EXTENDED YULE-WALKER IDENTIFICATION OF VARMA MODELS WITH SINGLE- OR MIXED-FREQUENCY DATA.

Peter A. Zadrozny

Bureau of Labor Statistics
2 Massachusetts Ave., NE, Room 3105
Washington, DC 201212
e-mail: zadrozny.peter@bls.gov

December 17, 2014

JEL Classification: C32, C80

Key words: linear and nonlinear estimation, state-space representation, spectral factorization

ABSTRACT

Chen and Zadrozny (1998) developed the linear extended Yule-Walker (XYW) method for determining the parameters of a vector autoregressive (VAR) model using available covariances of mixed-frequency data (MFD). XYW takes data covariances as inputs and determines AR parameters as outputs. If the covariance inputs are true population values and the outputs are unique, then, the outputs are true parameter values, i.e., the model is identified; if the covariance inputs are consistent sample estimates and the outputs are unique, then, the outputs are consistent parameter estimates. The present paper extends XYW to "extended XYW" (X^2YW) for determining all ARMA parameters of a VARMA model using available covariances of single-frequency data (SFD) or MFD. The paper also proves, under stated conditions on parameters, that the outputs are unique, so that the VARMA model is identified for true population-covariance inputs and is consistently estimated for consistent sample-covariance inputs. X^2YW solves one linear equation to determine the AR parameters and solves two linear equations and does one spectral factorization to determine the MA parameters.*

*The paper represents the author's views and does not necessarily represent any official positions of the Bureau of Labor Statistics. Affiliated as Research Fellow with the Center for Financial Studies (CFS), Johann Wolfgang Goethe University, Frankfurt, Germany, and with the Center for Economic Studies and Ifo Institute for Economic Research (CESifo), Munich, Germany. Thanks to Manfred Deistler, Roderick McCrorie, and Tucker McElroy for motivating and clarifying conversations.

1. Introduction.

Estimation of vector autoregressive moving-average (VARMA) models using mixed-frequency data (MFD) was considered first using nonlinear maximum likelihood estimation (MLE), which is effective only if good starting values can be found for the parameters to be estimated and this is difficult to do unless there are relatively few variables and parameters (Zadrozny 1988, 1990a,b). 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 ordinary least squares, and illustrated XYW's accuracy relative to MLE. The inference problem with MFD is that 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 takes as inputs data covariances presumed to be generated by a VAR model and determines its parameters as outputs. If the inputs are true population covariances and the outputs are unique, then, the outputs are the true model parameters, i.e., the VAR model is identified; if the inputs are consistent sample estimates and the outputs are unique, then, the outputs are consistent parameter estimates. Chen and Zadrozny (1998) did not prove either that XYW is feasible or that it determines unique VAR parameter outputs for true-population or consistent-sample covariance inputs. Anderson et al. (2012) did this for a general VAR model and a particular MFD case, for a generic set of parameters.

The paper makes two contributions. First, the paper extends steps 1 and 2 of XYW (XYW1-2) to steps 1 and 2 of "extended XYW" (X^2YW1-2), which determine all ARMA parameters of a VARMA model based on available covariances of single-frequency data (SFD) or MFD. Second, the paper proves, under stated and verifiable conditions on parameters (without the qualification "generically") that the parameter outputs are unique, so that the model is (locally) identified for true-population-covariance inputs and is consistently estimated for consistent-sample-covariance inputs. X^2YW1 solves one linear equation to determine the AR parameters and X^2YW2 solves two linear equations and does one spectral factorization to determine the MA parameters. The nonlinear spectral factorization does not vitiate the goal of avoiding iteration to convergence from starting values of nonlinear MLE, because spectral factorization can be done reliably, accurately, and quickly using a noniterative eigenvalue method (Zadrozny, 1998).

Identification can be local or global. The paper considers only local identification. By definition, parameters which generate identical population or sample covariances 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 has a single point, then, the model is globally identified. A particular form of global nonidentification of VARMA models called aliasing can occur with MFD. See Phillips (1973) and Hansen and Sargent (1983) for background discussions. In Anderson et al. (2012), parameter a_{ss} exhibits global aliasing nonidentification with MFD when its sign is undetermined.

The paper continues as follows. Section 2 states the VARMA model in original and state-space form, states assumptions (I)-(V) of stationarity, regularity, miniphase, controllability, and observability on the model, derives backward Yule-Walker equations (BYWE) of X^2YW1 for determining the AR parameters, and proves that under assumptions (I)-(V) the BYWE determine unique values of the AR parameters. Section 3 derives forward Yule-Walker equations (FYWE) of X^2YW2 for determining the MA parameters and proves that under assumptions (I)-(V) and an additional diagonalizability assumption (VI) the FYWE determine unique values of the MA parameters. The BYWE and FYWE are first derived for SFD in sections 2 and 3 and are, then, adapted for MFD in section 4. Section 5 concludes with a summary and brief discussions of necessity versus sufficiency of assumptions (I)-(VI), of possible extensions, and of identification of structural models.

2. BYWE solution of unique AR parameters for SFD.

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

(2.1) $$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + B_0 \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + + B_{p-1} \varepsilon_{t-p+1},$$

with components defined and assumptions (I)-(III) made on them as follows: y_t denotes an n×1 vector of observed outputs; 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, $A_r \neq 0_{n\times n}$, and intermediate (i = 1, ..., r-1) and trailing (i = r+1, ..., p) A_i, respectively, may be and are zero; *we assume (I) that VARMA model (2.1) is* *stationary, 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 absolute value (modulus);* B_j (j = 0, ..., p-1) denote n×n matrices of MA parameters, $B_q \neq 0_{n\times n}$, and intermediate (j = 1, ..., q-1) and trailing (j = q+1, ..., p-1) B_j, respectively, may be and are zero; *we assume (II) that VARMA model (2.1) is regular, i.e., B_0 is nonsingular and lower triangular; we assume (III) that VARMA model (2.1) is miniphase, i.e., if λ is a real- or complex-valued scalar root of the MA characteristic equation $|B(\lambda)| = |B_0\lambda^q + B_1\lambda^{q-1} + \ldots + B_q| = 0$, then, $|\lambda| \leq 1$;* ε_t denotes an n×1 vector of unobserved disturbances ~ $IID(0_{n\times 1}, I_n)$, where I_n denotes the n×n identity matrix; all quantities above and below are implicitly real valued, unless explicitly stated to be complex valued.

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

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

where $0_{n\times n}$ denotes the n×n zero matrix and x_t denotes the np×1 state vector, with state equation

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

Let $C_k = Ey_t y_{t-k}^T$, for k = 0, ±1, ±2, ..., denote the k-th true population covariance of y_t and y_{t-k} on the assumption that y_t is generated by the stationary VARMA model, where E denotes unconditional expectation and superscript T denotes vector or matrix transposition. C_k exists if the model is stationary and is skew symmetric, i.e., $C_k = C_{-k}^T$.

The BYWE and FYWE are first derived for SFD, respectively, in sections 2 and 3 and are, then, adapted for MFD in section 4. The BYWE for SFD are usually called YWE, but are here called BYWE to distinguished them from the related but different FYWE.

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

$$(2.4)\qquad \begin{bmatrix} C_0 \\ \vdots \\ C_{p-1}^T \\ C_p^T \\ \vdots \\ C_L^T \end{bmatrix} = \begin{bmatrix} C_p & \cdots & C_1 \\ \vdots & & \vdots \\ C_1 & \cdots & C_{p-2}^T \\ C_0 & \cdots & C_{p-1}^T \\ \vdots & & \vdots \\ C_{L-p}^T & \cdots & C_{L-1}^T \end{bmatrix} \begin{bmatrix} A_p^T \\ \vdots \\ A_1^T \end{bmatrix} + \begin{bmatrix} \sum_{i=0}^{p-1} \Psi_i B_i^T \\ \vdots \\ \Psi_0 B_{p-1}^T \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{bmatrix},$$

where $\Psi_i = HF^iG$ denotes the i-th coefficient matrix of the Wold MA(∞) representation of the model.

We want to solve BYWE (2.4) for unique values of the AR parameters, A_1, ..., A_p. To do this, we skip the first p blocks (k = 1, ..., p) with MA terms and consider only further blocks (k = p, ..., L) without MA terms,

$$(2.5)\qquad \begin{bmatrix} C_0 & \cdots & C_{p-1}^T \\ \vdots & & \vdots \\ C_{L-p}^T & \cdots & C_{L-1}^T \end{bmatrix} \begin{bmatrix} A_p^T \\ \vdots \\ A_1^T \end{bmatrix} = \begin{bmatrix} C_p^T \\ \vdots \\ C_L^T \end{bmatrix}.$$

To proceed, we need the assumptions of controllability and observability.

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

$$(2.6)\qquad C_K(F,G) = [G, \ldots, F^{K-1}G] = np\times nK.$$

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) rank$[C_{np}(F,G)]$ = np.

A VARMA model is said to be controllable iff the controllability matrix has full rank. For general state-space matrices F and G, Hautus (1969) proved that rank$[C_{np}(F,G)]$ = np iff, for any complex-valued scalar λ,

$$(2.7)\qquad \text{rank}[F - I_{np}\lambda,\ G] = np.$$

Kailath (1980, p. 135) calls condition (2.7) the "PBH test," although Lancaster and Rodman (1995, p. 88) state that the condition was first proved by Hautus (1969).

Because F has the block-companion form (2.3) and a left (row) eigenvector of F has the block-Vandermonde form (3.4), condition (2.7) holds iff, for i = 1, ..., np,

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

where λ_i is an eigenvalue of F, ξ_i is a nonzero left latent vector of $A(\lambda) = I_n\lambda^r - A_1\lambda^{r-1} - \ldots - A_r$, which satisfies $\xi_i^T A(\lambda_i) = 0_{1xn}$, and $B(\lambda) = B_0\lambda^q + \ldots + B_q$. The discussion in the appendix leading to equation (6.4) proves that $\mathrm{rank}[C_{np}(F,G)] = np$ and condition (2.8) are equivalent. If $r \leq q$, then, n(q-r+1) zero eigenvalues of F are not AR roots that satisfy $|A(\lambda)| = 0$, $(\lambda_i)^{\max(r-q-1,0)} = 1$ in condition (2.8), and AR roots being distinct from MA roots is sufficient but generally unnecessary for condition (2.8) to hold; if $r \geq q+1$, then, all eigenvalues of F are AR roots and must be nonzero in order for controllability and condition (2.8) to hold.

Thus, VARMA model (2.1) is controllable iff $\mathrm{rank}[C_{np}(F,G)] = np$ and conditions (2.7) and (2.8) hold. *We assume (IV) that VARMA model (2.1) is controllable*.

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 (3.4), where ξ_i is a left latent vector of $A(\lambda_i)$. See Dennis et al. (1976).

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

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

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 $\mathrm{rank}[O_{np}(F,H)] = np$. A VARMA model is said to be observable iff the observability matrix has full rank, hence, iff $\mathrm{rank}[F^T - I_{np}\lambda,\ H^T] = np$. Because F is asymmetric, it generally has different left and right eigenvectors for each eigenvalue, so there is no direct analogue of condition (2.8) for observability, obtained by replacing F and G with F^T and H^T in equation (2.7).

Controllability and observability come from dynamic system theory (Kwakernaak and Sivan, 1972; Anderson and Moore, 1979; Kailath, 1980). Controllability generally depends on all ARMA parameter values, regardless how the model's output (y_t) is observed. Observability generally depends only on AR parameter values and on how the model's output is 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, although it is generally necessary to assume that the model is observable for MFD, because in the latter case observability generally depends on the AR parameters.

Different lower bounds have been stated for L. In each case, the lower bound is a necessary but not necessarily sufficient condition for an observability condition to hold. However, because L has no upper limit in identification, we shall henceforth often more simply just state that "L is sufficiently large". Of course, in estimation, L is limited by sample size. *Thus, we assume (V) that VARMA model (2.1) is observable for a sufficiently large L, for the MFD being considered.*

Consider equation (2.5) as DX = E. The equation solves uniquely for AR parameters in X iff, for sufficiently large L, D has full (column) rank. It has been proved, for sufficiently large L and SFD, that D has full rank if assumptions (I)-(V) of stationarity, regularity, miniphase, controllability, and observability hold. See Akaike (1974), Hannan (1969, 1970, 1976), and Hannan and Deistler (1986). Nevertheless, to facilitate adapting the SFD results in sections 2 and 3 to MFD in section 4, we now give another proof.

State-space representation (2.2)-(2.3) implies that, for k = 0, 1, ...,

$$(2.10) \qquad C_k = HF^kVH^T = n\times n,$$

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,

$$(2.11) \qquad V = [C_{np}(F,G),\ \dots\ ][C_{np}(F,G),\ \dots\ ]^T = np\times np.$$

V is symmetric positive semidefinite by its structure. The Cayley-Hamilton theorem implies 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×np, lower triangular, nonsingular, and unique. Using $\tilde{F} = R^{-1}FR$, $\tilde{H} = HR$, and equation (2.10), system matrix D of equation (2.5) can be expressed as

$$(2.12)\quad D = \begin{bmatrix} HVH^T & \cdots & HV(F^T)^{p-1}H^T \\ \vdots & & \vdots \\ HV(F^T)^{L-p}H^T & \cdots & HV(F^T)^{L-1}H^T \end{bmatrix} = \begin{bmatrix} \tilde{H}\tilde{H}^T & \cdots & \tilde{H}(\tilde{F}^T)^{p-1}\tilde{H}^T \\ \vdots & & \vdots \\ \tilde{H}(\tilde{F}^T)^{L-p}\tilde{H}^T & \cdots & \tilde{H}(\tilde{F}^T)^{L-1}\tilde{H}^T \end{bmatrix}$$

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

D has full rank np, for sufficiently large L, iff $O_{L-p+1}(\tilde{F}^T,\tilde{H}) = R^{-1}C_{L-p+1}(F,VH^T)^T$ and $O_p(\tilde{F},\tilde{H}) = O_p(F,H)R$ do. Because R^{-1} 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. The appendix proves that $C_{np}(F,VH^T)$ has full rank np under assumptions (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 (2.5) solves uniquely for the AR parameters as

$$(2.13)\quad X = (D^TD)^{-1}D^TE = np\times n.$$

To summarize this section: for SFD, if VARMA model (2.1) satisfies assumptions (I)-(V) of stationarity, regularity, miniphase, controllability, and observability, then, equation (2.13) gives the unique X^2YW1 solution of the AR parameters, A_1, ..., A_p, which are true and identified parameter values for true SFD population covariances and are consistently estimated parameter values for consistent sample covariances. In the latter case, because sample covariances generated by a stationary VARMA model converge in probability to population covariances and because the mapping from covariances to AR parameters is differentiable and, hence, continuous, the continuous mapping theorem implies that the AR parameters are consistently estimated.

3. FYWE solution of unique MA parameters for SFD.

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

$$(3.1)\qquad C^T_k - \sum_{i=1}^{p} A_i C^T_{i+k} = \sum_{i=0}^{p-1} B_i \Psi^T_{i+k} = n\times n,$$

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

$$(3.2)\qquad [B_0, \ldots, B_{p-1}]C_p(F,G)^T O_{L+1}(F,H)^T = \overline{\Gamma}_L = n\times n(L+1),$$

where $\overline{\Gamma}_L = [\Gamma_0, \ldots, \Gamma_L]$ and $\Gamma_k = C^T_k - \sum_{i=1}^{p} A_i C^T_{i+k}$. Because $O_{L+1}(F,H)$ has full column rank for L ≥ p-1 and SFD, equation (3.2) can be written as

$$(3.3)\qquad \sum_{i=0}^{p-1} B_i[B^T_0, \ldots, B^T_{p-1}](F^T)^i = \overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1} = n\times np.$$

The first np BYWE with MA terms could be used together with FYWE (3.3) 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 involving population covariances hold whether or not the additional equations are used. However, using the additional equations for estimation with sample covariances would probably result in more accurate estimates on average 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

$$(3.4)\qquad z_i = (\lambda_i^{p-1}\xi^T_i, \ldots, \xi^T_i)^T = np\times 1,$$

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

$$(3.5)\qquad Z = \begin{bmatrix} Z_1\Lambda_1^{p-1} & \cdots & Z_p\Lambda_p^{p-1} \\ \vdots & & \vdots \\ Z_1 & \cdots & Z_p \end{bmatrix} = np\times np,$$

where, for $\ell = 1, \ldots, p$, $Z_\ell = [\xi_{(\ell-1)n+1}, \ldots, \xi_{(\ell-1)n+n}] = n\times n$, $\Lambda_\ell = \text{diag}(\lambda_{(\ell-1)n+1}, \ldots, \lambda_{(\ell-1)n+n}) = n\times n$, $\Lambda = \text{diag}(\Lambda_1, \ldots, \Lambda_p) = np\times np$, 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).

Use $F^T = Z\Lambda Z^{-1}$, multiply out $Z\Lambda^i$ and MZ at the level of detail of equation (3.5), and, for $\ell = 1, \ldots, p$, write equation (3.3) as

$$(3.6)\qquad \sum_{i=0}^{p-1}\sum_{j=0}^{p-1} B_i B_j^T Z_\ell \Lambda_\ell^{p-1+i-j} = N_\ell = n\times n,$$

where $N_\ell = \sum_{k=1}^{p} M_k Z_\ell \Lambda_\ell^{p-k}$, $M = [M_1, \ldots, M_p] = n\times np$ denotes the right side of equation (3.3), and M_k denotes the k-th $n\times n$ block of M.

Also assume temporarily that F is nonsingular, so that Λ is nonsingular. For $\ell = 1, \ldots, p$, postmultiply equation (3.6) by Λ_ℓ^{-p+1}, apply the vectorization rule $\text{vec}(ABC) = [C^T \otimes A]\text{vec}(B)$, where $\text{vec}(\cdot)$ denotes the left-to-right column vectorization of a matrix (Magnus and Neudecker, 1999, p. 30), and write the resulting equation as

$$(3.7)\qquad (Z_\ell^T \otimes I_n)\text{vec}(\textstyle\sum_{j=0}^{p-1} B_j B_j^T) + \sum_{i=1}^{p-1}[(\Lambda_\ell^{i} Z_\ell^T \otimes I_n) + (\Lambda_\ell^{-i} Z_\ell^T \otimes I_n)P]\text{vec}(\sum_{j=0}^{p-1-i} B_{i+j} B_j^T)$$

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

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

Write equation (3.7) more concisely as $Ax = \beta$, where

$$(3.8)\qquad A = \begin{bmatrix} \Lambda_1^{p-1} Z_1^T \otimes I_n & \cdots & Z_1^T \otimes I_n \\ \vdots & & \vdots \\ \Lambda_p^{p-1} Z_p^T \otimes I_n & \cdots & Z_p^T \otimes I_n \end{bmatrix} + \begin{bmatrix} (\Lambda_1^{-p+1} Z_1^T \otimes I_n)P & \cdots & (Z_1^T \otimes I_n)P \\ \vdots & & \vdots \\ (\Lambda_p^{-p+1} Z_p^T \otimes I_n)P & \cdots & (Z_p^T \otimes I_n)P \end{bmatrix} = n^2p\times n^2p,$$

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

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

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

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

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

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 $\sum_{j=0}^{p-1} B_jB_j^T$ is symmetric.

To simplify $Ax = \beta$ in order to verify that it solves uniquely for x, first,

$$\text{(3.9)} \qquad A = \begin{bmatrix} \Lambda_1^{p-1}Z_1^T & \cdots & Z_1^T \\ \vdots & & \vdots \\ \Lambda_p^{p-1}Z_p^T & \cdots & Z_p^T \end{bmatrix} \otimes I_n + \left[\begin{bmatrix} Z_1^T & \cdots & \Lambda_1^{-p+1}Z_1^T \\ \vdots & & \vdots \\ Z_p^T & \cdots & \Lambda_p^{-p+1}Z_p^T \end{bmatrix} Q \otimes I_n\right](I_p \otimes P),$$

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 equation (3.7). Use equation (3.5), premultiply equation (3.9) by $(Z^{-T}\Lambda^{p-1} \otimes I_n)$, and obtain

$$\text{(3.10)} \qquad (Z^{-T}\Lambda^{p-1} \otimes I_n)A = (F^{p-1} \otimes I_n) + S,$$

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

$$\text{(3.11)} \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \left(\Lambda_1^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_1\right) \\ \vdots \\ \left(\Lambda_p^{-p+1} \otimes I_n\right)\mathrm{vec}\left(N_p\right) \end{bmatrix} = \begin{bmatrix} \left(\Lambda_1^{-p+1} \otimes I_n\right)\sum_{k=1}^{p}\left(\Lambda_1^{p-k}Z_1^T \otimes I_n\right)\mathrm{vec}\left(M_k\right) \\ \vdots \\ \left(\Lambda_p^{-p+1} \otimes I_n\right)\sum_{k=1}^{p}\left(\Lambda_p^{p-k}Z_p^T \otimes I_n\right)\mathrm{vec}\left(M_k\right) \end{bmatrix}$$

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

Premultiply equation (3.11) by $(Z^{-T}\Lambda^{p-1} \otimes I_n)$, compare the result with equation (3.10), and obtain equation (3.7) as

$$(3.12) \qquad [(F^{p-1} \otimes I_n) + S]x = vec(M).$$

Because equation (3.12) 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 Vandermonde property of F and the permutation property of S.

There are two cases to consider in solving equation (3.12) for x: p = 1 and p ≥ 2.

If p = max(r,q+1) = 1, then, r = 1, q = 0, equation (3.12) is unnecessary, equation (3.3) directly gives

$$(3.13) \qquad B_0B_0^T = \overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1}H^T,$$

and B_0 can be determined uniquely from $B_0B_0^T$ by Cholesky factorization.

If p ≥ 2, then, $(F^{p-1} \otimes I_n) + S$ must be nonsingular in order to solve equation (3.12) uniquely for x. Because all eigenvalues of F have moduli less than one, because the model is stationary, and because all eigenvalues of S have moduli equal to one, because S is a permutation matrix which 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, hence, is nonsingular. Thus, we can solve equation (3.12) uniquely for x in terms of known covariances and previously determined AR parameters as

$$(3.14) \qquad x = [(F^{p-1} \otimes I_n) + S]^{-1}vec\{\overline{\Gamma}_L O_{L+1}(F,H)[O_{L+1}(F,H)^T O_{L+1}(F,H)]^{-1}\}.$$

Having derived linear steps (3.3) and (3.13)-(3.14) for determining the MA parameters, we now describe the final nonlinear spectral factorization step for determining the MA parameters. At this point, because we are dealing only with the MA part of the model and no longer need state-space representation (2.3)-(2.3), if p-1 ≥ q+1, we simplify expressions by imposing $B_i = 0_{nxn}$, for i ≥ q+1. Thus, we define the n×n characteristic polynomials

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

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

where, for i = 0, ..., q, upper-case X_i are unique n×n unvectorizations of the $n^2\times1$ lower-case x_i defined by equations (3.8) and λ is a complex-valued scalar. If p-1 ≥ q+1, then, $X_i = 0_{nxn}$, for i = q+1, ..., p-1, equation (3.12) 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 in the manner of equation (2.13).

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

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

exists. The factorization exists because X(λ) has been derived from covariances which are assumed to be generated by VARMA model (2.1). Hannan (1970, pp. 61-67) proved that X(λ) has a unique miniphase right factor B(λ) under assumptions (I)-(V) and a more restrictive version of assumption (VI) of diagonalizability (stated below equation (3.18)) that the MA roots are distinct. See also Hannan and Deistler (1988, pp. 20-30). If X(λ) is divided by 2π and λ is restriced to $e^{-i\omega}$, where $i = \sqrt{-1}$ and $-2\pi < \omega \le 2\pi$, then, X(λ) is the spectral density of the MA part of VARMA model (2.1).

With the next discussion of simultaneous identification in mind, we also consider diagonalizability assumption (VI). Define the block-companion form matrix

(3.18) $$\overline{B} = \begin{bmatrix} B_0^{-1}B_1 & I_n & \cdots & 0_{n\times n} \\ \vdots & 0_{n\times n} & \ddots & \vdots \\ \vdots & \vdots & \ddots & I \\ B_0^{-1}B_q & 0_{n\times n} & \cdots & 0_{n\times n} \end{bmatrix} = nq\times nq.$$

We assume (VI) that $\overline{B}$ *is diagonalizable*. It is conjectured (Zadrozny, 2014) that X(λ) has a unique miniphase right factor B(λ) iff $\overline{B}$ is diagonalizable. A unique miniphase right factor B(λ) can be computed using an eigenvalue method of undetermined coefficients, which was developed for a more general and analogous $\overline{B}$ that does not depend on B_0 being nonsingular (Zadrozny, 1998).

The parameters of VARMA model (2.1) have so far been proved to be identified sequentially under assumptions (I)-(VI): first the AR parameters were proved to be identified; then, conditional on AR parameters being identified, the MA parameters were proved to be identified. However, does this sequential identification preclude redundancy between AR and MA parameters and guarantee their simultaneous identification? We now prove that the AR and MA parameters are also simultaneously identified.

Let Ξ and Υ denote nq×nq matrices of eigenvalues and eigenvectors of $\overline{B}$. Diagonalizable $\overline{B}$ has eigenvalue decomposition $\overline{B}^T = \Upsilon\Xi\Upsilon^{-1}$. Because $\overline{B}$ has the same block-companion form as F, Υ has the same block-Vandermonde form as Z in equation (3.5). For i = 1, ..., q, let Υ_i and Ξ_i denote analogues of Z_i and Λ_i. It is conjectured (Zadrozny, 2014) that if $\overline{B}$ is diagonalizable, then, Υ_i are nonsingular for any ordering of Υ_i and Ξ_i and any orderings of their columns, and conversely. Thus, for i = 1, ..., q, solvents $W_i(\lambda) = I_n\lambda - \Upsilon_i\Xi_i\Upsilon_i^{-1}$ of $B(\lambda)$ exist. See Dennis et al. (1976) for a discussion of solvents of lambda matrices. It is conjectured (Zadrozny, 2014) that if $\overline{B}$ is diagonalizable, then, $B(\lambda)$ factors as,

(3.19) $$B(\lambda) = W_1(\lambda) \cdots W_q(\lambda),$$

where the $W_i(\lambda)$ and their columns can be in any order.

For some j = 1, ..., q, suppose that $U_j(\lambda) = I_n\lambda - Z_j\Lambda_jZ_j^{-1}$ is a left solvent of $A(\lambda) = I_n\lambda^r - A_1\lambda^{r-1} - \ldots - A_r$, so that

(3.20) $$A(\lambda) = U_j(\lambda)\tilde{A}_j(\lambda),$$

where $\tilde{A}_j(\lambda)$ is an n×n lambda matrix of order r-1 in λ. Let λ_k be a root of $U_j(\lambda)$ and, hence, of $A(\lambda)$ and let ξ_k be a matching left latent vector which satisfies $\xi_k^TU_j(\lambda_k) = 0_{1xn}$ and, hence, $\xi_k^TA(\lambda_k) = 0_{1xn}$. If $U_j(\lambda) \equiv W_1(\lambda)$, then, $\xi_k^TB(\lambda_k) = 0_{1xn}$ and controllability is contradicted. Thus, under controllability, AR and MA parameters cannot be redundant with respect to each other and, hence, are simultaneously identified. Generally speaking, AR and MA parameters are simultaneously identified iff $A(\lambda)$ and $B(\lambda)$ have no common left factors and, in such case, $A(\lambda)$ and $B(\lambda)$ are said to be left coprime (Hannan, 1969, 1976).

In the above analysis, $W_1(\lambda)$ exists because $\overline{B}$ is assumed to be diagonalizable, but $U_j(\lambda)$ may not exist because no assumptions have been made to ensure this result. Thus, the simultaneous identification condition that "$U_j(\lambda)$ and $W_1(\lambda)$ do not both exist such that $U_j(\lambda) \equiv W_1(\lambda)$" may hold simply because $A(\lambda)$ has no left factor $U_j(\lambda)$.

To summarize this section: for SFD, if VARMA model (2.1) satisfies assumptions (I)-(VI) of stationarity, regularity, miniphase, controllability, observability, and diagonalizability, then, the miniphase right factor $B(\lambda)$ of equation (3.17), based on x given by equation (3.14), is the unique X^2YW2 solution of the MA parameters, B_0, ..., B_q, which are true and identified parameter values for true SFD population covariances and are consistently estimated parameter values for consistent sample covariances. As in X^2YW1 in section 2, sample covariances generated by a stationary model converge in probability to population covariances and, because the AR parameters are consistently estimated and the mapping from covariances and AR parameters to MA parameters is differentiable and, hence, continuous, the continuous mapping theorem implies that the MA parameters are also consistently estimated.

4. BYWE and FYWE solution of unique ARMA parameters for MFD.

The key step in XYW for MFD is deleting YW equations with missing covariances due to low-frequency variables being observed at less than the highest frequency. Anderson et al. (2012) pointed out that only high-frequency autocovariances of low-frequency variables are missing for MFD. Accordingly, they simplified the XYW row-deletion step to deleting rows of the observation matrix, H, which map state variables to output variables observed at low frequencies (in their notation, deleting columns of K). Describing this simplification 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 highest frequency every period and remaining variables are observed at the lower frequency every other period. Chen and Zadrozny (1998) and Anderson et al. (2012) both used this simplest case to analyze XYW estimation of bivariate models. By studying generalizations of equation (13) in Chen and Zadrozny (1998), one can see that the following two-part partition of H covers most MFD cases, except in unusual cases in which some intermediate AR and MA coefficient matrices are restricted to zero.

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 other period. Then, H can be written as $H = [H_1^T, H_2^T]^T$, where $H_1 = [e_1^T, ..., e_{n_1}^T]^T = n_1 \times np$, $H_2 = [e_{n_1+1}^T, ..., e_n^T]^T = n_2 \times np$, and, for $i = 1, ..., n$, $e_i = (0, ..., 0, 1, 0, ..., 0)^T$ denotes the $np \times 1$ vector with one in position i and zeros elsewhere.

To adapt the X^2YW1 solution of section 2 for the AR parameters from SFD to MFD, simply reduce H to H_1 in system matrix D of equation (2.5) as written out in equation (2.12) and use the resulting reduced D in the AR solution equation (2.13). To adapt the X^2YW2 solution of section 3 for MA parameters from SFD to MFD, similarly, reduce H to H_1 in $O_{L+1}(F,H)^T$ on the left side of equation (3.2), correspondingly reduce columns of $\overline{\Gamma}_L$ on the right side of the equation, and proceed as in the SFD case to factor equation (3.17) for the MA parameters. The adaptation works iff $\text{rank}[O_{L+1}(F,H_1)] = np$. This requires two things. First, the reduction of sample information from removing YW equations must be compensated for by increasing L, although this by itself is generally insufficient to maintain $\text{rank}[O_{L+1}(F,H_1)] = np$, because for MFD observability generally also depends on the AR parameters. Because controllability of a model does not depend on how its outputs are observed, it is unaffected by moving from SFD to MFD.

Consider verifying identification conditions (I)-(VI) for the bivariate ARMA(1,1) model estimated by MLE and monthly-quarterly data by Zadrozny (1990a,b) and Chen and Zadrozny (1998). The verification proves that the set of models and MFD for which the identification conditions hold is nonempty. In the model, the first and high-frequency variable is monthly U.S. aggregate employment and the second and low-frequency variable is quarterly U.S. real GNP, both in standardized percentage-growth form. In the present notation, the estimated AR and MA coefficient matrices are

$$(4.1) \qquad \hat{A}_1 = \begin{bmatrix} .799 & .417 \\ .203 & .353 \end{bmatrix}, \quad \hat{B}_0 = \begin{bmatrix} 2.37 & 0.00 \\ .634 & 1.34 \end{bmatrix}, \quad \hat{B}_1 = \begin{bmatrix} -.615 & -.697 \\ 1.72 & -.613 \end{bmatrix},$$

which indicate that the AR roots are .942 and .209 and that the MA roots are $.289 \pm .643\sqrt{-1}$ (modulus = .705). The model is stationary (I), because the AR roots have moduli less than one; the model is regular (II), because B_0 is nonsingular; the model is miniphase (III), because the MA roots have moduli less than one; the model is controllable (IV), because the AR roots are distinct from the MA roots; the model is observable (V), because element (1,2) of $\hat{A}_1$ is nonzero ($\hat{a}_{12}$ = .417); and, $\bar{B}$ is diagonalizable (VI), because the MA roots are distinct. Thus, the parameters of model (4.1) are simultaneously identified. By converging to a locally unique maximum, the MLE of the model also numerically verified the simultaneous identification.

The MIDAS method for estimating a distributed-lag model with MFD by regressing low-frequency variables on high-frequency variables, developed by Ghysels et al. (2004, 2006), has subsequently been applied widely. For MFD, MIDAS has advantages and disadvantages compared to estimation of a VARMA model. For example, MIDAS is less likely to be numerically uresolved due to multicollinearity, which can prevent MLE of a VARMA model from converging, and MIDAS executes much faster. On the other hand, an estimated VARMA model includes lagged feedbacks from both high- and low-frequency variables. In model (4.1), the first and second columns of $\hat{A}_1$ and $\hat{B}_1$ give the AR and MA lagged feedbacks from, respectively, the high- and the low-frequency variables. By design, MIDAS estimates only lagged feedbacks from the high-frequency variables. Thus, for example, MIDAS cannot check observability condition $\hat{a}_{12} \neq 0$ of model (4.1).

5. Conclusion.

The paper has contributed two main results. First, steps 1 and 2 of X^2YW (X^2YW1-2) have been derived. Second, the paper has proved that if VARMA model (2.1) satisfies assumptions (I)-(VI) of stationarity, regularity, miniphase, controllability, observability, and diagonalizability, then, for SFD or MFD, its parameters are simultaneously identified by X^2YW1-2 based on population covariances and are consistently estimated by X^2YW1-2 based on consistent sample covariances.

The paper has proved that assumptions (I)-(VI) are sufficient for the stated results, but did not attempt to prove that the assumptions are necessary for the results. Nevertheless, the discussion in the paper suggests that some of assumptions might be relaxed to be closer to necessity or to be necessary. For example, the regularity assumption that B_0 is nonsingular is used in section 3 to define $\bar{B}$ and in the appendix to prove that $\tilde{M}$ is nonsingular. An analogous $\bar{B}$ can be defined that does not depend on regularity and the nonsingularity of $\tilde{M}$ is sufficient but unnecessary for the proof in the appendix, which suggests that regularity could be relaxed. However, we leave a fuller discussion of this issue for the future.

Several extensions of X^2YW come to mind. Straightforward extensions would be: (1) investigate further which commonly considered VARMA(r,q) models are identified for commonly occuring MFD; (2) explain how to handle linear restrictions on parameters, in particular, zero restrictions; and, (3) use real and simulated data to investigate X^2YW1-2 estimation accuracy relative to MLE. More difficult extensions would be: (4) extend X^2YW1-2 estimation to a statistically more efficient generalized method of moments (GMM) step 3 or X^2YW3; and, (5) determine the asymptotic distributions of X^2YW1-2 and, if available, of X^2YW1-3 for one iteration, multiple iterations, or converged iterations over the steps (if the iterations indeed converge under some conditions) and relate these distributions to MLE's asymptotic distribution. Although the identification proof in the paper indicates a computational method for consistently estimating VARMA parameters with MFD, simply by replacing population covariances with consistent sample covariances, the X^2YW method is not necessarily recommended as a substitute for MLE until more testing with real and simulated data shows X^2YW to be competitive in accuracy with MLE.

Structural linear dynamic stochastic general equililbrium (LDSGE) models or linear approximations of DSGE models (Smets and Wouters, 2003) are now widely used in macroeconomic analysis. LDSGE models often have VARMA reduced forms, hence, have state-space reduced forms. Komunjer and Ng (2011) and Kocięcki and Kolasa (2013) discussed identification of structural LDSGE models in terms of state-space reduced forms. Suppose that an LDSGE model has a structure that imposes restrictions on VARMA reduced-form parameters in vector φ in terms of structural parameters in vector θ by a differentiable mapping $\varphi = f(\theta)$. Let $g(\varphi)$ denote a twice-differentiable unrestricted likelihood function in terms of the reduced-form parameters and let $g(f(\theta))$ denote the restricted likelihood function in terms of the structural parameters. Suppose that MLE with SFD or MFD successfully finds the unique global maximum of $g(f(\theta))$ at $\hat{\theta}$. Let $\nabla^2 g(\varphi)$ denote the unrestricted Hessian matrix of second-partial derivatives of $g(\cdot)$ evaluated at φ and let $\nabla f(\theta)$ denote the Jacobian matrix of first-partial derivatives of $f(\cdot)$ evaluated at θ. Unique MLE implies that the structural parameters are identified, hence, that the restricted Hessian matrix $\nabla f(\hat{\theta})^T \nabla^2 g(f(\hat{\theta}))\, \nabla f(\hat{\theta})$ of $g(f(\theta))$ evaluated at $\hat{\theta}$ is negative definite. Thus, if a model is identified, then, $\nabla f(\hat{\theta})$ must have full column rank, although $\nabla^2 g(\hat{\varphi})$ could be (i) nonsingular (and negative definite) or (ii) singular (and negative semidefinite). In case (i), if an unrestricted VARMA model with identity mapping $f(\cdot)$ is identified at $\hat{\varphi}$ by the above VARMA-identification conditions (I)-(VI), then, unrestricted $\nabla^2 g(\hat{\varphi})$ is negative definite. In case (ii), a sufficiently restricted structural model could be identified, so that restricted $\nabla f(\hat{\theta})^T \nabla^2 g(f(\hat{\theta}))\, \nabla f(\hat{\theta})$ is negative definite, even if unrestricted $\nabla^2 g(\hat{\varphi})$ is singular. Case (ii) occurs, for example, when some variables in a structural model are never observed by the economic analyst, such as expectations of inflation in Zadrozny (1997) or production capital and technology in Chen and Zadrozny (2009, 2013). The present paper has in effect established conditions on VARMA model (2.1) under which unrestricted $\nabla^2 g(\hat{\varphi})$ is negative definite.

6. Appendix.

This appendix proves that, under assumptions (I)-(IV) of stationarity, regularity, miniphase, and controllability, $C_{np}(F, VH^T)$ has full rank np. This result contributes to the proof in the text below equation (2.12) that matrix D in equation (2.13) has full rank np.

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

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

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

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 $j \geq q+1$,

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

where $B(\lambda_i) = \sum_{j=0}^{q} B_j\lambda_i^{q-j}$. If $r \leq q$, then, $p = q+1$ and

$$(6.3)\qquad z_i^T C_{np}(F,G) = [\xi_i^T B(\lambda_i),\ \ldots,\ \xi_i^T B(\lambda_i)\lambda_i^{np-1}],$$

so that, for any r and q,

$$(6.4)\qquad z_i^T C_{np}(F,G) = (\lambda_i)^{\max(r-q-1,0)}[\xi_i^T B(\lambda_i),\ \ldots,\ \xi_i^T B(\lambda_i)\lambda_i^{np-1}].$$

Then, equations (2.11) and (6.4) imply that

$$(6.5)\qquad z_i^T VH^T = (\lambda_i)^{\max(r-q-1,0)}[\xi_i^T B(\lambda_i),\ \ldots,\ \xi_i^T B(\lambda_i)\lambda_i^{np-1},\ \ldots\ ]\begin{bmatrix} G^T \\ \vdots \\ G^T(F^T)^{np-1} \\ \vdots \end{bmatrix} H^T$$

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

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 (6.5) holds. Thus, because controllability implies that $(\lambda_i)^{\max(r-q-1,0)}\xi_i^T B(\lambda_i) \neq 0_{1xn}$, it follows that $z_i^T VH^T \neq 0_{1xn}$ if, but not only if, $\tilde{M} = H[I_{np} - \lambda_i F^T]^{-1}G$ is nonsingular.

It remains to prove that $\tilde{M}$ is nonsingular. Consider observation equation (2.2), state equation (2.3) modified as $x_t = \lambda_i F x_{t-1} + G\varepsilon_t$, where F and G are unchanged from equation (2.3), and the state vector is partitioned into n×1 subvectors as $x_t = (x_{1,t}^T,\ \ldots,\ x_{p,t}^T)^T$. Accordingly, the modified state equation is written out as

$$(6.6)\qquad \begin{aligned} x_{1,t} &= \lambda_i A_1 x_{1,t-1} + \lambda_i x_{2,t-1} + B_0\varepsilon_t, \\ &\vdots \\ x_{p-1,t} &= \lambda_i A_{p-1} x_{1,t-1} + \lambda_i x_{p,t-1} + B_{p-2}\varepsilon_t, \\ x_{p,t} &= \lambda_i A_p x_{1,t-1} + B_{p-1}\varepsilon_t. \end{aligned}$$

Replace $x_{p,t-1}$ on the right side of the next-to-last equation in (6.6) for $x_{p-1,t}$ with the right side of the last equation in (6.6) 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

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

Consider equation (6.7) 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 (6.7) is $\bar{y} = \tilde{N}\bar{\varepsilon} = B_0\bar{\varepsilon}$ and assumption (II) of regularity means that $\tilde{N} = B_0$ 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 (6.7) 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 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 implies that $B(\lambda_i^{-1})$ is nonsingular, so that $\tilde{N}$ is nonsingular. Because state-space representation (2.2) and (6.6) implies that $\bar{y} = \tilde{M}\bar{\varepsilon}$, $\tilde{M} = \tilde{N}$. Therefore, for any λ_i, $\tilde{M}$ is nonsingular and $C_{np}(F,VH^T)$ has full rank np, as was to be shown.

7. 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., M. Deistler, E. Felsenstein, B. Funovits, P. 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. and G 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.

Chen, B. and P.A. Zadrozny (2009), "Estimated U.S. Manufacturing Production Capital and Technology Based on an Estimated Dynamic Structural Economic Model," Journal of Economic Dynamics and Control 33: 1398-1418.

Chen, B. and P.A. Zadrozny (2013), "Further Model-Based Estimates of U.S. Total Manufacturing Production Capital and Technology, 1949-2005," Journal of Productivity Analysis 39: 61-73.

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.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models," working paper, CIRANO, Montreal, Canada.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006), "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies," Journal of Econometrics 131: 59-95.

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. 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.

Kocięcki, 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: J. Wiley & Sons.

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

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.

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. (1997) "An Econometric Analysis of Polish Inflation Dynamics with Learning About Rational Expectations," Economics of Planning 30: 221-238.

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.

Zadrozny, P.A. (2014), "Necessary and Sufficient Conditions for Existence of Unique VARX Solutions of EVARMAX Linear Rational Expectations Models," paper being prepared.