$. From (7.13) we obtain

$$A_1 T_1 + A_2 T_2 + ... + A_p T_p = T_1 \bar{\mathcal{A}} \qquad T_i = T_{i+1} \bar{\mathcal{A}}, \quad i = 1, ..., p-1. \tag{7.14}$$

Using the fact that we know T_1 from

$$\bar{C}_b \bar{A}_b^{-1} = \begin{pmatrix} T_1 \\ (I_{n_f}, 0)\, T_2 \end{pmatrix} \tag{7.15}$$

we can calculate the remaining T_i from $T_i = T_1 \bar{\mathcal{A}}^{-i+1}$, $i = 2, .., p$. Finally, we obtain the desired companion form $\mathcal{A} = T \bar{\mathcal{A}} T^{-1}$, where the free system parameters are in the first n rows and are uniquely determined. Thus we have shown part of the following theorem:

THEOREM 9. *Under the assumptions of Theorem 7 and the additional assumptions that the pair $\left(\left(I_{n_f}\ 0 \cdots 0 \right), \mathcal{A} \right)$ is observable and that the eigenvalues λ_i of $\mathcal{A}$ are distinct, the system parameters $\left(A_1, \ldots, A_p \right)$ as well as the noise parameters are uniquely determined from the population second moments of the observations.*

It remains to show uniqueness of the noise parameters Σ_ν. Note that given the system parameters $\left(A_1, ..., A_p \right)$ the noise parameters Σ_ν can be generically determined as in Section 4.2. However, under our assumptions, we can show that

Σ_ν is uniquely determined as follows: As easily can be shown that (compare Section 4.1)

$$\gamma(h) = \mathcal{G}\mathcal{A}^h \Gamma_p \mathcal{G}^T \tag{7.16}$$

holds where again $\mathcal{G} = (I_n, 0, \ldots, 0)$. Thus

$$\begin{pmatrix} \gamma(0) \\ \gamma(2) \\ \gamma(4) \\ \vdots \\ \gamma(2(np-1)) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathcal{G} \\ \mathcal{G}\mathcal{A}^2 \\ \mathcal{G}\mathcal{A}^4 \\ \vdots \\ \mathcal{G}\mathcal{A}^{2(np-1)} \end{pmatrix}}_{\mathcal{O}_2} \Gamma_p \mathcal{G}^T \tag{7.17}$$

holds and under our assumptions (compare (F.1) in appendix 8, note that $\mathcal{O}_2\mathcal{A}^2 = \mathcal{O}$) the matrix $\mathcal{O}_2$ has full column rank and therefore $\Gamma_p \mathcal{G}^T$ is unique. Hence using (7.16) all missing second moments can be reconstructed. Now Σ_ν can be obtained via the "high frequency Yule Walker equation":

$$\Sigma_\nu = \gamma(0) - (A_1, \ldots, A_p) \begin{pmatrix} \gamma(-1) \\ \vdots \\ \gamma(-p+1) \end{pmatrix}. \tag{7.18}$$

Note that the assumptions of Theorem 9 do not determine a largest set where identifiability holds: Consider again the set described in Theorem 1 for the two dimensional AR(1) case. For $a_{ss} = 0$, $a_{sf} = 0$ the assumption that A_p is nonsingular is violated.

The approach via blocking is again constructive as is the approach via XYW. Note that we have implicitly proposed an algorithm for obtaining a unique parameter which is identifiable in Θ. It is straightforward to show that θ depends on the second moments of $(\tilde{y}_t)$ in a continuous way: First note that by equations (G.1) and (G.2) $(\bar{A}_b, \bar{C}_b)$ continuously depends on the second moments of the process $(\tilde{y}_t)$. Next we show that $(A_1, \ldots, A_p)$ continuously depends on $(\bar{A}_b, \bar{C}_b)$. In order to see this, note that the eigenvalues λ_i^2 are continuous functions of the matrix entries and under our assumptions the eigenvectors have the same property, when the eigenvectors are suitably normalized. By equations (7.11), (7.12) also the λ_i are continuous functions of $(\bar{A}_b, \bar{C}_b)$ and by (7.13) the same holds for $\bar{\mathcal{A}}$. The equation (7.15) and the argument below show that it is also true for T and thus for $\mathcal{A}$. The continuous dependence of Σ_ν on the second moments of the observations is easily seen from (7.16), (7.17), and (7.18).

Note that $(\bar{A}_b, \bar{C}_b)$ is contained in an Euclidean space of dimension $np(np + n + n_f)$ whereas $(A_1, \ldots, A_p)$ has $n^2 p$ free parameters, see Section 3. The set of all stable and observable $(\bar{A}_b, \bar{C}_b)$ corresponding to our parameter space is a set of the form $\bar{A}_b = T^{-1}\mathcal{A}^2 T$, $\bar{C}_b = \begin{pmatrix} I_n & 0 & 0\cdots 0 \\ 0 & (I_{n_f},0) & 0\cdots 0 \end{pmatrix} \mathcal{A}^2 T$, where $\mathcal{A}$ corresponds to $\theta \in \Theta$ and $\det(T) \neq 0$. Note that an estimate of $(\bar{A}_b, \bar{C}_b)$ is not necessarily contained in this set and thus has to be projected.

We have not been able to describe a relation between the generic set corresponding to XYW and the generic set corresponding to blocking.

One advantage for blocking, as has been already said, seems to be that all second moments of the observations are used in contrast to the case of XYW. Accordingly, one could hope for better results when working with real data.

8. OUTLOOK AND CONCLUSIONS

This paper deals with identifiability of system and noise parameters of a multivariate (high frequency) AR system when only mixed frequency observations are available. No a priori assumptions are imposed on the high frequency system except for its degree, stability and the rank of the innovation variance. The main result is that identifiability holds, (possibly) except for the parameters in a "thin" set, which is the union of two proper algebraic sets (which are lower dimensional). Thus "typically" we have identifiability and, in particular, we have identifiability on a set containing an open and dense subset of the parameter space. This holds for all "aggregation intervals" N for the slow component of the observations and whenever there is at least one fast component in the observations. Our approach is different from the well known MIDAS approach (see Ghysels, 2012) and we also do not use a relation to continuous-time systems and more generally, let us repeat, we do not impose a priori restrictions in addition to the general ones mentioned above. Despite the fact that for most parts of the paper we discuss the case of stock variables, we show that the central results hold for very general (linear) aggregation schemes for the slow variables, including stock and flow variables as special cases.

The final aim of identifiability analysis is to obtain estimators. Note that, as has been pointed out above, the identifiability analysis given in the previous sections is "constructive", i.e., it not only ensures generic uniqueness of the AR-parameters for given population second moments (2.8), but also provides an algorithm for actually calculating these parameters from the population second moments. In system theory, following terminology introduced by Kalman (see Ho and Kalman, 1966), such an algorithm is called a realization algorithm. Continuity of these algorithms leads to consistent estimators. This suggests two estimation algorithms: one based on the extended Yule Walker equations as suggested by Chen and Zadrozny (1998), and the second using a subspace algorithm for the blocked observations, see Deistler, Peternell, and Scherrer (1995). Both algorithms may also be used for obtaining initial estimators for numerically maximizing the Gaussian likelihood.

To summarize, we consider the identifiability analysis given here to be important for the overall problem of parameter estimation for the mixed frequency case.

Despite the fact that we have not been able to explicitly describe the set of nonidentifiable systems (except for the AR(1) case), it has been shown that we have identifiability for "almost all" parameters. In addition, by its constructive nature, our realization algorithms lead to consistent estimation procedures. We admit - and this is ongoing work - that there are a number of important open questions concerning estimation. This includes a comparison of the estimators discussed above by Monte Carlo simulation and by their asymptotic variances. Of special interest is also, on the one hand, an analysis of the information loss due to mixed frequency data relative to the case of high frequency data and, on the other hand, of the information gain when using mixed frequency data compared to the use of data at the slow rate only.

REFERENCES

Akaike, H. (1974) Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control* 19(6), 667–674.

Amemiya, T. & R.Y. Wu (1972) The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* 67(339), 628–632.

Anderson, B.D.O., M. Deistler, E. Felsenstein, B. Funovits, P.A. Zadrozny, M. Eichler, W. Chen, & M. Zamani (2012) Identifiability of regular and singular multivariate autoregressive models. In Proceedings of the 51th IEEE Conference on Decision and Control (CDC), pp. 184–189. *Query 2*

Aruoba, S.B., F.X. Diebold, & C. Scotti (2007) Real-time measurement of business conditions. *Journal of Business and Economic Statistics* 27(4), 417–427.

Bai, J., E. Ghysels, & J.B. Wright (2013) State space models and MIDAS regressions. *Econometric Reviews* 32(7), 779–813.

Bergstrom, A. (1988) The history of continuous-time econometric models. *Econometric Theory* 4, 365–383.

Bernanke, B.S., M. Gertler, M.W. Watson, C.A. Sims, & B.M. Friedman (1997) Systematic monetary policy and the effects of oil price shocks. In *Brookings Papers on Economic Activity*, pp. 91–157.

Bittanti, S., P. Colaneri, & G. De Nicolao (1988) The difference periodic Riccati equation for the periodic prediction problem. *IEEE Transactions on Automatic Control* 33(8), 706–712.

Bochnak, J., M. Coste, & M.-F. Roy (1998) *Real Algebraic Geometry*. Springer Verlag.

Brockwell, P.J. (1995) A note on the embedding of discrete time ARMA processes. *Journal of Time Series Analysis* 16(5), 451–460.

Chambers, M.J. & M.A. Thornton (2012) Discrete time representation of continuous time ARMA processes. *Econometric Theory* 28, 219–238.

Chen, W., B.D.C. Anderson, M. Deistler, & A. Filler (2012) Properties of blocked linear systems. *Automatica* 48, 2520–2525.

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

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

Deistler, M., A. Filler, & B. Funovits (2011) AR systems and AR processes: The singular case. *Communications in Information and Systems* 11(3), 225–236.

Deistler, M., K. Peternell, & W. Scherrer (1995) Consistency and relative efficiency of subspace methods. *Automatica* 31(12), 1865–1875.

Deistler, M. & H.-G. Seifert (1978) Identifiability and consistent estimability in econometric models. *Econometrica* 46(6), 969–980.

Doz, C., D. Giannone, & L. Reichlin (2011) A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164(1), 188–205.
Felsenstein, E. (2014) Regular and singular AR and ARMA models: The single and the mixed frequency case. PhD thesis, Vienna University of Technology.
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., M. Hallin, M. Lippi, & L. Reichlin (2000) The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics* 82(4), 540–554.
Ghysels, E. (2012) Macroeconomics and the Reality of Mixed Frequency Data. Working paper, Department of Economics, University of North Carolina.
Ghysels, E., J.B. Hill, & K. Motegi (2014) Testing for Granger Causality in Mixed Frequency. Working paper.
Ghysels, E., P. Santa-Clara, & R. Valkanov (2006) Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131(1), 59–95.
Ghysels, E., A. Sinko, & R. Valkanov (2007) MIDAS regressions: Further results and new directions. *Econometric Reviews* 26(1), 53–90.
Ghysels, E. & J.H. Wright (2009) Forecasting professional forecasters. *Journal of Business and Economic Statistics* 27(4), 504–516.
Hannan, E.J. (1970) *Multiple Time Series*. Wiley.
Hannan, E.J. & M. Deistler (2012) *The Statistical Theory of Linear Systems*. SIAM Classics in Applied Mathematics.
Query 3
Hansen, L.P. & T.J. Sargent (1983) The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51(2), 377–387.
Harvey, A.C. & R.G. Pierse (1984) Estimating missing observations in economic time series. *Journal of the American Statistical Association* 79(385), 125–131.
Ho, B.L. & R.E. Kalman (1966) Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* 14, 545–548.
Kailath, T. (1980) *Linear Systems*. Prentice Hall.
Kalman, R.E. (1965) Irreducible realizations and the degree of a rational matrix. *Journal of the Society for Industrial & Applied Mathematics* 13(2), 520–544.
Kohn, R. & C.F. Ansley (1986) Estimation, prediction and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association* 81(395), 751–761.
Lee, E.B. & L. Markus (1967) *Foundations of Optimal Control Theory*. Wiley.
Marcellino, M. (1998) Temporal disaggregation, missing observations, outliers, and forecasting: A unifying non-model based procedure. *Advances in Econometrics* 13, 181–202.
Marcellino, M. & C. Schumacher (2010) Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statictics* 72(4), 518–550.
Mariano, R.S. & Y. Murasawa (2003) A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18(4), 427–443.
Nijman, T.E. (1985) Missing observations in dynamic macroeconomic modeling. PhD thesis, VU University Amsterdam.
Phillips, P.C.B. (1973) The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 1(4), 351–362.
Phillips, P.C.B. (1974) The estimation of some continuous time models. *Econometrica* 42(5), 803–823.
Phillips, P.C.B. (1976) Some computations based on observed data series of the exogenous variable component in continuous systems. In *Statistical Inference in Continuous Time Economic Models*, pp. 175–214.
Rozanov, Y.A. (1967) *Stationary Random Processes*. Holden-Day.
Stock, J.H. & M.W. Watson (2002) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.

Tiao, G.C. (1972) Asymptotic behaviour of temporal aggregates of time series. *Biometrika* 59(3), 525–531.

Wohlrabe, K. (2008) Forecasting with mixed-frequency time series models. PhD thesis, Ludwig-Maximilians-Universitt Mnchen.

Wonham, W.M. (1985) *Linear Multivariable Control - A Geometric Approach*, 3rd ed. Springer Verlag.

Zadrozny, P.A. (1988) Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4(1), 108–124.

Zadrozny, P.A. (1990a) Estimating a multivariate ARMA model with mixed frequency data: An application to forecasting U.S. GNP at monthly intervals. *Federal Reserve Bank of Atlanta*, 90(6). Query 4

Zadrozny, P.A. (1990b) Forecasting US GNP at monthly intervals with an estimated bivariate time series model. *Economic Review*, 2–15. Query 5

Zamani, M. (2014) *Modeling multivariable time series using regular and singular autoregressions*. PhD thesis, The Australian National University.

APPENDIX A: Proof of Theorem 1

Note that $\tilde{\mathcal{A}}$ (3.3) does not depend on σ_{sf} and thus is unaffected whether σ_{sf} is set equal to zero or not. Only (3.4) is changed to

$$\Sigma_{\tilde{\nu}} = \begin{pmatrix} \sigma_{ff} & \sigma_{sf} & 0 \\ \sigma_{sf} & \sigma_{ss} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{ff} & \sigma_{sf} \\ \sigma_{sf} & \sigma_{ss} \end{pmatrix} \begin{pmatrix} a_{ff} & a_{sf} & 1 \\ a_{fs} & a_{ss} & 0 \end{pmatrix}. \tag{A.1}$$

Thus a_{ff}, a_{fs}, and σ_{ff} are unique for given $\tilde{\mathcal{A}}$, $\Sigma_{\tilde{\nu}}$.

We are left with the problem to uniquely solve equation systems (3.3) and (A.1) in the variables a_{sf}, a_{ss}, σ_{sf}, and σ_{ss}. For this purpose we distinguish two cases, namely the case $a_{fs} = 0$ and the case $a_{fs} \neq 0$, considering that we already know a_{fs}.

We start with the case $a_{fs} \neq 0$. It is easy to see that the missing parameters a_{sf}, a_{ss}, σ_{sf} can be recovered using (3.3) and (A.1). Subsequently σ_{ss} can be recovered using equation (A.1).

In the event that $a_{fs} = 0$, then $\mathcal{A}^2$ is lower triangular, with $(2,2)$ entry a_{ss}^2. First, observe that the $(2,3)$ and $(2,1)$ entries of $\Sigma_{\tilde{\nu}}$ are respectively $\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}$ and $a_{ff}\left(\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}\right) + \sigma_{sf}$. It is immediate that σ_{sf} is available.

Next, if $a_{ss} = 0$, a fact which is immediately known from the $(2,2)$ entry of $\mathcal{A}^2$, then the $(2,3)$ entry of $\Sigma_{\tilde{\nu}}$ is simply $\sigma_{ff}a_{sf}$ and since σ_{ff} is obviously nonzero, the value of a_{sf} can be obtained. Also, the $(2,2)$ entry of $\Sigma_{\tilde{\nu}}$ is $a_{sf}\left(\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss}\right) + \sigma_{ss}$ and one immediately has σ_{ss}.

It therefore remains to consider the situation where $a_{ss} \neq 0$. The following quantities α and β, corresponding to the $(2,1)$ entry of $\mathcal{A}^2$ and the $(2,3)$ entry of $\Sigma_{\tilde{\nu}}$, are known:

$$a_{sf}a_{ff} + a_{ss}a_{sf} = \alpha,$$
$$\sigma_{ff}a_{sf} + \sigma_{sf}a_{ss} = \beta.$$

By eliminating a_{sf}, we obtain

$$-a_{ss}^2\sigma_{sf} + \left(\beta - a_{ff}\sigma_{sf}\right)a_{ss} + a_{ff}\beta = \alpha\sigma_{ff}$$

Using this equation and the value for a_{ss}^2 available from $\mathcal{A}^2$, it follows that a_{ss} is uniquely determined if and only if

$$\beta - a_{ff}\sigma_{sf} \neq 0.$$

Introducing the expression above for β, this yields:

$$a_{sf}\sigma_{ff} + a_{ss}\sigma_{sf} - a_{ff}\sigma_{sf} \neq 0.$$

To sum up, identification is possible except for parameters satisfying

$$\begin{aligned} a_{fs} &= 0, \\ a_{ss} &\neq 0, \\ a_{sf}\sigma_{ff} + a_{ss}\sigma_{sf} - a_{ff}\sigma_{sf} &= 0. \end{aligned}$$

The set of nonidentifiable points as described by equations (3.5) is a so-called semi-algebraic set, see Bochnak et al. (1998, p. 24, Def. 2.1.4), i.e., a set of (multivariate) polynomial zeros where in addition inequalities may be imposed. Here, in particular, the set of all identifiable parameters, which is a complement of the semi-algebraic set above, contains a generic subset of the parameter space, viz. the complement of the set defined by the zeros of the polynomial equalities alone.

APPENDIX B: Proof of Theorem 2

The proof uses the following well known result, see e.g. Lee and Markus (1967), Wonham (1985), and Bochnak et al. (1998):

Let $f : \Theta \to \mathbb{R}$ be a polynomial function. If there exists a $\theta^* \in \Theta$ such that $f(\theta^*) \neq 0$, then the set of zeros of f is a proper algebraic set and in particular its complement in Θ is generic.

In a first step, we have to show that Z is a rational function of $\theta \in \Theta$. It follows immediately that Z is rational if we can show that K is a rational function of $\theta \in \Theta$. Vectorizing the Lyapunov equation (2.6) we obtain

$$\text{vec}\Gamma_p = (\mathcal{A} \otimes \mathcal{A})\,\text{vec}\Gamma_p + \text{vec}\mathcal{B}\mathcal{B}^T$$

and thus

$$\text{vec}\Gamma_p = \left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1} \text{vec}\mathcal{B}\mathcal{B}^T. \tag{B.1}$$

Note that the absolute value of all eigenvalues λ_j of $\mathcal{A}$ is smaller than one by the stability assumption (2.2). Therefore the same holds for the eigenvalues of $(\mathcal{A} \otimes \mathcal{A})$ since the eigenvalues of $(\mathcal{A} \otimes \mathcal{A})$ are $\lambda_i \lambda_j$ $i, j = 1, \ldots, np$ and thus $\left(I_{(np)^2} - \left(\mathcal{A} \otimes \mathcal{A}\right)\right)$ is nonsingular. This implies that $\text{vec}\Gamma_p$ is a rational function in $\left(\left(A_1, \ldots, A_p\right), \Sigma_\nu\right)$ having no poles in Θ. Thus K and $\mathcal{A}^j K$ and subsequently Z are rational in $\left(\left(A_1, \ldots, A_p\right), \Sigma_\nu\right)$ on Θ. Without loss of generality we may restrict ourselves to the case where K is a vector and thus Z is square. Multiplying Z by $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ we obtain a polynomial in the entries of $\left(\left(A_1, \ldots, A_p\right), \Sigma_\nu\right)$ since $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ has no zeros. Thus the set of zeros of the determinant of the polynomial matrix $\det\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)Z$ is the same as the set of zeros of the determinant of Z and thus is an algebraic set in Θ, compare Bochnak et al. (1998, p. 23).

Now consider a point θ^* in Θ given by

$$\mathcal{A} = \begin{pmatrix} 0 & \cdots & 0 & \rho C \\ I_n & & & \\ & \ddots & & \\ & & I_n & 0 \end{pmatrix}, \ \mathcal{B} = \mathcal{E}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{B.2}$$

where $\rho \in (0, 1)$ and

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

is a so-called circulant matrix and define $e_1 \in \mathbb{R}^n$, where the first component is one and all others are zero. We will show that for this point in the parameter space, $\det(Z) \neq 0$ holds.

Note that for an AR(p) process w_t of a system with parameters θ^* the covariance $\gamma(0)$ is diagonal and the covariances $\gamma(j)$, $j = 1, .., p-1$ are zero, which is easily seen by looking at the Wold decomposition $w_t = \sum_{j=0}^{\infty} \rho^j C^j e_1 \varepsilon_{t-jp}$. Obviously, $\gamma(0) = \sum_{j=0}^{\infty} \rho^{2j} C^j e_1 e_1^T \left(C^j\right)^T$ is nonsingular. Thus $\Gamma_p > 0$ holds and this implies that $\left(\mathcal{B}, \mathcal{AB}, \ldots, \mathcal{A}^{np-1}\mathcal{B}\right)$ is of full row rank, see (2.7). Now it is immediate that Z is of full row rank since, as Γ_p is diagonal, $Z = \left(\Gamma_p \mathcal{E}_1, \mathcal{A}\Gamma_p \mathcal{E}_1, \ldots, \mathcal{A}^{np-1}\Gamma_p \mathcal{E}_1\right)$ is a multiple of $\left(\mathcal{B}, \mathcal{AB}, \ldots, \mathcal{A}^{np-1}\mathcal{B}\right)$. Thus $\det(Z) \neq 0$ holds.

Thus the set of zeros of $\det(Z)$ is a proper algebraic set, i.e.,, an algebraic set of dimension smaller than the dimension of Θ. Therefore its complement in the parameter space, which corresponds to all controllable pairs, is the complement of a proper algebraic set and hence is open and dense in the parameter space.

APPENDIX C: Proof of Theorem 3

We commence from identifiable system parameters $\left(A_1, \ldots, A_p\right)$. Through columnwise vectorization of

$$\gamma(0) = \mathbb{E}\left(y_t y_t^T\right) = \mathcal{G}\Gamma_p \mathcal{G}^T$$

we obtain

$$\text{vec}\gamma(0) = (\mathcal{G} \otimes \mathcal{G})\text{vec}\Gamma_p.$$

This together with (B.1) gives

$$\text{vec}\gamma(0) = (\mathcal{G} \otimes \mathcal{G})\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}(\mathcal{G}^T \otimes \mathcal{G}^T)\text{vec}\Sigma_\nu, \tag{C.1}$$

where we used that $\mathcal{B}\mathcal{B}^T = \mathcal{G}^T \Sigma_\nu \mathcal{G}$.

Note that $\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)$ is nonsingular. For $A_1 = \cdots = A_p = 0$, the matrix $\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}$ is triangular with ones on its diagonal. Thus, in view of the particular form of $\mathcal{G}$, $(\mathcal{G} \otimes \mathcal{G})\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}\left(\mathcal{G}^T \otimes \mathcal{G}^T\right)$ is a principal submatrix of $\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}$ with the same property and is therefore nonsingular. $(\mathcal{G} \otimes \mathcal{G})\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}\left(\mathcal{G}^T \otimes \mathcal{G}^T\right)$ is a function rational in $\left(A_1, \ldots, A_p\right)$ having no poles. Thus the set of zeros of this function is a proper algebraic set on Θ not depending on Σ_ν. On the complement of this proper algebraic set we have

$$\operatorname{vec}\Sigma_\nu = \left((\mathcal{G} \otimes \mathcal{G})\left(I_{(np)^2} - (\mathcal{A} \otimes \mathcal{A})\right)^{-1}\left(\mathcal{G}^T \otimes \mathcal{G}^T\right)\right)^{-1} \operatorname{vec}\gamma(0). \qquad \textbf{(C.2)}$$

APPENDIX D: Proof of Theorem 5

The proof is along the same lines as the proof of Theorem 2. A point $\theta^* \in \Theta_{(p_1 \ldots p_n)}$ where $\bar{Z}$ has full row rank is constructed as follows: Let

$$a_i(z) = e_i - \left[A_1\right]_{\bullet,i} z - \cdots - \left[A_p\right]_{\bullet,i} z^p$$

denote the i-th column of $a(z)$ and let

$$C_E = \left(\left[A_{p_1}\right]_{\bullet,1}, \ldots, \left[A_{p_n}\right]_{\bullet,n}\right)$$

be the so-called column-end matrix of $a(z)$. Then we take

$$C_E = \rho C, \quad \rho \in (0,1),$$

where C is the circulant defined in the proof of Theorem 2 and

$$\left[A_k\right]_{\bullet,i} = 0, \quad 0 < k < p_i;\; i = 1, \ldots, n$$

with $b = e_1$. Then again, $\bar{\Gamma}_p = S\Gamma_p S^T$ can be shown to be diagonal and nonsingular and thus $\det\left(\bar{Z}\right) \neq 0$ holds. Once the system parameters are unique, Σ_ν is obtained in the same way as in the proof of Theorem 3.

APPENDIX E: Proof of Theorem 6

We first consider the AR process $\left(y_t^r\right)$ with parameters not prescribed zero $(\bar{A}_1^r, \ldots, \bar{A}_p^r) = S_1(\bar{A}_1, \ldots, \bar{A}_p)$. Let us define $\bar{x}_t^r$ as the state of a quasi-companion form of y_t^r. Obviously, using Theorem 5 we have g-identifiability of $(\bar{A}_1^r, \ldots, \bar{A}_p^r)$ since

$$\begin{aligned}\bar{Z}^r &= \mathbb{E}\left(\bar{x}_t^r\left(\left(y_{t-1}^{rf}\right)^T, \ldots, \left(y_{t-np+s}^{rf}\right)^T\right)\right)\\ &= \left(\bar{K}^r, \bar{\mathcal{A}}^r\bar{K}^r, \left(\bar{\mathcal{A}}^r\right)^2\bar{K}^r, \ldots, \left(\bar{\mathcal{A}}^r\right)^{np-s-1}\bar{K}^r\right),\end{aligned}$$

where $\bar{K}^r = \bar{\Gamma}_p^r\begin{pmatrix} I_{(n_r)_f} \\ 0 \end{pmatrix}$ and $\bar{\Gamma}_p^r = \mathbb{E}\left(\bar{x}_t^r(\bar{x}_t^r)^T\right)$, is generically of full column rank. Thus we are left to show g-identifiability of the remaining rows $(\bar{A}_1^z, \ldots, \bar{A}_p^z) = S_2(\bar{A}_1, \ldots, \bar{A}_p)$, which is easily done since $y_t^z = (\bar{A}_1^z, \dots, \bar{A}_p^z)\bar{x}_t^r + b^z \varepsilon_t$, where b^z are the rows of b corresponding to y_t^z, and thus

$$\mathbb{E}\left(y_t^z \left(\left(y_{t-1}^{rf}\right)^T, \dots, \left(y_{t-np+s}^{rf}\right)^T\right)\right) = \left(\bar{A}_1^z, \dots, \bar{A}_p^z\right) \underbrace{\mathbb{E}\left(\bar{x}_t^r \left(\left(y_{t-1}^{rf}\right)^T, \dots, \left(y_{t-np+s}^{rf}\right)^T\right)\right)}_{=\bar{Z}^r}.$$

Therefore $(\bar{A}_1, \dots, \bar{A}_p)$ is identifiable if $\bar{Z}^r$ has full row rank which is generic. Once the system parameters are unique, Σ_ν is obtained in the same way as in the proof of Theorem 3.

APPENDIX F: Proof of Theorem 7

First note that the McMillan degree of $k(z^2)$ is equal to the rank of the Hankel matrix of covariances $\mathcal{H}^\gamma$: Let $k(z^2) = \sum_{j=0}^{\infty} k_{2j} z^{2j}$ be the power series expansion of $k(z^2)$. As $\tilde{\gamma}(2j) = \mathbb{E}(\tilde{y}_{2j}\tilde{y}_0^T) = \sum_{i=0}^{\infty} k_{2j+2i} k_{2i}^T$ we have

$$\underbrace{\begin{pmatrix} \tilde{\gamma}(2) & \tilde{\gamma}(4) & \tilde{\gamma}(6) & \cdots \\ \tilde{\gamma}(4) & \tilde{\gamma}(6) & \tilde{\gamma}(8) & \cdots \\ \tilde{\gamma}(6) & \tilde{\gamma}(8) & \tilde{\gamma}(10) & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^\gamma} = \underbrace{\begin{pmatrix} k_2 & k_4 & k_6 & \cdots \\ k_4 & k_6 & k_8 & \cdots \\ k_6 & k_8 & k_{10} & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k} \begin{pmatrix} k_0^T & & & \\ k_2^T & k_0^T & & \\ k_4^T & k_2^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix},$$

where the second matrix on the right hand side is of full row rank since $k(z^2)$ is miniphase. Thus we have that $\mathrm{rk}\mathcal{H}^\gamma = \mathrm{rk}\mathcal{H}^k$ holds. Since a spectral miniphase factor of the spectral density $f_Y(z^2)$ of $(Y_t | t \in 2\mathbb{Z})$ has McMillan degree np, the McMillan degree of $k(z^2)$ must be smaller than or equal to np.

Thus it remains to prove that a finite submatrix of $\mathcal{H}^\gamma$ has rank np:

$$\begin{aligned} \mathcal{H}_{np}^\gamma &= \begin{pmatrix} \tilde{\gamma}(2) & \tilde{\gamma}(4) \dots & \tilde{\gamma}(2np) \\ \tilde{\gamma}(4) & \tilde{\gamma}(6) \dots & \\ \vdots & \vdots \ \ddots & \\ \tilde{\gamma}(2np) & & \tilde{\gamma}(4np-2) \end{pmatrix} \in \mathbb{R}^{(n+n_f)np \times (n+n_f)np} \\ &= \mathbb{E}\begin{pmatrix} \tilde{y}_{t+2} \\ \vdots \\ \tilde{y}_{t+2np} \end{pmatrix} \left(\tilde{y}_t^T \cdots \tilde{y}_{t-2(np-1)}^T\right) \\ &= \mathbb{E}\begin{pmatrix} y_{t+2} \\ y_{t+1}^f \\ y_{t+4} \\ y_{t+3}^f \\ \vdots \\ y_{t+2np} \\ y_{t+2np-1}^f \end{pmatrix} \left(y_t^T \left(y_{t-1}^f\right)^T \cdots y_{t-2np+2}^T \left(y_{t-2np+1}^f\right)^T\right) \end{aligned}$$

$$
= \left(\begin{array}{cc|cc|c|cc}
\gamma(2) & [\gamma(3)]_{\cdot,1:n_f} & \gamma(4) & [\gamma(5)]_{\cdot,1:n_f} & \cdots & \gamma(2np) & [\gamma(2np+1)]_{\cdot,1:n_f} \\
[\gamma(1)]_{1:n_f,\cdot} & \gamma^{ff}(2) & [\gamma(3)]_{1:n_f,\cdot} & \gamma^{ff}(4) & \cdots & [\gamma(2np-1)]_{1:n_f,\cdot} & \gamma^{ff}(2np) \\
\hline
\gamma(4) & [\gamma(5)]_{\cdot,1:n_f} & \gamma(6) & [\gamma(7)]_{\cdot,1:n_f} & \cdots & \gamma(2np+2) & [\gamma(2np+3)]_{\cdot,1:n_f} \\
[\gamma(3)]_{1:n_f,\cdot} & \gamma^{ff}(4) & [\gamma(5)]_{1:n_f,\cdot} & \gamma^{ff}(6) & \cdots & [\gamma(2np+1)]_{1:n_f,\cdot} & \gamma^{ff}(2np+2) \\
\hline
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
\hline
\gamma(2np) & [\gamma(2np+1)]_{\cdot,1:n_f} & \gamma(2np+2) & [\gamma(2np+3)]_{\cdot,1:n_f} & \cdots & \gamma(4np-2) & [\gamma(4np-1)]_{\cdot,1:n_f} \\
[\gamma(2np-1)]_{1:n_f,\cdot} & \gamma^{ff}(2np) & [\gamma(2np+1)]_{1:n_f,\cdot} & \gamma^{ff}(2np+2) & \cdots & [\gamma(4np-3)]_{1:n_f,\cdot} & \gamma^{ff}(4np-2)
\end{array}\right)
$$

1

$$
= \left(\begin{array}{cc|cc|c|cc}
\left(I_n\ 0\right)\mathcal{A}^{2}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{3}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{4}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{5}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \cdots & \left(I_n\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+1}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} \\
\left(I_{n_f}\ 0\right)\mathcal{A}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{3}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{4}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \cdots & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np-1}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} \\
\hline
\left(I_n\ 0\right)\mathcal{A}^{4}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{5}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{6}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{7}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & & \left(I_n\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+3}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} \\
\left(I_{n_f}\ 0\right)\mathcal{A}^{3}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{4}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{5}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{6}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np+1}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} \\
\hline
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
\hline
\left(I_n\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+1}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+3}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \cdots & \left(I_n\ 0\right)\mathcal{A}^{4np-2}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{4np-1}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} \\
\left(I_{n_f}\ 0\right)\mathcal{A}^{2np-1}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np+1}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix} & \cdots & \left(I_{n_f}\ 0\right)\mathcal{A}^{4np-3}\Gamma_p\begin{pmatrix} I_n \\ 0 \end{pmatrix} & \left(I_{n_f}\ 0\right)\mathcal{A}^{4np-2}\Gamma_p\begin{pmatrix} I_{n_f} \\ 0 \end{pmatrix}
\end{array}\right)
$$

We look at the following submatrix of $\mathcal{H}_{np}^{\gamma}$

$$\begin{pmatrix} \left(I_n\ 0\right)\mathcal{A}^2\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^4\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^6\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \cdots & \left(I_n\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} \\ \left(I_n\ 0\right)\mathcal{A}^4\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^6\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^8\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \cdots & \left(I_n\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} \\ \left(I_n\ 0\right)\mathcal{A}^6\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^8\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{10}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \cdots & \\ \vdots & \vdots & \vdots & & \\ \left(I_n\ 0\right)\mathcal{A}^{2np}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & \left(I_n\ 0\right)\mathcal{A}^{2np+2}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} & & & \left(I_n\ 0\right)\mathcal{A}^{4np-2}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \left(I_n\ 0\right)\mathcal{A}^2 \\ \left(I_n\ 0\right)\mathcal{A}^4 \\ \left(I_n\ 0\right)\mathcal{A}^6 \\ \vdots \\ \left(I_n\ 0\right)\mathcal{A}^{2np} \end{pmatrix} \left(\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix}\ \mathcal{A}^2\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix}\ \mathcal{A}^4\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix}\ \cdots\ \mathcal{A}^{2np-2}\Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix}\right) = \mathcal{O}\mathcal{C} = \mathcal{H} \quad \textbf{(F.1)}$$

Since we assumed that the eigenvalues are nonzero, $\lambda_i \neq 0$, and for eigenvalues $\lambda_i \neq \lambda_j$ of $\mathcal{A}$ $\lambda_i^2 \neq \lambda_j^2$ holds, it is easy to see that q_i is an eigenvector of $\mathcal{A}^2$ if and only if q_i is an eigenvector of $\mathcal{A}$.

Observe that $\mathcal{O}$ is of full column rank. To show this we are using the PBH Test, see Kailath (1980), and the fact that for any right eigenvector q_i of $\mathcal{A}$ the first n components are not all equal to zero, as shown in Anderson et al. (2012, Lemma 2):

$$\left(\begin{pmatrix}\mathcal{A}^2 - \lambda_i^2 I\end{pmatrix} \\ \left(I_n\ 0\right)\mathcal{A}^2\right) q_i = \begin{pmatrix} 0 \\ \left[\lambda_i^2 q_i\right]_{1:n} \neq 0 \end{pmatrix}.$$

Also observe that $\mathcal{C}$ is of full row rank if $\Gamma_p > 0$. Again we are using the PBH Test: We have to test for all left eigenvectors q_i^T of $\mathcal{A}^2$ or equivalently $\mathcal{A}$ that

$$q_i^T\left(\left(\mathcal{A}^2 - \lambda_i^2 I\right), \Gamma_p\begin{pmatrix}I_n\\0\end{pmatrix}\right) = \left(0, q_i^T\begin{pmatrix}\gamma(0)\\ \vdots \\ \gamma(1-p)\end{pmatrix}\right) \neq 0.$$

Thus if q_i is orthogonal to $\begin{pmatrix}\gamma(0)\\ \vdots \\ \gamma(1-p)\end{pmatrix}$ also

$$q_i^T\mathcal{A}^j\begin{pmatrix}\gamma(0)\\ \vdots \\ \gamma(1-p)\end{pmatrix} = 0, \forall j \in \mathbb{N},$$

holds which implies $q_i^T\Gamma_p = 0$ which is in contradiction to $\Gamma_p > 0$. Therefore for all eigenvectors q_i of $\mathcal{A}^2$

$$\left(0, q_i^T\begin{pmatrix}\gamma(0)\\ \vdots \\ \gamma(1-p)\end{pmatrix}\right) \neq 0.$$

Now, according to Hannan and Deistler (2012, Thm. 2.3.2) $\mathcal{H}$ has rank np and therefore the same holds for $\mathcal{H}^k$.

APPENDIX G: Proof of Theorem 8

First we show, using a notation more general than the notation in Theorem 8, how to determine a minimal state space realization of a stable transfer function $k(z)$ from the Hankel matrix of this transfer function, see Ho and Kalman (1966), Akaike (1974) or Hannan and Deistler (2012, Chap. 2). Let (ε_t) be the inputs such that $y_t = k(z)\varepsilon_t$. Clearly, we have

$$\begin{pmatrix} y_t \\ y_{t+1} \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} k_1 & k_2 & k_3 & \cdots \\ k_2 & k_3 & k_4 & \cdots \\ k_3 & k_4 & k_5 & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix} + \begin{pmatrix} k_0 & & & \\ k_1 & k_0 & & \\ k_2 & k_1 & k_0 & \\ \vdots & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t+1} \\ \vdots \end{pmatrix}.$$

A minimal state space system can be obtained as follows: Let S be a selector matrix selecting basis rows of the row space of $\mathcal{H}^k$. Define the state as

$$x_t = S \underbrace{\begin{pmatrix} k_1 & k_2 & k_3 & \cdots \\ k_2 & k_3 & k_4 & \cdots \\ k_3 & k_4 & k_5 & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k_\alpha} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix}.$$

A state equation of a minimal state space system can be obtained as follows:

$$\begin{aligned} x_{t+1} &= S \underbrace{\begin{pmatrix} k_1 & k_2 & k_3 & \cdots \\ k_2 & k_3 & k_4 & \cdots \\ k_3 & k_4 & k_5 & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k_\alpha} \begin{pmatrix} \varepsilon_{t} \\ \varepsilon_{t-1} \\ \vdots \end{pmatrix} \\ &= S \underbrace{\begin{pmatrix} k_2 & k_3 & k_4 & \cdots \\ k_3 & k_4 & k_5 & \cdots \\ k_4 & k_5 & k_6 & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k_{\alpha+n}} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix} + S \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \vdots \end{pmatrix} \varepsilon_t \\ &= \underbrace{A\, S \underbrace{\begin{pmatrix} k_1 & k_2 & k_3 & \cdots \\ k_2 & k_3 & k_4 & \cdots \\ k_3 & k_4 & k_5 & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}^k_\alpha} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix}}_{x_t} + S \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \vdots \end{pmatrix} \varepsilon_t . \end{aligned}$$

Thus the state transition matrix A can be uniquely obtained as a solution of the linear equation system

$$\mathcal{H}^k_{\alpha+n} = A\mathcal{H}^k_{\alpha}. \tag{G.1}$$

The observation equation can be obtained as follows:

$$\begin{aligned} y_t &= \begin{pmatrix} k_1 & k_2 & k_3 & \cdots \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix} + k_0\varepsilon_t \\ &= C \underbrace{\mathcal{H}^k_{\alpha} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix}}_{x_t} + k_0\varepsilon_t. \end{aligned}$$

Therefore C can be determined as the unique solution of the linear equation system

$$\begin{pmatrix} k_1 & k_2 & k_3 & \cdots \end{pmatrix} = C\mathcal{H}^k_{\alpha}. \tag{G.2}$$

In the next step we discuss the connection of the row spaces of the Hankel matrices of the transfer functions $k(z)$ and $\tilde{k}(z) = \sum_{j=0}^{\infty} \tilde{k}_j z^j$, respectively, where now $k(z)$ is miniphase and $\tilde{k}(z)$ corresponds to the same spectrum, has the same McMillan degree as $k(z)$, but is not necessarily miniphase and has not necessarily the same number of columns as $k(z)$. We use the relation of the Hankel matrices of the transfer functions $k(z)$ and $\tilde{k}(z)$ and the Hankel matrix $\mathcal{H}^{\gamma}$ of the covariances of $(y_t \,|\, t \in \mathbb{Z})$:

$$\mathcal{H}^{\gamma} = \mathcal{H}^{k} \begin{pmatrix} k_0^T & & & \\ k_1^T & k_0^T & & \\ k_2^T & k_1^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

and

$$\mathcal{H}^{\gamma} = \mathcal{H}^{\tilde{k}} \begin{pmatrix} \tilde{k}_0^T & & & \\ \tilde{k}_1^T & \tilde{k}_0^T & & \\ \tilde{k}_2^T & \tilde{k}_1^T & \tilde{k}_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}.$$

Note that the left kernels of $\mathcal{H}^k$ and $\mathcal{H}^{\tilde{k}}$ are subsets of the left kernel of $\mathcal{H}^{\gamma}$. Thus we have the following: Let S be a selector matrix selecting a basis of the row space of $\mathcal{H}^{\gamma}$. Then S also selects bases of the row spaces of $\mathcal{H}^k$ and $\mathcal{H}^{\tilde{k}}$ since the McMillan degrees of k and $\tilde{k}$ are the same and equal to the rank of $\mathcal{H}^k$, $\mathcal{H}^{\tilde{k}}$, and $\mathcal{H}^{\gamma}$, respectively, see (Hannan and Deistler, 2012).

Let $\mathcal{H}_\alpha^k = S\mathcal{H}^k$ and $\mathcal{H}_{\alpha+n}^k$ be defined as above. Define $\mathcal{H}_\alpha^{\tilde{k}} = S\mathcal{H}^{\tilde{k}}$ and $\mathcal{H}_{\alpha+n}^{\tilde{k}}$ accordingly. Then

$$\mathcal{H}_\alpha^k \begin{pmatrix} k_0^T & & & \\ k_1^T & k_0^T & & \\ k_2^T & k_1^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix} = \mathcal{H}_\alpha^{\tilde{k}} \begin{pmatrix} \tilde{k}_0^T & & & \\ \tilde{k}_1^T & \tilde{k}_0^T & & \\ \tilde{k}_2^T & \tilde{k}_1^T & \tilde{k}_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

and

$$\mathcal{H}_{\alpha+n}^k \begin{pmatrix} k_0^T & & & \\ k_1^T & k_0^T & & \\ k_2^T & k_1^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix} = \mathcal{H}_{\alpha+n}^{\tilde{k}} \begin{pmatrix} \tilde{k}_0^T & & & \\ \tilde{k}_1^T & \tilde{k}_0^T & & \\ \tilde{k}_2^T & \tilde{k}_1^T & \tilde{k}_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

hold. Let $\tilde{A}$ be the state transition matrix of a minimal state space realization of $\tilde{k}(z)$ such that

$$\mathcal{H}_{\alpha+n}^{\tilde{k}} = \tilde{A}\mathcal{H}_\alpha^{\tilde{k}}.$$

This implies

$$\underbrace{\mathcal{H}_{\alpha+n}^k \begin{pmatrix} k_0^T & & & \\ k_1^T & k_0^T & & \\ k_2^T & k_1^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}_{\alpha+n}^\gamma} = \tilde{A}\, \underbrace{\mathcal{H}_\alpha^k \begin{pmatrix} k_0^T & & & \\ k_1^T & k_0^T & & \\ k_2^T & k_1^T & k_0^T & \\ \vdots & \vdots & & \ddots \end{pmatrix}}_{\mathcal{H}_\alpha^\gamma}.$$

Since k_0 has full column rank, the last equation above implies

$$\mathcal{H}_{\alpha+n}^k = \tilde{A}\mathcal{H}_\alpha^k.$$

Thus $\tilde{A}$ is also the state transition matrix of a minimal state space realization of $k(z)$.

Analogously, we treat the observation equation and obtain a unique $\tilde{C}$.

As is well known, for every minimal realization of $\tilde{k}(z)$ the matrices A^* and C^* are related to $\tilde{A}$ and $\tilde{C}$ via a basis change. This proves the theorem.

Author Query Form

Journal: Econometric Theory

Article: S0266466615000043

Dear Author, During the preparation of your manuscript for publication, the questions listed below have arisen. The numbers pertain to the numbers in the margin of the proof. Please attend to these matters and return the form with this proof.

Many thanks for your assistance.

Query no.	Query	Response
Q1	Please note that the acknowledgment section has been moved to the footnote on the first page as per journal style. Please check.	
Q2	Please provide editors name (if any) and publisher name for the references "Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, & Zamani (2012), Bernanke, Gertler, Watson, Sims, & Friedman (1997) and Phillips (1976)".	
Q3	Please provide the publisher name for the reference "Hannan & Deistler (2012)".	
Q4	Please provide the page range for the reference "Zadrozny (1990a)".	
Q5	Please provide the volume number for the reference "Zadrozny (1990b)".	