CESifo Working Papers

www.cesifo.org/wp

Extended Yule-Walker Identification of Varma Models with Single- or Mixed-Frequency Data.

Peter A. Zadrozny

CESIFO WORKING PAPER NO. 5884
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
MAY 2016

ISSN 2364-1428

Extended Yule-Walker Identification of Varma Models with Single- or Mixed-Frequency Data.

Abstract

Chen and Zadrozny (1998) developed the linear extended Yule-Walker (XYW) method for determining the parameters of a vector autoregressive (VAR) model with available covariances of mixed-frequency observations on the variables of the model. If the parameters are determined uniquely for available population covariances, then, the VAR model is identified. The present paper extends the original XYW method to an extended XYW method for determining all ARMA parameters of a vector autoregressive moving-average (VARMA) model with available covariances of single- or mixed-frequency observations on the variables of the model. The paper proves that under conditions of stationarity, regularity, miniphaseness, controllability, observability, and diagonalizability on the parameters of the model, the parameters are determined uniquely with available population covariances of single- or mixed-frequency observations on the variables of the model, so that the VARMA model is identified with the single- or mixed-frequency covariances.

JEL-Codes: C320, C800.

Keywords: block-Vandermonde eigenvectors of block-companion state-transition matrix of state-space representation, matrix spectral factorization.

Peter A. Zadrozny
Bureau of Labor Statistics
2 Massachusetts Ave., NE, Room 3105
USA – Washington, DC 201212
zadrozny.peter@bls.gov

April 13, 2016
The paper represents the author's views and does not necessarily represent any official positions of the Bureau of Labor Statistics. The paper was presented at the following conferences and seminars: NBER-NSF Time Series, Washington, Sep. 2013; $(EC)^2$ on "Mixed-Frequency Econometrics," Nicosia, Dec. 2013; SNDE, New York City, Apr. 2014; CEF, Oslo, June 2014; JSM, Boston, Aug. 2014; KOF, Zurich, Dec. 2014; CFE, Pisa, Dec. 2014; NBP on "Identification in Macroeconomics," Warsaw, Dec. 2014. Affiliated as Research Fellow with the Center for Financial Studies (CFS), Goethe University, Frankfurt, Germany, and with the Center for Economic Studies and Ifo Institute for Economic Research (CESifo), Munich, Germany. The paper has benefitted from comments by Manfred Deistler, Eric Ghysels, Roderick McCrorie, Tucker McElroy, and anonymous referees. Forthcoming in Journal of Econometrics issue on "Econometric Analysis of Mixed Frequency Data Sampling."

1. Introduction.

There has always correctly been a "real-time" interest in using all available information in an econometric analysis. Until recently, econometric analysis of data indexed by discrete-time periods has focused almost exclusively on single-frequency data (SFD), in which all variables are indexed at the same time interval. However, in practice, different variables have been available at different time intervals, i.e., at mixed frequencies. As more variables are available at higher frequencies, data sets with greater mixtures of frequencies are available for econometric analysis. The desire to use the best available analysis has motivated research on econometric methods for mixed-frequency data (MFD). There is also the real-time matter of different variables being available at different lags, but this will not be considered here. For further discussion of these issues, see Ghysels (2012).

Initially, only regression was used in econometric analysis with MFD, usually monthly-quartely data (Friedman, 1962; Miller and Chin, 1996). More recently, univariate and multivariate or vector autoregressive moving-average (ARMA and VARMA) models have been increasingly used in econometric analysis with SFD or MFD. In econometrics, VARMA models were first estimated with MFD using maximum likelihood (Zadrozny, 1988, 1990a,b), but maximum likelihood estimation (MLE) is effective only if good starting values are available for the parameters to be estimated so that iterations converge correctly and this is often difficult to do unless the model is "small" and has relatively few variables and, hence, few parameters to be estimated. In response, Chen and Zadrozny (1998) developed the linear extended Yule-Walker (XYW) method for determining the parameters of a VAR model, which uses available covariances of MFD and has the computational simplicity of least squares, and illustrated XYW's accuracy relative to MLE. The XYW method overcomes the computational problem of not being able to evaluate standard Yule-Walker equations because autocovariances at high-frequency lags of variables observed at low frequencies are unavailable. Although VAR models now dominate linear multivariate models used for modelling and analyzing economic time series, including an MA term in a model often allows it to fit data more accurately and parsimoniously (Box and Jenkins, 1976).

XYW can be thought of not just as an AR-parameter estimation method, but more generally as a method that takes data covariances presumed to be generated by a VAR model as inputs and determines the AR parameters as outputs. If the covariance inputs are true population covariances and the outputs are unique, then, the outputs are the true model parameters and the VAR model is identified; if the covariance inputs are consistent sample estimates and the outputs are unique, then, the outputs are consistent parameter estimates. Chen and Zadrozny (1998) introduced XYW as an estimation method with sample covariances but did not prove, under certain conditions, that XYW is feasible (computationally implementable) or that XYW determines unique AR parameter outputs for true-population or consistent-sample covariance inputs. Anderson et al. (2012) proved this for a general VAR model and a particular MFD case, but only for a "generic" set of parameters.

The present paper makes two contributions. First, the paper extends the original XYW method to an extended XYW method that determines all ARMA parameters of a VARMA model with available covariances of its variables observed with SFD or MFD. Second, the paper proves that if the parameters of the model satisfy conditions I-VI of stationarity, regularity, miniphaseness, controllability, observability, and diagonalizability, then, the extended XYW method produces unique ARMA parameter values and the VARMA model is identified (not just "generically") with population covariances of its variables. Although the paper is not directly concerned with parameter estimation, the extended XYW method becomes a consistent method for estimating VARMA parameters simply by replacing population covariances with consistent sample covariances. However, experience with the XYW method (Chen and Zadrozny, 1998) suggests that such a consistent estimation method is unlikely to be accurate in small samples but that a generalized method of moments (GMM) extension of the method could be accurate in small samples. However, such an extension is beyond the scope of this paper and is left for the future.

The extended XYW method solves one linear system to determine the AR parameters and solves two linear systems and does one matrix spectral factorization to determine the MA parameters. Spectral factorization is a linear operation except for an initial step of computing eigenvalues, which can be done reliably, accurately, and quickly using the QR algorithm (Golub and Van Loan, 1996; Zadrozny, 1998). The key to the proof in the paper is exploiting the block-Vandermonde structure of eigenvectors of a block-companion-form state-transition matrix of a state-space representation of a VARMA model.

Identification can be local or global. By definition, different sets of parameters of a model that generate identical covariances of variables of the model are observationally equivalent. If a point of a set of observationally equivalent parameters is isolated in the set, then, the model is locally identified at that point; if the set of observationally equivalent parameters is a single point, then, the model is globally identified. The paper assumes that global identification problems have been resolved by other assumptions and proves that, if the model satisfies conditions I-VI, then, the model's parameters are locally (and globally) identified with population covariances of variables of the model observed with SFD or MFD.

The result is first proved for SFD and is, then, adapted to MFD. The adaptation is straightforward, because it requires only reducing derived equations and requires no additional derivations. In the paper, SFD means that all variables of a model are observed at the same discrete-time frequency at which the model operates and MFD means that some of the variables are observed at the same discrete-time frequency at which the model operates and others are observed at one or more lower frequencies. Although the paper considers only the above definition of SFD, SFD could also mean that all variables are observed at the same discrete-time frequency which is lower than the frequency at which the model operates.

For the second definition of SFD or for MFD, in the limit as its operating frequency goes to infinity, a discrete-time model approaches a continuous-time model observed with discrete-time data (Zadrozny, 1988). Both discrete- and continuous-time models can be locally identified but not globally identified due to aliasing. Although aliasing has been considered mostly for continuous-time models observed with discrete-time data (Phillips, 1973; Hansen and Sargent, 1983), aliasing can also occur in discrete-time models observed with discrete-time data. Aliasing occurs when state-transition matrices of different but observationally equivalent models have different eigenvalues.

One general resolution of aliasing is to choose the "least noisy" observationally-equivalent model in the sense of having the least spectral power at high frequencies. For example, in Anderson et al.'s (2012) model, a_{ss} is a parameter whose absolute value but not sign is identified, hence, is locally identified but not globally identified when disturbances are uncorrelated. Because a_{ss} is also an eigenvalue of a state-transition matrix of the model, the global unidentification is also an aliasing unidentification. Because positive a_{ss} contributes spectral power at the zero frequency and negative a_{ss} contributes spectral power at the Nyquist frequency, choosing a_{ss} to be positive results in the least noisy and globally identified model. Of course, an application's subject matter could offer a more compelling reason for resolving unidentification, including aliasing. Henceforth in the paper, any global unidentification is assumed to have been resolved with additional assumptions, so that "identification" means both local and global identification.

Priestley (1981, pp. 800-804) reviewed the literature on identification of a VARMA model with population covariances of its variables observed with (first-definition) SFD and attributed results principally to Hannan (1969, 1970, 1976, 1979) and secondarily to Akaike (1974). See also Hannan and Deistler (1986). Like here, Hannan assumed that the parameters of the model satisfy conditions I-IV of stationarity, regularity, miniphaseness, and controllability. Hannan didn't and didn't need to assume observability condition V, because, as explained in section 2, observability holds for any VARMA model observed with SFD. Hannan proved that, under these conditions, a VARMA model is identified with population covariances of its variables observed with SFD. Hannan's proof is different from the present one: whereas Hannan used mathematical analysis, we use only linear algebra. Hannan didn't state and use some version of diagonalizability condition VI, which appears to be necessary in the proof here. Using the same conditions I-IV and essentially the same argument as here, Akaike (1974) proved that the AR parameters of a VARMA model are identified by population covariances of its variables observed with SFD and asserted, but didn't prove, that the MA parameters of the model are identified by unique spectral factorization. The present paper contributes to this literature by being the first one to prove that, under conditions I-VI of stationarity, regularity, miniphaseness, controllability, observability, and diagonalizability, a VARMA model is identified (without the qualification "generically") by population covariances of its variables observed with MFD. Although the conditions are individually necessary for identification in different parts of the proof, the paper proves only that the conditions as a whole are sufficient for identification. The question of necessity of the conditions for identification is discussed further in concluding section 5.

The paper continues as follows. Section 2 states the general VARMA model in original and state-space form and states conditions I-VI assumed for the model. Section 3 derives backward Yule-Walker equations (BYWE) for a model observed with SFD, proves that under conditions I-V the BYWE can be solved for unique values of the AR parameters of the model, and adapts the BYWE and their solution to MFD. Section 4 derives forward Yule-Walker equations (FYWE) for a model observed with SFD, proves that under conditions

I-VI the FYWE can be solved for unique values of the MA parameters of the model, and adapts the FYWE and their solution to MFD. The paper concludes in section 5 with discussion of common left AR and MA factors, necessity of the identifying conditions, numerical illustration of identification, and identifcation of structural parameters.

2. Statement of VARMA model and assumptions on it.

We write a general VARMA(r,q) model in VARMA(p,p-1) form as

$$(2.1)\qquad y_t = A_1y_{t-1} + \ldots + A_py_{t-p} + B_0\varepsilon_t + B_1\varepsilon_{t-1} + \ldots + + B_{p-1}\varepsilon_{t-p+1}$$

and define its components as follows: y_t denotes an n×1 vector of observed variables; r and q denote any assumed nonnegative integers, such that at least one of r or q is positive; p = max(r,q+1); A_i (i = 1, ..., p) denote n×n matrices of AR parameters, $A_r \neq 0_{nxn}$ ($O_{j\times k}$ denotes the j×k zero matrix), and intermediate (i = 1, ..., r-1) and trailing (i = r+1, ..., p) A_i, respectively, may be and are zero; B_j (j = 0, ..., p-1) denote n×n matrices of MA parameters, $B_q \neq 0_{nxn}$, and intermediate (j = 1, ..., q-1) and trailing (j = q+1, ..., p-1) B_j, respectively, may be and are zero; ε_t denotes an n×1 vector of unobserved disturbances ~ IID(0_{nx1},I_n), where I_n denotes the n×n identity matrix. All quantities in the paper are real valued except possibly eigenvalues, eigenvectors, latent roots, and latent vectors, which may be complex valued.

We assume that the model satisfies conditions I-III of stationarity, regularity, and miniphaseness:

<u>Condition</u> <u>I</u>: VARMA model (2.1) is <u>stationary</u>, i.e., if λ is a real- or complex-valued scalar root of the AR characteristic equation $|A(\lambda)| = |I_n\lambda^r - A_1\lambda^{r-1} - \ldots - A_r| = 0$, then, $|\lambda| < 1$, where $|\cdot|$ denotes a determinant or an absolute value (modulus);

<u>Condition</u> <u>II</u>: VARMA model (2.1) is <u>regular</u>, i.e., B_0 is lower triangular and nonsingular;

<u>Condition III</u>: VARMA model (2.1) is <u>miniphase</u>, i.e., if λ is a real- or complex-valued scalar root of the MA characteristic equation $|B(\lambda)| = |B_0\lambda^q + B_1\lambda^{q-1} + \dots + B_q| = 0$, then, $|\lambda| \le 1$.

Miniphaseness extends invertibility to allow MA roots on the unit circle. An estimated VARMA model almost never has MA roots on the unit circle unless restrictions on it imply them. For example, suppose that n variables in $\tilde{y}_t = (\tilde{y}_{1t}^T, \tilde{y}_{2t}^T)^T$ (superscript T denotes vector or matrix transposition) are generated by an unrestricted (except for conditions I-VI) VARMA model estimated using data in which the first n_1 variables are observed directly as $y_{1t} = \tilde{y}_{1t}$ and the last n_2 variables are observed temporally aggregated as $y_{2t} = \tilde{y}_{2t} + \dots + \tilde{y}_{2,t-m}$, for some $m \ge 1$. Then, to be estimated with the partly aggregated data, the model must be extended to a VARMA(r,q+m) model with the same AR part and an MA part with characteristic equation $B(\lambda)D(\lambda)$, where $D(\lambda) = I_n\lambda^m + D_1\lambda^{m-1} + \dots + D_m$ and $D_i = \begin{bmatrix} 0_{n_1\times n_1} & 0_{n_1\times n_2} \\ 0_{n_2\times n_1} & I_{n_2} \end{bmatrix}$, for $i = 1, \dots, m$. $D(\lambda)$ adds mn MA roots, mn_1 zero roots and mn_2 roots on the unit circle.

VARMA(p,p-1) form (2.1) has the following state-space representation comprising observation equation

$$(2.2) \qquad y_t = Hx_t, \quad H = [I_n, 0_{n\times n}, \dots, 0_{n\times n}] = n\times np,$$

where x_t denotes the np×1 state vector, and state equation

$$(2.3) \qquad x_t = Fx_{t-1} + G\varepsilon_t, \quad F = \begin{bmatrix} A_1 & I_n & \cdots & 0_{n\times n} \\ \vdots & 0_{n\times n} & \ddots & \vdots \\ \vdots & \vdots & \ddots & I_n \\ A_p & 0_{n\times n} & \cdots & 0_{n\times n} \end{bmatrix} = np\times np, \quad G = \begin{bmatrix} B_0 \\ \vdots \\ \vdots \\ B_{p-1} \end{bmatrix} = np\times n.$$

For K = 1, 2, ..., define

$$(2.4) \qquad C_K(F,G) = [G, \dots, F^{K-1}G] = np\times nK.$$

For K = np, $C_{np}(F,G)$ is called the controllability matrix. By the Cayley-Hamilton theorem, which says that every square matrix satisfies its own characteristic equation, $C_K(F,G)$ has maximum rank when $K = np$, so that $C_K(F,G)$ has full rank np for some K if and only if (iff) $\mathrm{rank}[C_{np}(F,G)] = np$.

A VARMA model is said to be controllable iff its controllability matrix has full rank, i.e., $\mathrm{rank}[C_{np}(F,G)] = np$. Hautus (1969) proved that $\mathrm{rank}[C_{np}(F,G)] = np$ iff, for any real- or complex-valued scalar λ,

$$(2.5) \qquad \mathrm{rank}[F - I_{np}\lambda,\ G] = np.$$

Controllability is often more easily proved by checking condition (2.5) than by checking $\mathrm{rank}[C_{np}(F,G)] = np$ directly. Kailath (1980, p. 135) called condition (2.5) the "PBH test," although Lancaster and Rodman (1995, p. 88) state that it was first proved by Hautus (1969).

The block-Vandermonde form (4.4) of the left (row) eigenvectors of the block-companion state-transition matrix F implies that condition (2.5) is equivalent to condition

$$(2.6) \qquad (\lambda_i)^{\max(r-q-1,0)}\, \xi_i^T B(\lambda_i) \neq 0_{1xn},$$

for $i = 1, \ldots, np$, where λ_i is an eigenvalue of F, ξ_i is a nonzero left latent (row) vector of $A(\lambda) = I_n\lambda^r - A_1\lambda^{r-1} - \ldots - A_r$ that satisfies $\xi_i^T A(\lambda_i) = 0_{1xn}$, and $B(\lambda) = B_0\lambda^q + \ldots + B_q$. The derivation of equation (6.2) in the appendix implies that conditions (2.5) and (2.6) are equivalent. Therefore, the conditions $\mathrm{rank}[C_{np}(F,G)] = np$, (2.5), and (2.6) are equivalent.

We assume that the model satisfies condition IV of controllability:

<u>Condition</u> <u>IV</u>: VARMA model (2.1) is <u>controllable</u>.

If $r \le q$, then, $n(q-r+1)$ zero eigenvalues of F are not AR roots that satisfy $|A(\lambda)| = 0$ but, in condition (2.6), $(\lambda_i)^{\max(r-q-1,0)} = 1$ for any zero or nonzero AR roots. If $r \ge q+1$, then, all eigenvalues of F are AR roots and must be nonzero for controllability to hold. In both cases, when AR roots are nonzero, their being distinct from MA roots is sufficient, but unnecessary, for condition (2.6) to hold. Controllability holds in most applications because AR roots are distinct from MA roots.

We have called ξ_i "latent" according to the theory of matrix polynomials. In this theory, the AR characteristic polynomial $A(\lambda)$ is called a lambda matrix. A root λ_i of the characteristic equation $|A(\lambda)| = 0$ is called a latent root. Just as an eigenvalue of a square matrix has a matching nonzero left (row) eigenvector, a latent root λ_i of $A(\lambda)$ has a matching nonzero left (row) latent vector ξ_i that satisfies $\xi_i^T A(\lambda_i) = 0_{1xn}$. Because F has the block-companion form (2.3), every latent root of $A(\lambda)$ is also an eigenvalue λ_i of F and vice versa if $r \geq q+1$; and, every left eigenvector z_i of F has the block-Vandermonde form (4.4), where ξ_i is a left latent vector of $A(\lambda_i)$. See Dennis et al. (1976).

Analogous to controllability, for L = 1, 2, ..., we define

$$(2.7) \qquad O_L(F,H) = [H^T, \ldots, (F^T)^{L-1}H^T]^T = nL\times np.$$

For L = np, $O_{np}(F,H)$ is called the observability matrix. By the Cayley-Hamilton theorem, $O_L(F,H)$ has maximum rank when L = np, so that $O_L(F,H)$ has full rank np, for some L, iff rank$[O_{np}(F,H)]$ = np. A VARMA model is said to be observable iff the observability matrix has full rank, hence, iff rank$[F^T - I_{np}\lambda, H^T]$ = np. Because F is asymmetric, it generally has different left and right eigenvectors for each eigenvalue, so there is generally no direct analogue of condition (2.6) for observability, obtained by replacing F and G with F^T and H^T in equation (2.5).

Controllability and observability come from dynamic system theory (Kwakernaak and Sivan, 1972; Anderson and Moore, 1979; Kailath, 1980). Controllability generally depends on all ARMA parameters, regardless how the model's variables are observed. Observability generally depends only on AR parameters and on how the model's variables are observed. For SFD, every VARMA model is observable, regardless of its AR parameter values, because $O_L(F,H)$ is unit lower triangular for $L \geq p$. Thus, it is unnecessary to assume that VARMA model (2.1) is observable for SFD, but it is generally necessary to assume that the model is observable for MFD.

Different lower bounds have been stated for L. In each case, the lower bound is a necessary but not necessarily a sufficient condition for an observability condition to hold. However, because L has no upper limit in identification, we may henceforth more simply state that "L is sufficiently large". Of course, in estimation, L is limited by sample size.

We assume that the model satisfies condition V of observability when its variables are observed with MFD:

<u>Condition V</u>: VARMA model (2.1) is <u>observable</u> for a sufficiently large L, for the MFD being considered.

Define the block-companion-form matrix

$$(2.8) \qquad \overline{\overline{B}} = \begin{bmatrix} \overline{B}_1 & I_n & \cdots & 0_{n\times n} \\ \vdots & 0_{n\times n} & \ddots & \vdots \\ \vdots & \vdots & \ddots & I \\ \overline{B}_q & 0_{n\times n} & \cdots & 0_{n\times n} \end{bmatrix} = nq\times nq,$$

where, for i = 1, ..., q, $\overline{B}_i = -B_i B_0^{-1}$. We assume that the model satisfies condition VI of diagonalizability:

Condition VI: $\overline{\overline{B}}$ is diagonalizable, i.e., has a linearly independent set of eigenvectors.

Distinct MA roots, equivalently distinct eigenvalues of $\overline{\overline{B}}$, imply that $\overline{\overline{B}}$ has a full set of nq linearly independent eigenvectors. For this reason, diagonalizability should hold in most applications.

Conditions I-VI are conventional and can be expected to hold in practice for all but a singular (measure zero) set of parameters.

3. Identification of AR parameters with backward Yule-Walker equations.

Let $C_k = Ey_t y_{t-k}^T$, for k = 0, ±1, ±2, ..., denote the k-th population covariance matrix of y_t and y_{t-k} generated by VARMA model (2.1), where E denotes unconditional expectation. C_k exists because the model is stationary and is skew symmetric, i.e., $C_k = C_{-k}^T$.

To obtain the backward Yule-Walker equations (BYWE) for SFD, postmultiply VARMA model (2.1) by "backward in time" y_{t-k}^T, for k = 0, ..., L ≥ 2p-1, take unconditional expectations, and obtain

$$(3.1) \qquad \begin{bmatrix} C_0 \\ \vdots \\ C_{p-1}^T \\ C_p^T \\ \vdots \\ C_L^T \end{bmatrix} = \begin{bmatrix} C_p & \cdots & C_1 \\ \vdots & & \vdots \\ C_1 & \cdots & C_{p-2}^T \\ C_0 & \cdots & C_{p-1}^T \\ \vdots & & \vdots \\ C_{L-p}^T & \cdots & C_{L-1}^T \end{bmatrix} \begin{bmatrix} A_p^T \\ \vdots \\ A_1^T \end{bmatrix} + \begin{bmatrix} \sum_{i=0}^{p-1} \Psi_i B_i^T \\ \vdots \\ \Psi_0 B_{p-1}^T \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{bmatrix},$$

where $\Psi_i = HF^iG$ denotes the i-th coefficient matrix of the Wold infinite MA representation of the model. We want to solve BYWE (3.1) for unique values of the AR parameters, $A_1, \ldots, A_p$. To do this, we skip the first p blocks ($k = 0, \ldots, p-1$) with MA terms and consider only further blocks ($k = p, \ldots, L$) without MA terms,

$$(3.2) \qquad \begin{bmatrix} C_0 & \cdots & C_{p-1}^T \\ \vdots & & \vdots \\ C_{L-p}^T & \cdots & C_{L-1}^T \end{bmatrix} \begin{bmatrix} A_p^T \\ \vdots \\ A_1^T \end{bmatrix} = \begin{bmatrix} C_p^T \\ \vdots \\ C_L^T \end{bmatrix}.$$

Consider equation (3.2) as $DX = E$. The equation can be solved for unique AR parameter values in X iff, for sufficiently large L, D has full (column) rank. A proof of this result goes as follows. State-space representation (2.2)-(2.3) implies that, for $k = 0, 1, \ldots,$

$$(3.3) \qquad C_k = HF^kVH^T = n\times n,$$

where, because the model is stationary, $V = Ex_tx_t^T$ exists and satisfies $V = \sum_{k=0}^{\infty} F^kGG^T(F^T)^k$ or, equivalently,

$$(3.4) \qquad V = [C_{np}(F,G), \ldots][C_{np}(F,G), \ldots]^T = np\times np.$$

V is symmetric positive semidefinite by its structure. Equation (3.4) and the Cayley-Hamilton theorem imply that V is positive definite iff the VARMA model is controllable, which has been assumed.

Because V is positive definite, it has the Cholesky factorization $V = RR^T$, where R is $np\times np$, lower triangular, nonsingular, and unique. Using $\tilde{F} = R^{-1}FR$, $\tilde{H} = HR$, and equation (3.3), system matrix D of equation (3.2) can be expressed as

$$(3.5) \qquad D = \begin{bmatrix} HVH^T & \cdots & HV(F^T)^{p-1}H^T \\ \vdots & & \vdots \\ HV(F^T)^{L-p}H^T & \cdots & HV(F^T)^{L-1}H^T \end{bmatrix} = \begin{bmatrix} \tilde{H}\tilde{H}^T & \cdots & \tilde{H}(\tilde{F}^T)^{p-1}\tilde{H}^T \\ \vdots & & \vdots \\ \tilde{H}(\tilde{F}^T)^{L-p}\tilde{H}^T & \cdots & \tilde{H}(\tilde{F}^T)^{L-1}\tilde{H}^T \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{H} \\ \tilde{H}\tilde{F}^T \\ \vdots \\ \tilde{H}(\tilde{F}^T)^{L-p} \end{bmatrix} \begin{bmatrix} \tilde{H}^T & \tilde{F}^T\tilde{H}^T & \cdots & (\tilde{F}^T)^{p-1}\tilde{H}^T \end{bmatrix} = O_{L-p+1}(\tilde{F}^T,\tilde{H})O_p(\tilde{F},\tilde{H})^T = n(L-p+1)\times np.$$

D has full rank np, for sufficiently large L, iff $O_{L-p+1}(\tilde{F}^T,\tilde{H}) = C_{L-p+1}(F,VH^T)^TR^{-T}$ and $O_p(\tilde{F},\tilde{H}) = O_p(F,H)R$ do. Because R is nonsingular, $O_{L-p+1}(\tilde{F}^T,\tilde{H})$ has full rank np, for sufficiently large L, iff $C_{L-p+1}(F,VH^T)$ does. $O_p(\tilde{F},\tilde{H})$ has full rank np, because R is nonsingular and because $O_p(F,H)$ has full rank np for any VARMA model and SFD, because it is lower-unit triangular. The appendix proves that $C_{np}(F,VH^T)$ has full rank np under conditions I-IV, so that $C_{L-p+1}(F,VH^T)$ has full rank np, for sufficiently large L. Thus, for sufficiently large L, $O_{L-p+1}(\tilde{F}^T,\tilde{H})$, $O_p(\tilde{F},\tilde{H})$, and D have full rank np and equation (3.2) can be solved for unique AR parameter values as

(3.6) $\quad X = (D^TD)^{-1}D^TE = np\times n.$

By virtue of the structure of D and E and the Cayley-Hamilton theorem, solution (3.6) satisfies equation (3.2) exactly, because, once D achieves full rank for sufficiently large L, the columns of E are in the space spanned by the columns of D.

The key step in the original XYW method for MFD is deleting Yule-Walker equations with missing high-frequency autocovariances of low-frequency variables. Anderson et al. (2012) pointed out that the deletions can be implemented by deleting from H rows mapping into unobserved variables (in their notation, deleting columns of G to obtain K). Describing such deletions for general MFD would be difficult and is not attempted here. However, this is practically unnecessary because most MFD cases can be handled as in the simplest MFD case in which some variables are observed at the high frequency every period and remaining variables are observed at the low frequency every other period. Both Chen and Zadrozny (1998) and Anderson et al. (2012) used this simplest case to analyze, respectively, XYW estimation and identification of bivariate VAR models. By studying generalizations of equation (13) in Chen and Zadrozny (1998), one can see that the two-part partition $H = [H_1^T,\ H_2^T]$ covers most MFD cases, except unusual ones in which some intermediate AR and MA coefficient matrices are restricted to zero.

Similarly describing three or more observation frequencies doesn't change this structure and only complicates notation.

Therefore, consider a VARMA model of $n = n_1 + n_2$ variables, whose first n_1 variables are high-frequency variables observed in every period and whose last n_2 variables are low-frequency variables observed every certain number of periods, so that $H = [H_1^T,\ H_2^T]^T$, where $H_1 = [e_1^T,\ \ldots,\ e_{n_1}^T]^T = n_1 \times np$, $H_2 = [e_{n_1+1}^T,\ \ldots,\ e_n^T]^T = n_2 \times np$, and, for $i = 1, \ldots, n$, $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ denotes the $np \times 1$ vector with one in position i and zeros elsewhere. Then, the deletion of unusable Yule-Walker equations with missing high-frequency autocovariances of low-frequency variables can be implemented simply by replacing H everywhere with H_1. Thus, in most circumstances, adapting solution equation (3.6) from SFD to MFD amounts to replacing D with $D_1 = O_{L-p+1}(\tilde{F}^T, \tilde{H}_1)O_p(\tilde{F}, \tilde{H}_1)^T$ and replacing E with $E_1 = O_{L-p+1}(\tilde{F}^T, \tilde{H}_1)(\tilde{F}^T)^p\tilde{H}_1^T$, where $\tilde{H}_1 = H_1R$, so that equation (3.6) becomes

$$(3.7) \qquad X = (D_1^T D_1)^{-1} D_1^T E_1 = np \times n.$$

The adaptation works iff $\mathrm{rank}[D_1] = np$, which requires two things. First, the reduction of sample information from removing Yule-Walker equations must be compensated for by increasing L, although this by itself is generally insufficient to maintain $\mathrm{rank}[D_1] = np$, because for MFD observability generally also depends on the AR parameters, as illustrated in section 5.3. The other identifying conditions, I-IV and VI, are unaffected by the move from SFD to MFD.

This section has proved that the AR parameters of VARMA model (2.1) are identified with SFD or typical MFD under conditions I-V of stationarity, regularity, miniphaseness, controllability, and observability.

4. Identification of MA parameters with forward Yule-Walker equations.

The effective disturbance covariance matrix of the model is $B_0B_0^T$, parameterized in the elements of lower-triangular B_0. Treating the disturbance covariance matrix as a part of the MA part of a model, even if the model is a pure VAR model with $q = 0$ and putting the first $(k = 0)$ block in both backward equations (3.1) and forward equations (4.1) simplifies derivations.

To obtain the forward Yule-Walker equations (FYWE) for SFD, postmultiply VARMA model (2.1) by "forward in time" y_{t+k}^T, for k = 0, ..., L ≥ p-1, take unconditional expectations, and obtain

$$(4.1)\qquad C_k^T - \sum_{i=1}^{p} A_i C_{i+k}^T = \sum_{i=0}^{p-1} B_i \Psi_{i+k}^T = n\times n,$$

which, using $\Psi_i = HF^iG$, can be written as

$$(4.2)\qquad [B_0, \ldots, B_{p-1}]C_p(F,G)^T O_{L+1}(F,H)^T = \overline{\Gamma}_L = n\times n(L+1),$$

where $\overline{\Gamma}_L = [\Gamma_0, \ldots, \Gamma_L]$ and $\Gamma_k = C_k^T - \sum_{i=1}^{p} A_i C_{i+k}^T$. Because $O_{L+1}(F,H)$ has full column rank for L ≥ p-1 and SFD, equation (4.2) can be rewritten as

$$(4.3)\qquad \sum_{i=0}^{p-1} B_i[B_0^T, \ldots, B_{p-1}^T](F^T)^i = \overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1} = n\times np.$$

The first np BYWE with MA terms in equation (3.1) could be used together with FYWE (4.3) to determine the MA parameters but are not. Not using the first np BYWE equations for identification makes no difference, because relevant full-rank conditions based on population covariances hold regardless whether the additional equations are used. However, using the additional equations for estimation when population covariances are replaced by sample covariances should result in more accurate estimates because more sample information would be used.

Assume temporarily that F is diagonalizable as $F^T = Z\Lambda Z^{-1}$. Because F has the block-companion form (2.3), its left (row) eigenvectors have the block-Vandermonde form

$$(4.4)\qquad z_i = (\lambda_i^{p-1}\xi_i^T, \ldots, \xi_i^T)^T = np\times 1,$$

where, for i = 1, ..., np, λ_i is an eigenvalue of F. Then, the np×np matrix Z of right (column) eigenvectors of F^T has the block-Vandermonde form

$$(4.5)\qquad Z = \begin{bmatrix} Z_1\Lambda_1^{p-1} & \cdots & Z_p\Lambda_p^{p-1} \\ \vdots & & \vdots \\ Z_1 & \cdots & Z_p \end{bmatrix} = np\times np,$$

where, for $\ell = 1, \ldots, p$, $Z_\ell = [\xi_{(\ell-1)n+1}, \ldots, \xi_{(\ell-1)n+n}] = n\times n$, $\Lambda_\ell = \mathrm{diag}(\lambda_{(\ell-1)n+1}, \ldots, \lambda_{(\ell-1)n+n}) = n\times n$, $\Lambda = \mathrm{diag}(\Lambda_1, \ldots, \Lambda_p) = np\times np$, and, for $i = 1, \ldots, np$, λ_i is a latent root of $A(\lambda)$ and ξ_i is a matching nonzero left latent vector of $A(\lambda)$. See Dennis et al. (1976) and Zadrozny (1998).

Let M denote the right side of equation (4.3). Use $F^T = Z\Lambda Z^{-1}$, multiply out $Z\Lambda^i$ and MZ at the level of detail of equation (4.5), and, for $\ell = 1, \ldots, p$, write equation (4.3) as

$$\text{(4.6)} \qquad \sum_{i=0}^{p-1}\sum_{j=0}^{p-1} B_iB_j^TZ_\ell\Lambda_\ell^{p-1+i-j} = N_\ell = n\times n,$$

where $N_\ell = \sum_{k=1}^{p} M_kZ_\ell\Lambda_\ell^{p-k}$, $M = [M_1, \ldots, M_p] = n\times np$, and M_k denotes the k-th n×n block of M.

Also assume temporarily that F is nonsingular, so that Λ is nonsingular. For $\ell = 1, \ldots, p$, postmultiply equation (4.6) by Λ_ℓ^{-p+1}, apply the vectorization rule $\mathrm{vec}(ABC) = [C^T \otimes A]\mathrm{vec}(B)$, where $\mathrm{vec}(\cdot)$ denotes the left-to-right column vectorization of a matrix (Magnus and Neudecker, 1999, p. 30), and write the resulting equation as

$$\text{(4.7)} \qquad (Z_\ell^T \otimes I_n)\mathrm{vec}(\sum_{j=0}^{p-1} B_jB_j^T) + \sum_{i=1}^{p-1} [(\Lambda_\ell^iZ_\ell^T \otimes I_n) + (\Lambda_\ell^{-i}Z_\ell^T \otimes I_n)P]\mathrm{vec}(\sum_{j=0}^{p-1-i} B_{i+j}B_j^T)$$

$$= \left(\Lambda_\ell^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_\ell\right),$$

for $\ell = 1, \ldots, p$, where, for any n×n matrix X, P denotes the $n^2\times n^2$ permutation matrix defined by $\mathrm{vec}(X^T) = P\mathrm{vec}(X)$.

Write equation (4.7) more concisely as $Ax = \beta$, where

$$\text{(4.8)} \qquad A = \begin{bmatrix} \Lambda_1^{p-1}Z_1^T \otimes I_n & \cdots & Z_1^T \otimes I_n \\ \vdots & & \vdots \\ \Lambda_p^{p-1}Z_p^T \otimes I_n & \cdots & Z_p^T \otimes I_n \end{bmatrix} + \begin{bmatrix} (\Lambda_1^{-p+1}Z_1^T \otimes I_n)P & \cdots & (Z_1^T \otimes I_n)P \\ \vdots & & \vdots \\ (\Lambda_p^{-p+1}Z_p^T \otimes I_n)P & \cdots & (Z_p^T \otimes I_n)P \end{bmatrix} = n^2p\times n^2p,$$

$$x = (x_{p-1}^T, \ldots, x_0^T)^T = n^2p\times 1,$$

$$x_i = \mathrm{vec}(\sum_{j=0}^{p-1-i} B_{i+j}B_j^T) = n^2\times 1 \;\; (i = p-1, \ldots, 1),$$

$$x_0 = \mathrm{vec}(\textstyle\sum_{j=0}^{p-1} B_j B_j^T)/2,$$

$$\beta = (\beta_1^T, \ldots, \beta_p^T)^T = n^2p\times 1,$$

$$\beta_\ell = \left(\Lambda_\ell^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_\ell\right) = n^2\times 1 \quad (\ell = 1, \ldots, p),$$

and P can postmultiply $Z_\ell^T \otimes I_n$, for $\ell = 1, \ldots, p$, in the last block column of the second part of A because $\sum_{j=0}^{p-1} B_j B_j^T$ is symmetric.

To simplify $Ax = \beta$ in order to verify that it can be solved for a unique value of x, first, write A as

$$(4.9) \qquad A = \begin{bmatrix} \Lambda_1^{p-1}Z_1^T & \cdots & Z_1^T \\ \vdots & & \vdots \\ \Lambda_p^{p-1}Z_p^T & \cdots & Z_p^T \end{bmatrix} \otimes I_n + \left[\begin{bmatrix} Z_1^T & \cdots & \Lambda_1^{-p+1}Z_1^T \\ \vdots & & \vdots \\ Z_p^T & \cdots & \Lambda_p^{-p+1}Z_p^T \end{bmatrix} Q \otimes I_n\right](I_p \otimes P),$$

where Q denotes the np×np permutation matrix that permutes blocks of n columns of $\Lambda^{-p+1}Z^T$ and P is the same permutation matrix as in equations (4.7)-(4.8). Use equation (4.5), premultiply equation (4.9) by $(Z^{-T}\Lambda^{p-1} \otimes I_n)$, and obtain

$$(4.10) \qquad (Z^{-T}\Lambda^{p-1} \otimes I_n)A = (F^{p-1} \otimes I_n) + S,$$

where $S = \left(Q \otimes I_n\right)\left(I_p \otimes P\right)$ is an $n^2p\times n^2p$ permutation matrix. Similarly,

$$(4.11) \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \left(\Lambda_1^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_1\right) \\ \vdots \\ \left(\Lambda_p^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_p\right) \end{bmatrix} = \begin{bmatrix} \left(\Lambda_1^{-p+1} \otimes I_n\right)\sum_{k=1}^{p}\left(\Lambda_1^{p-k}Z_1^T \otimes I_n\right)\mathrm{vec}\left(M_k\right) \\ \vdots \\ \left(\Lambda_p^{-p+1} \otimes I_n\right)\sum_{k=1}^{p}\left(\Lambda_p^{p-k}Z_p^T \otimes I_n\right)\mathrm{vec}\left(M_k\right) \end{bmatrix}$$

$$= \left(\Lambda^{-p+1} \otimes I_n\right)\begin{bmatrix} \Lambda_1^{p-1}Z_1^T \otimes I_n & \cdots & Z_1^T \otimes I_n \\ \vdots & & \vdots \\ \Lambda_p^{p-1}Z_p^T \otimes I_n & \cdots & Z_p^T \otimes I_n \end{bmatrix}\begin{bmatrix} \mathrm{vec}(M_1) \\ \vdots \\ \mathrm{vec}(M_p) \end{bmatrix} = \left(\Lambda^{-p+1}Z^T \otimes I_n\right)\mathrm{vec}(M).$$

Premultiply equation (4.11) by $(Z^{-T}\Lambda^{p-1} \otimes I_n)$, compare the result with equation (4.10), and obtain equation (4.7) as

(4.12) $[(F^{p-1} \otimes I_n) + S]x = \mathrm{vec}(M)$.

Because equation (4.12) is valid whether F is diagonalizable and nonsingular or not, having derived the equation, we no longer need these assumptions and, therefore, can and do withdraw them. Thus, in the end, these assumptions are unnecessary and serve only to reveal the derivation of equation (4.12) from equation (4.3). Figuring out how to do this without allowing F to be diagonalizable and nonsingular would be difficult.

There are two cases to consider in solving equation (4.12) for x: $p = 1$ and $p \geq 2$.

If $p = \max(r,q+1) = 1$, then, $r = 1$, $q = 0$, equation (4.12) is unnecessary and equation (4.3) reduces to

(4.13) $B_0 B_0^T = \overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1}$.

Because B_0 is nonsingular, $B_0 B_0^T$ is positive definite and B_0 can be determined uniquely from $B_0 B_0^T$ by Cholesky factorization.

If $p \geq 2$, then, $(F^{p-1} \otimes I_n) + S$ must be nonsingular in order to solve equation (4.12) uniquely for x. Because all eigenvalues of F have moduli less than one (because the model is stationary) and all eigenvalues of S have moduli equal to one (because S is a permutation matrix that maps vectors on the unit hypersphere back to the unit hypersphere), theorem 5.1.1 of Lancaster and Rodman (1995, p. 98) implies that $(F^{p-1} \otimes I_n) + S$ has nonzero eigenvalues and is nonsingular. Thus, we can solve equation (4.12) for a unique value of x in terms of previously determined AR parameters, as

(4.14) $x = [(F^{p-1} \otimes I_n) + S]^{-1}\mathrm{vec}\{\overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1}\}$.

We now describe the final steps for determining the MA parameters from x. We already have the n×n MA characteristic polynomial

(4.15) $B(\lambda) = B_0\lambda^q + B_1\lambda^{q-1} + \ldots + B_{q-1}\lambda + B_q$

and now also define the n×n characteristic polynomial

(4.16) $X(\lambda) = X_q\lambda^q + \ldots + X_1\lambda + 2X_0 + X_1^T\lambda^{-1} + \ldots + X_q^T\lambda^{-q}$,

where, for $i = 0, \ldots, q$, upper-case X_i are unique $n \times n$ unvectorizations of the $n^2 \times 1$ lower-case x_i defined by equations (4.8) and λ is a complex-valued scalar. If $p-1 \geq q+1$, then, $X_i = 0_{nxn}$, for $i = q+1, \ldots, p-1$, equation (4.12) could be reduced by deleting the first p-q-1 columns of $(F^{p-1} \otimes I_n) + S$ and the first p-q-1 elements of x, and solving equation (4.12) in the manner of equations (3.6)-(3.7).

Multiplying out $B(\lambda^{-1})B(\lambda)^T$ and comparing the resulting coefficients of λ with those of $X(\lambda)$ verifies that the factorization

(4.17) $X(\lambda) = B(\lambda^{-1})B(\lambda)^T$

exists. The factorization exists because $X(\lambda)$ has been derived based on covariances of variables which are assumed to be generated by VARMA model (2.1). If $X(\lambda)$ is divided by 2π and λ is restriced to $e^{-i\omega}$, where $i = \sqrt{-1}$ and $-2\pi < \omega \leq 2\pi$, then, $X(\lambda)$ becomes the spectral density of the MA part of VARMA model (2.1).

Zadrozny (1998) described an eigenvalue method of undetermined coefficients for solving a linear rational expectations model. The first step of doing this is computing the factorization $C(\lambda) = K(\lambda)\Phi(\lambda)$, such that $\Phi(\lambda)$ contains the smallest np roots of $C(\lambda)$, usually the stationary roots inside the unit circle. Because $B(\lambda)^T$ in $X(\lambda) = B(\lambda^{-1})B(\lambda)^T$ corresponds to $\Phi(\lambda)$ in $C(\lambda) = K(\lambda)\Phi(\lambda)$, it can be also be computed using the eigenvalue method of undetermined coefficients. In fact, the method applies without modification to computing $B(\lambda)$, because ensuring that $B(\lambda)$ is miniphase is the same as ensuring that $\Phi(\lambda)$ contains the np smallest roots of $C(\lambda)$. Here, skew symmetry of $X(\lambda)$ implies that $X(\lambda)$ has 2nq roots in nq reciprocal pairs. If a pair of roots is off the unit circle, then, the root inside the unit circle is chosen for the MA solution. If a pair of roots is on the unit circle, then, additional assumptions must be introduced to decide which root and associated latent vector (for repeated roots) should be chosen for the MA solution, akin to introducing additional assumptions for resolving global unidentification, as discussed in section 1.

Let $\overline{\overline{B}}^T U = U\Omega$ denote the right (column) eigenvalue decomposition of $\overline{\overline{B}}^T$, where Ω is an $nq \times nq$ diagonal matrix of eigenvalues and U is an $nq \times nq$ matrix of right eigenvectors. Because $\overline{\overline{B}}$ has the block-companion form of F in equation (2.3), the columns of U have the block-Vandermonde form of the left (row) eigenvectors of F in equation (4.4). Then, following Zadrozny (1998, pp. 1358-1359), the upper $n \times nq$ part of $\overline{\overline{B}}^T U = U\Omega$ is

$$(4.18) \qquad [\overline{B}_1^T, \ldots, \overline{B}_q^T]U = [U_1\boldsymbol{\Omega}_1^q, \ldots, U_q\boldsymbol{\Omega}_q^q], \quad U = \begin{bmatrix} U_1\boldsymbol{\Omega}_1^{q-1} & \cdots & U_q\boldsymbol{\Omega}_q^{q-1} \\ \vdots & & \vdots \\ U_1 & \cdots & U_q \end{bmatrix},$$

where, for i = 1, ..., q, U_i and Ω_i are defined analogously to Z_i and Λ_i in equation (4.5).

Equation (4.18) can be solved for $\overline{B}_i$, for i = 1, ..., q, because diagonalizability condition VI means that U is nonsingular. Because U and Ω are intermediate, not given, values, it might seem that the $\overline{B}_i$ could be nonunique. We now prove that the $\overline{B}_i$ are unique. First, Ω is unique because its nq diagonal elements are chosen from the 2nq eigenvalues of $\overline{\overline{B}}$ by a determinate rule, such as that the chosen eigenvalues have minimal moduli. Second, if the eigenvalues in Ω are distinct, then, U is unique (Wilkinson, 1965, p. 5). Third, if some eigenvalues in Ω are repeated, then, right eigenvectors of $\overline{\overline{B}}^T$ in U of repeated eigenvalues are nonunique. Let $\tilde{U}$ denote another matrix of right eigenvectors of $\overline{\overline{B}}^T$. For given U and $\tilde{U}$, there is an $nq \times nq$ nonsingular matrix M such that $\tilde{U} = UM$, because U and $\tilde{U}$ are nonsingular by diagonalizability. Because $\overline{\overline{B}}^T = \tilde{U}\Omega\tilde{U}^{-1} = UM\Omega M^{-1}U^{-1} = U\Omega U^{-1}$, because $\overline{\overline{B}}^T$ has the same eigenvalues for any eigenvalue decomposition, it follows that $\overline{B}_i$, for i = 1, ..., q, that satisfy equation (4.18) are unique, whether or not eigenvalues in Ω are distinct.

It remains to determine unique values of B_i, for i = 0, ..., q. Factorization (4.17) can be restated as $X(\lambda) = \overline{B}(\lambda^{-1})B_0B_0^T\overline{B}(\lambda)^T$, where $\overline{B}(\lambda) = I_n\lambda^q + \overline{B}_1\lambda^{q-1} + \ldots + \overline{B}_{q-1}\lambda + \overline{B}_q$. Suppose that λ_0 is not a root of $\overline{B}(\lambda)$, so that $|\overline{B}(\lambda_0)| \neq 0$. Then, $B_0B_0^T$ can be determined as

$$(4.19) \qquad B_0B_0^T = \overline{B}(\lambda_0^{-1})^{-1}X(\lambda_0)\overline{B}(\lambda_0)^{-T},$$

where superscript -T denotes inversion and transposition. Because B_0 is nonsingular, a unique value of B_0 can be determined from positive definite

$B_0B_0^T$ by Cholesky factorization, whereupon unique values of $B_i = -\overline{B}_i B_0$, for i = 1, ..., q, are determined.

Therefore, unique values of B_i, for i = 0, ..., q, are determined for given x from equation (4.13) or from equations (4.14)-(4.19).

We adapt the above solution for the MA parameters from SFD to MFD in essentially the same way as we adapted the solution for the AR parameters from SFD to MFD at the end of section 3. Consider the same partition $H = [H_1^T, H_2^T]$ as in section 3. Similarly, replace H with H_1 in $O_{L+1}(F,H)^T$ on the left side of equation (4.2), correspondingly reduce columns of $\overline{\Gamma}_L$ on the right side of the equation, and proceed as in the SFD case, from equations (4.13)-(4.14) to equation (4.19). The adaptation to MFD works iff $\text{rank}[O_{L+1}(F,H_1)] = np$, which, as in the AR case, generally also requires increasing L and imposing additional restrictions on the AR parameters.

This section has proved that the MA parameters of VARMA model (2.1) are identified with SFD or MFD under conditions I-VI of stationarity, regularity, miniphaseness, controllability, observability, and diagonalizability, conditional on the AR parameters having been identified.

5. Concluding discussion.

The paper concludes with discussions of common left AR and MA factors, necessity of the identifying conditions, numerical illustration of identification, and extension to identifcation of structural parameters.

5.1. Common left AR and MA factors.

All parameters of VARMA model (2.1) have been proved to be identified when the model satisfies conditions I-VI and its variables are observed with either single-frequency data (SFD) according to the first definition in section 1 or with mixed-frequency data (MFD). Because the AR parameters were proved to be identified independently of the MA parameters, all ARMA parameters were proved to be identified. Has redundancy between the AR and MA parameters been precluded in the sense that the AR and MA characteristic equations have no common left factors? Hannan (1969) emphasized the absence of this condition as a condition for identifying AR and MA parameters. However, because controllability condition (2.6) precludes common AR and MA pairs of latent roots and left latent vectors, it precludes common left AR and MA factors.

5.2. Necessity of the identifying conditions.

The paper proved that identifying conditions I-VI are sufficient to identify the parameters of a VARMA model, but did not prove that the conditions are as a whole necessary for identification. Each condition is necessary or appears to be necessary in some part of the proof. Stationarity (I) is necessary for otherwise the identification problem is not well posed. Regularity (II) appears to be necessary for identifying the MA parameters, although Zadrozny (1998) defined an analogue of $\overline{\overline{B}}$ that doesn't require regularity, which suggests that it may be unnecessary. Miniphaseness (III) is necessary for otherwise some MA parameters with roots on the unit circle are ruled out and cannot be identified. Controllability (IV) appears to be necessary for separately identifying the AR and MA parameters. Observability (V) appears to be necessary for identifying the AR parameters. Diagonalizability (VI) appears to be necessary for identifying the MA parameters. The present proof follows a particular method for determining the parameters from data covariances. However, a general proof of whether conditions I-VI as a whole are necessary for identifying the parameters must be independent of any particular method for determining them.

5.3. Numerical illustration of identification.

Sections 3-4 proved that parameters of VARMA model (2.1) are identified for SFD if $C_L(F,VH^T)$ and $O_p(F,H)$ have full rank and are identified for MFD if $C_L(F,VH_1^T)$ and $O_p(F,H_1)$ have full rank, in both cases for sufficiently large L. As an illustration of identification for MFD, consider the estimated bivariate ARMA(1,1) model of monthly employment and quarterly GNP in Zadrozny (1990a,b). The model has estimated coefficient matrices $\hat{A}_1 = \begin{bmatrix} .799 & .417 \\ .203 & .353 \end{bmatrix}$, $\hat{B}_0 = \begin{bmatrix} 2.37 & 0.00 \\ .634 & 1.34 \end{bmatrix}$, $\hat{B}_1 = \begin{bmatrix} -.615 & -.697 \\ 1.72 & -.613 \end{bmatrix}$, the model is stationary (I), miniphase (III), controllable (IV), and diagonalizable (VI), because AR roots of .942 and .209 and MA roots of $.289 \pm .643\sqrt{-1}$ are less than one in modulus and are distinct, and the model is regular (II) because $\hat{B}_0$ is nonsingular. For L = p

= 2 and H_1 = (1,0,0,0), $C_L(F,VH_1^T)^T = \begin{bmatrix} v_{11} & v_{21} & v_{31} & v_{41} \\ a_{11}v_{11}+a_{12}v_{21}+v_{31} & a_{21}v_{11}+a_{22}v_{21}+v_{41} & 0 & 0 \end{bmatrix}$ and $O_p(F,H_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{11} & a_{12} & 1 & 0 \end{bmatrix}$, where a_{ij} and v_{ij} denote (i,j) elements of A_1 and V. $C_L(F,VH_1^T)$ has full rank 2 for the parameter values and the structure of $O_p(F,H_1)$ implies that it has full rank 2 for all values of a_{11} and a_{12}, so that the model is identified for the MFD.

Observability is the only identifying condition that depends on how variables of a model are observed. Section 3 proved that observability holds for SFD and all parameter values. For MFD, observability may or may not depend on the parameters. In the above model, observability holds for the MFD and all parameter values. However, if we drop the MA part and the bottom half of the state vector in order to maintain controllability, then, $C_L(F,VH_1^T)^T = \begin{bmatrix} v_{11} & v_{21} \\ a_{11}v_{11}+a_{12}v_{21} & a_{21}v_{11}+a_{22}v_{21} \end{bmatrix}$ and $O_p(F,H_1) = \begin{bmatrix} 1 & 0 \\ a_{11} & a_{12} \end{bmatrix}$ continue to have full rank 2 for the parameter values, but now $O_p(F,H_1)$ has full rank 2 iff $a_{12} \neq 0$.

Adding parameters usually complicates estimation and identification, but independence of observability from a_{12} simplifies identification and could result in AR parameters being estimated more accurately. This would be unusual because MA parameters are usually the more difficult parameters to estimate accurately and their presence often reduces precision of estimation of AR parameters. The condition for $C_L(F,VH_1^T)$ to have full rank is difficult to interpret, but observability condition $a_{12} \neq 0$ means simply that quarterly GNP, the low-frequency variable, feeds back at monthly intervals on monthly employment, the high-frequency variable.

5.4. Identification of structural parameters.

If a VARMA model has no underlying structure, then, its parameters are equivalently structural and reduced-form. Consider now structural VARMA models with underlying structural parameters in vector θ that are mapped to reduced-form parameters in vector φ by some differentiable function $f(\theta)$. Usually, an estimation method produces an estimate $\hat{\theta}$ of θ that minimizes a twice-differentiable composite function $g(f(\theta))$ (in reduced-form estimation, $\varphi \equiv \theta$). Proposed structural and reduced-form methods for estimating VARMA models with MFD are maximum likelihood (Zadrozny, 1988, 1990a,b), extended

Yule-Walker (Chen and Zadrozny, 1998), Bayesian (Eraker et al., 2015), MIDAS (Ghysels et al., 2007), and stacking (Ghysels, 2012). After putting more details into $g(\cdot)$, in theory θ is identified if the smallest nonnegative eigenvalue of $J_0^T K_0 J_0$ is positive, where J_0 and K_0 are Jacobian and Hessian matrices of $f(\theta)$ and $g(\varphi)$ evaluated at true values θ_0 and $\varphi_0 = f(\theta_0)$; and, in practice, $\hat{\theta}$ is estimated accurately or is "strongly identified" if the smallest nonnegative eigenvalue of $\hat{J}^T \hat{K} \hat{J}$ is sufficiently positive, where $\hat{J}$ and $\hat{K}$ are evaluated at estimated values $\hat{\theta}$ and $\hat{\varphi} = f(\hat{\theta})$. Dynamic stochastic general equilibrium (DSGE) models (Smets and Wouters, 2003) are now commonly used structural models in macroeconomic analysis. Recent research on identification of structural VARMA models has been motivated by this use. For example, Komunjer and Ng (2011) and Kociecki and Kolasa (2013) studied some, but not all, necessary conditions for identification of DSGE models with VARMA reduced forms. The present paper contributes to this literature by effectively proving that true or estimated K is positive definite if, respectively, true or estimated ARMA parameters satisfy conditions I-VI.

6. Appendix.

The appendix proves that, under conditions I-IV of stationarity, regularity, miniphaseness, and controllability, $C_{np}(F,VH^T)$ has full rank np, which contributes to the proof in the text below equation (3.5) that matrix D in equations (3.5) and (3.6) has full rank np.

For i = 1, ..., np, consider

$$(6.1)\qquad z_i^T C_{np}(F,G) = (\lambda_i^{p-1}\xi_i^T, \ \ldots, \ \xi_i^T)^T \left[\begin{bmatrix} B_0 \\ \vdots \\ B_{p-1} \end{bmatrix}, F\begin{bmatrix} B_0 \\ \vdots \\ B_{p-1} \end{bmatrix}, \cdots, F^{np-1}\begin{bmatrix} B_0 \\ \vdots \\ B_{p-1} \end{bmatrix} \right]$$

$$= \xi_i^T \left[\textstyle\sum_{j=0}^{p-1} B_j \lambda_i^{p-1-j}, \ \sum_{j=0}^{p-1} B_j \lambda_i^{p-j}, \ \ldots, \ \sum_{j=0}^{p-1} B_j \lambda_i^{np+p-2-j} \right],$$

where z_i is a left eigenvector of F and λ_i is its matching eigenvalue.

There are two cases: $r \geq q+1$ and $r \leq q$. If $r \geq q+1$, then, $p = \max(r,q+1) = r$ and, because $B_j = 0_{nxn}$ for $j \geq q+1$,

$$(6.2)\qquad z_i^T C_{np}(F,G) = (\lambda_i)^{\max(r-q-1,0)}[\xi_i^T B(\lambda_i),\ \xi_i^T B(\lambda_i)\lambda_i\ ,\ \ldots,\ \xi_i^T B(\lambda_i)\lambda_i^{np-1}],$$

where $B(\lambda_i) = \sum_{j=0}^{q} B_j\lambda_i^{q-j}$. If $r \le q$, then, $p = q+1$ and equation (6.2) continues to hold with the understanding that $(\lambda_i)^{\max(r-q-1,0)} = 1$ for any zero or nonzero λ_i. Then, equations (3.4) and (6.2) imply that

$$(6.3)\qquad z_i^T VH^T = (\lambda_i)^{\max(r-q-1,0)}[\xi_i^T B(\lambda_i),\ \ldots,\ \xi_i^T B(\lambda_i)\lambda_i^{np-1},\ \ldots]\begin{bmatrix} G^T \\ \vdots \\ G^T(F^T)^{np-1} \\ \vdots \end{bmatrix}H^T$$

$$= (\lambda_i)^{\max(r-q-1,0)}\xi_i^T B(\lambda_i)G^T\sum_{j=0}^{\infty}(\lambda_i F^T)^j H^T = (\lambda_i)^{\max(r-q-1,0)}\xi_i^T B(\lambda_i)G^T[I_{np}-\lambda_i F^T]^{-1}H^T,$$

where stationarity implies that $\sum_{j=0}^{\infty}(\lambda_i F^T)^j$ exists and equals $[I_{np}-\lambda_i F^T]^{-1}$, so that the last equality in equation (6.3) holds. Thus, because controllability implies that $(\lambda_i)^{\max(r-q-1,0)}\xi_i^T B(\lambda_i) \ne 0_{1xn}$, it follows that $z_i^T VH^T \ne 0_{1xn}$ if, but not necessarily only if, $\tilde{M} = H[I_{np} - \lambda_i F]^{-1}G$ is nonsingular.

To prove that $\tilde{M}$ is nonsingular, consider observation equation (2.2), state equation (2.3) modified hypothetically as $x_t = \lambda_i F x_{t-1} + G\varepsilon_t$, where F and G are unchanged from equation (2.3); for $i = 1, \ldots, np$, λ_i continues to denote an eigenvalue of F; and, the state vector is partitioned into $n\times 1$ subvectors as $x_t = (x_{1,t}^T,\ \ldots,\ x_{p,t}^T)^T$. The modified state equation may be written out as

$$(6.4)\qquad \begin{aligned} x_{1,t} &= \lambda_i A_1 x_{1,t-1} + \lambda_i x_{2,t-1} + B_0\varepsilon_t, \\ &\ \ \vdots \\ x_{p-1,t} &= \lambda_i A_{p-1} x_{1,t-1} + \lambda_i x_{p,t-1} + B_{p-2}\varepsilon_t, \\ x_{p,t} &= \lambda_i A_p x_{1,t-1} + B_{p-1}\varepsilon_t. \end{aligned}$$

Replace $x_{p,t-1}$ on the right side of the next-to-last equation in (6.4) for $x_{p-1,t}$ with the right side of the last equation in (6.4) for $x_{p,t}$ lagged one period; then, replace $x_{p-1,t-1}$ on the right side of the next-to-next-to-last equation for $x_{p-2,t}$ with the right side of the just obtained equation for $x_{p-1,t}$

lagged one period; continue like this; after using observation equation (2.2) to replace $x_{1,t}$ with y_t, obtain

$$(6.5) \qquad y_t = \lambda_i A_1 y_{t-1} + \ldots + \lambda_i^p A_p y_{t-p} + B_0 \varepsilon_t + \lambda_i B_1 \varepsilon_{t-1} + \ldots + \lambda_i^{p-1} B_{p-1} \varepsilon_{t-p+1}.$$

Consider equation (6.5) at the steady-state output $\bar{y}$ for any constant input $\bar{\varepsilon}$. There are two cases: $\lambda_i = 0$ and $\lambda_i \neq 0$. If $\lambda_i = 0$, then, the steady state of equation (6.5) is $\bar{y} = \tilde{N}\bar{\varepsilon}$, where $\tilde{N} = B_0$, so that regularity condition II implies that $\tilde{N}$ is nonsingular. If $\lambda_i \neq 0$, then, because $A_j = 0_{nxn}$ for $j \geq r+1$ and $B_k = 0_{nxn}$ for $k \geq q+1$, the steady state of equation (6.5) is $A(\lambda_i^{-1})\bar{y} = \lambda_i^{q-r} B(\lambda_i^{-1})\bar{\varepsilon}$, where $A(\lambda_i^{-1}) = I_n \lambda_i^{-r} - A_1 \lambda_i^{-r+1} - \ldots - A_r$ and $B(\lambda_i^{-1}) = B_0 \lambda_i^{-q} + B_1 \lambda_i^{-q+1} + \ldots + B_q$. Because stationarity condition I implies that $A(\lambda_i^{-1})$ is nonsingular, $\bar{y} = \tilde{N}\bar{\varepsilon}$, where $\tilde{N} = \lambda_i^{q-r} A(\lambda_i^{-1})^{-1} B(\lambda_i^{-1})$. Miniphase condition III implies that $B(\lambda_i^{-1})$ is nonsingular, so that $\tilde{N}$ is nonsingular. Because state-space representation (2.2) and (6.4) implies that $\bar{y} = \tilde{M}\bar{\varepsilon}$, where $\tilde{M} = \tilde{N}$, it follows that $\tilde{M}$ is nonsingular. Therefore, for any λ_i, $\tilde{M}$ is nonsingular and $C_{np}(F, VH^T)$ has full rank np, as was to be shown.

REFERENCES

Akaike, H. (1974), "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes," Annals of the Institute of Statistical Mathematics 26: 363-387.

Anderson, B.D.O., M. Deistler, E. Felsenstein, B. Funovits, P.A. Zadrozny, M. Eichler, W. Chen, and M. Zamani (2012), "Identifiability of Regular and Singular Multivariate Autoregressive Models from Mixed-Frequency Data," Proceedings of 51st IEEE Conference on Decision and Control, pp. 184-189.

Anderson, B.D.O. and J.B. Moore (1979), Optimal Filtering, Englewood Cliffs, NJ: Prentice-Hall.

Box, G.E.P. and G.M. Jenkins (1976), Time Series Analysis: Forecasting and Control, San Francisco, CA: Holden Day.

Chen, B. and P.A. Zadrozny (1998), "An Extended Yule-Walker Method for Estimating Vector Autoregressive Models with Mixed-Frequency Data," Advances in Econometrics 13: 47-73.

Dennis, J.E., Jr., J.F. Traub, and R.P. Weber (1976), "The Algebraic Theory of Matrix Polynomials," SIAM Journal on Numerical Analysis 13: 831-845.

Eraker, B., C.W. Chiu, A.T. Foerster, T.B. Kim, and H.D. Seoane (2015), "Bayesian Mixed-Frequency VARs", Journal of Financial Econometrics 13: 698-721.

Friedman, M. (1962), "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association 57: 729-757.

Ghysels, E. (2012), "Macroeconomics and the Reality of Mixed Frequency Data," working paper, Dept. of Economics, University of North Carolina, Chapel Hill, NC.

Ghysels, E., A. Sinko, and R. Valkanov (2007), "MIDAS Regressions: Further Results and New Directions," Econometric Reviews 26: 53-90.

Golub, G. and C. Van Loan (1996), Matrix Computations, 3rd ed., Baltimore, MD: Johns Hopkins University Press.

Hannan, E.J. (1969), "The Identification of Mixed Autoregressive-Moving Average Systems," Biometrika 56: 223-225.

Hannan, E.J. (1970), Time Series Analysis, New York, NY: John Wiley & Sons.

Hannan, E.J. (1976), "The Identification and Parameterization of ARMAX and State Space Forms," Econometrica 44: 713-723.

Hannan, E.J. (1979), "The Statistical Theory of Linear Systems," pp. 83-121 in Developments of Statistics, P.R. Krishnaiah (ed.), New York, NY: Academic Press.

Hannan, E.J. and M. Deistler (1986), Statistical Theory of Linear Systems, John Wiley & Sons: New York, NY (reprinted in 2012 by SIAM, Philadelphia, PA).

Hansen, L.P. and T.J. Sargent (1983), "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," Econometrica 51: 377-387.

Hautus, M.L.J. (1969), "Controllability and Observability Conditions of Linear Autonomous Systems," Indagationes Mathematicae 12: 443-448.

Kailath, T. (1980), Linear Systems, Englewood Cliffs, NJ: Prentice-Hall.

Kociecki, A. and M. Kolasa (2013), "Global Identification of Linearized DSGE Models," working paper, National Bank of Poland, Warsaw, Poland.

Komunjer, I. and S. Ng (2011), "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," Econometrica 79: 1995-2032.

Kwakernaak, H. and R. Sivan (1972), Linear Optimal Control Systems, New York, NY: Wiley-Interscience.

Lancaster, P. and L. Rodman (1995), Algebraic Riccati Equations, Oxford, UK: Clarendon Press.

Magnus, J. and H. Neudecker (1999), Matrix Differential Calculus with Applications in Statistics and Econometrics, revised edition, Chichester, UK: John Wiley & Sons.

Miller, P.J. and D.M. Chin (1996), "Using Monthly Data to Improve Quarterly Model Forecasts," Federal Reserve Bank of Minneapolis Quarterly Review 20 (2): 16-33.

Phillips, P.C.B. (1973), "The Problem of Identification in Finite-Parameter Continuous-Time Models," Journal of Econometrics 1: 351-362.

Priestley, M.B. (1981), Spectral Analysis and Time Series, Vol. 2, New York, NY: Academic Press.

Smets, F. and R. Wouters (2003), "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," Journal of the European Economic Association 1: 1123-1175.

Wilkinson, J.H. (1965), The Algebraic Eigenvalue Problem, Oxford, UK: Clarendon Press.

Zadrozny, P.A. (1988), "Gaussian Likelihood of Continuous-Time ARMAX Models when Data are Stocks and Flows at Different Frequencies," Econometric Theory 4: 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," working paper, Research Dept., Federal Reserve Bank of Atlanta, Atlanta, GA and Center for Economic Studies, U.S. Bureau of Census, Washington, DC.

Zadrozny, P.A. (1990b), "Forecasting U.S. GNP at Monthly Intervals with an Estimated Bivariate Time Series Model," Federal Reserve Bank of Atlanta Economic Review 75: 2-15 (1990).

Zadrozny, P.A. (1998), "An Eigenvalue Method of Undetermined Coefficients for Solving Linear Rational Expectations Models," Journal of Economic Dynamics and Control 22: 1353-1373.