AN EXTENDED YULE-WALKER METHOD FOR ESTIMATING A VECTOR AUTOREGRESSIVE MODEL WITH MIXED-FREQUENCY DATA*

Baoline Chen
Department of Economics
Rutgers University-Camden
Camden, NJ 08102
e-mail: baoline@crab.rutgers.edu

Peter A. Zadrozny
Bureau of Labor Statistics
2 Massachusetts Ave., NE, Room 4915
Washington, DC 20212
e-mail: zadrozny_p@bls.gov

August 21, 1998

ABSTRACT

Zadrozny (1990) proposed and illustrated a nonlinear Kalman-filtering (KF) method for estimating a vector autoregressive moving-average (VARMA) model with mixed-frequency and partly temporally-aggregated data. The present paper proposes an optimal three-step linear instrumental variable method for estimating a VAR model with mixed-frequency data. The method compensates for missing data arising from mixed frequencies by using restrictions implied by extended Yule-Walker (XYW) equations, beyond the usual minimum YW equations for estimating a VAR model with fully-observed single-frequency data. The theory of generalized method of moments is used to derive an asymptotically efficient XYW estimator, determine its asymptotic distribution, and provide asymptotic tests of overidentifying restrictions. The KF method can simultaneously handle missing data, temporal aggregation, measurement errors, reporting delays, and revisions, but performs poorly or not at all on large models with many parameters. The XYW method can handle any pattern of missing data (subject to parameters being identified), but it is not yet clear how it might handle the other mentioned data problems. Having the computational complexity of generalized least squares, the XYW method can handle much larger models. The XYW method is illustrated and compared to the KF method using real and simulated macroeconomic monthly-quarterly data. The large number of available macroeconomic and financial time series at observation frequencies ranging from annual to "tick by tick" offer wide possibilities for applications of the XYW method.

1. INTRODUCTION

A longstanding problem in time-series econometrics has been estimating models with mixed-frequency data. When applied to mixed-frequency data, standard methods like ordinary least squares (OLS) are unable to estimate some high-frequency feedback effects. To understand this point, consider the following example. Two variables are observed, respectively, at monthly and quarterly intervals over a sample period. The variables are assumed to be generated jointly by a monthly process. Therefore, from a monthly perspective about two-thirds of potential observations on the quarterly variable are missing. The objective is to estimate a bivariate, monthly, third-order, vector autoregressive or VAR(3) model with the sample. Let $a_{k,ij}$ denote the coefficient in equation $i$ of variable $j$ at lag $k$. OLS cannot estimate $a_{k,ij}$ for $j=2$ and $k=1, 2$, i.e., OLS cannot estimate coefficients of the quarterly variable at intra-quarterly lags. In general, an estimation method designed for single-frequency data cannot estimate high-frequency feedbacks of low-frequency variables.

Zadrozny (1990) proposed and illustrated a nonlinear Kalman-filtering (KF) method, for estimating a vector autoregressive moving-average (VARMA) model with mixed-frequency data, which overcomes this problem. The method uses the Kalman filter to compute the Gaussian likelihood function in estimation and to help compute forecasts with the estimated model. The illustration concerned estimating a bivariate monthly model of U.S. aggregate employment and real GNP and developing and evaluating post-sample monthly forecasts of both variables.

The present paper proposes an optimal three-step linear instrumental variable method for estimating a VAR model with mixed-frequency data. The method compensates for missing data arising from mixed frequencies by using restrictions implied by extended Yule-Walker (XYW) equations, beyond the usual minimum YW equations for estimating a VAR model with fully-observed
single-frequency data. The standard theory of generalized method of moments (GMM) is used to derive an asymptotically efficient XYW estimator, determine its asymptotic distribution, and provide asymptotic tests of overidentifying YW equations.

Cadzow (1982) proposed and illustrated a closely related method which he termed overdetermined estimation. While we address the problem of estimating a correctly specified model with incomplete data, Cadzow addressed the problem of estimating a misspecified model with complete data. Nevertheless, the two methods can be merged to simultaneously address both problems. The common insight underlying the methods is that incomplete information, whether ignorance of the true data generating process or incomplete samples caused by mixed-frequency observations, can be partially replaced by "higher-order" covariance information beyond "minimal-order" covariance information used by standard estimation methods such as OLS. Cadzow was aware that overdetermining restrictions require weights in order to produce unique estimates, but was unaware of optimal weighting schemes (Hansen, 1982; Stoica and Soderstrom, 1983; Zou and Du, 1991) and used plausible, but generally asymptotically suboptimal, weights.

The KF method can simultaneously handle missing data, temporal aggregation, measurement errors, reporting delays, and revisions, but performs poorly or not at all on large models with many parameters. The XYW method can handle any pattern of missing data, subject to parameters being identified, but it is not yet clear how it might handle the other mentioned data problems. Having the computational complexity of generalized least squares, the XYW method can handle much larger models. The XYW method can be used to determine rank conditions for identifying VAR parameters with any pattern of mixed-frequency data. The recursive Kalman filter obscures the algebra of the model and conditions under which it is identified.
Section 2 uses standard GMM theory to derive an asymptotically efficient, three-step, XYW method for estimating VAR coefficients with a mixed-frequency sample: step one computes consistent estimates of autoregressive coefficients; step two uses the coefficient estimates to compute consistent estimates of disturbance covariances; step three uses the disturbance covariance estimates to compute asymptotically efficient coefficient estimates. Section 3 derives consistent estimates of disturbance covariances. Section 4 illustrates the XYW method with real and simulated U.S. monthly employment and quarterly GNP data and compares its performance with the KF method. Section 5 makes concluding remarks.
2. ESTIMATION OF AUTOREGRESSIVE COEFFICIENTS

Let $y_t$ denote an $m \times 1$ vector of observed, stationary, mean-adjusted variables, assumed to be generated by the vector autoregressive (VAR) process,

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \epsilon_t,$$

where $A_1, \ldots, A_p$ are $m \times m$ matrices of autoregressive coefficients and $\epsilon_t$ is an $m \times 1$ vector of unobserved disturbances, distributed identically, independently of past $\epsilon$'s and $y$'s, with a zero mean vector, and with a positive definite covariance matrix, i.e., $\epsilon_t \sim \text{IID}(0, \Sigma_\epsilon)$, with $\Sigma_\epsilon > 0$. The characteristic equation of the process, $a(\lambda) = |I_m \lambda^p - A_1 \lambda^{p-1} - \ldots - A_p| = 0$, where $|\cdot|$ denotes the determinant and $I_m$ denotes the $m \times m$ identity matrix, has $mp$ roots, $\lambda_1, \ldots, \lambda_{mp}$, some of which may repeat. The process is stationary if and only if all the roots have moduli less than one, i.e., $|\lambda_i| < 1$, for $i = 1, \ldots, mp$.

VAR process (2.1) operates at a basic, fixed, time interval normalized to unity. In each basic period, the process generates a value for each variable. Each variable is observed repeatedly at some regular time interval, which is an integral multiple of the basic interval. Thus, each variable could be observed at its own different frequency. Unless it has missing values for other reasons, the most frequently observed variable is observed every period, as in column one in table 2.1. The XYW method can also be applied to irregularly sampled data, if the basic interval represents a sufficiently short time span, so that every sampling interval is approximately an integral multiple of the basic interval. In such case, even the most frequently observed variable will have repeatedly missing values. At the cost of doing nonlinear estimation, irregular sampling can be accounted for exactly with a continuous-time process (e.g., Zadrozny, 1988).

Consider equation (2.1) in the transposed regression form
where \( B = [A_1, ..., A_p]^T \) is \( mp \times m \), \( x_t = (y_{t-1}^T, ..., y_{t-p}^T)^T \) is \( mp \times 1 \), and superscript \( T \) denotes transposition. Also, consider the \( mr \times 1 \) vector of instrumental variables \( z_t = (y_{t-1}^T, ..., y_{t-r}^T)^T \) where \( r \geq p \). Premultiply equation (2.2) by \( z_t^T \), to obtain

\[
z_t y_t^T = z_t^T x_t^T B + z_t^T e_t,
\]

and, then, average over the sample dates \( t = 1, ..., N \), to obtain

\[
\frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y_{t-1} y_t^T \\ \vdots \\ y_{t-r} y_t^T \\ y_{t-1} y_{t-1}^T \\ \vdots \\ y_{t-r} y_{t-r}^T \end{bmatrix} = \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y_{t-1} y_{t-1}^T & \cdots & y_{t-1} y_{t-p}^T \\ \vdots & \ddots & \vdots \\ y_{t-r} y_{t-1}^T & \cdots & y_{t-r} y_{t-p}^T \end{bmatrix} B + \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y_{t-1}^T \\ \vdots \\ y_{t-r}^T \end{bmatrix}.
\]

The covariances of \( y_t \) in equation (2.4), \( \frac{1}{N} \sum_{t=1}^{N} y_{t-k} y_{t-\ell}^T \) for \( k = 1, ..., r \) and \( \ell = 0, ..., p \), are not all computable because the sample has missing observations due to mixed frequencies. The objective of the next step is to define computable sample covariances, which use as many observations as possible, and to restate equation (2.4) in terms of these covariances. However, before doing this, it will be useful to introduce the population counterpart of equation (2.4) and to illustrate precisely the nature of the mixed-frequency missing-data problem.

The assumption that VAR process (2.1) is stationary implies that, for \( t=1, ..., N \) and \( k=0, \pm 1, \pm 2, ..., \), \( C_k = E \left[ y_t y_{t-k}^T \right] \) exists and is independent of \( t \). Also, the assumption that \( E e_t y_{t-k}^T = 0 \), for \( t=1, ..., N \) and \( k = 1, ..., r \), implies that the expected value of the error term in equation (2.4) is zero. Therefore, taking the unconditional expectation of equation (2.4) and using the skew symmetry of
$C_k$, i.e., $C_k = C_{-k}^T$, results in the population counterpart of equation (2.4),

$$
egin{bmatrix}
C_1^T \\
\vdots \\
C_p^T \\
C_r^T
\end{bmatrix} = 
\begin{bmatrix}
C_0 & \cdots & C_{p-1} \\
\vdots & \ddots & \vdots \\
\vdots & \ddots & \vdots \\
C_{p-1}^T & \cdots & C_0
\end{bmatrix} B. 
$$

(2.5)

The sequence of $m \times m$ equations within equation (2.5) constitutes the Yule-Walker equations, for $k = 1, \ldots, r$. The representative Yule-Walker equation $k$ is

$$
C_k^T = C_{k-1}^T A_1^T + \cdots + C_1^T A_{k-1}^T + C_0^T A_k^T + C_1^T A_{k+1}^T + \cdots + C_{p-k}^T A_p^T.
$$

(2.6)

To illustrate the nature of the mixed-frequency problem, suppose that $y_t$ is bivariate ($m = 2$), so that $y_t = (y_{1t}, y_{2t})^T$, the first variable, $y_{1t}$, is observed every period, and the second variable, $y_{2t}$, is observed every other period. In fact, consider a sample of observations on $y_t$, depicted schematically as

Table 2.1: Hypothetical Mixed-Frequency Sample
where "0" denotes a missing value and "X" denotes an observed value.

To illustrate the problem of computing the covariances in equation (2.4), consider computing the contemporaneous covariance matrix of $y_t$,

$$
\begin{bmatrix}
\frac{1}{N} \sum_{t=1}^{N} y_{1t}^2 & \frac{1}{N} \sum_{t=1}^{N} y_{1t} y_{2t} \\
\frac{1}{N} \sum_{t=1}^{N} y_{2t} y_{1t} & \frac{1}{N} \sum_{t=1}^{N} y_{2t}^2 \\
\end{bmatrix}
$$

(2.7)

with the sample in table 2.1. Evidently, all variance terms of $y_{1t}$ are computable, but only even dated covariance terms and variance terms of $y_{2t}$ are computable. In general, for $k = 0, 1, ..., r$, we define the computable covariance matrix, $\hat{C}_k$, by

$$
\hat{C}_{k,ij} = \frac{1}{N_{k,ij}} \sum_{t=k+1}^{N} \#(y_{it}) \#(y_{jt,k}),
$$

(2.8)

where, for $i, j = 1, ..., m$, $\hat{C}_{k,ij}$ denotes element $(i, j)$ of $\hat{C}_k$, $\#(y_{it}) = y_{it}$ when $y_{it}$ is observed and 0 otherwise, and $N_{k,ij}$ denotes the number of observed terms in the sum. It can be shown that if the VAR process is stationary and ergodic (the latter assumption being a harmless one which cannot be rejected
with typical, single realization, economic data), then, for \( k = 0, 1, \ldots, r \), the computable covariances, \( \hat{C}_k \), will approach the population covariances, \( C_k \), in probability as \( N \to \infty \), i.e., \( \text{plim}_{N \to \infty} \hat{C}_k = C_k \).

Restating equation (2.4) in terms of the computable covariances, we obtain

\[
\begin{bmatrix}
\hat{C}_1^T \\
\vdots \\
\hat{C}_p^T \\
\hat{C}_r^T
\end{bmatrix}
= \begin{bmatrix}
\hat{C}_0 & \ldots & \hat{C}_{p-1}
\end{bmatrix}
B
+ \begin{bmatrix}
\hat{U}_1 \\
\vdots \\
\hat{U}_r
\end{bmatrix}
\tag{2.9}
\]

The errors, \( \hat{U}_k \), for \( k = 1, \ldots, r \), in equation (2.9) are sums of the original errors in equation (2.4), \( U_k = \frac{1}{N} \sum_{t=1}^{N} y_{t-k}^T \epsilon_t \), and additional errors, \( \hat{V}_k \), which are introduced when the full-sample covariances, \( \frac{1}{N} \sum_{t=1}^{N} y_{t-k} y_{t-k}^T \), are replaced by the computable covariances, \( \hat{C}_{t-k} \), i.e., \( \hat{U}_k = U_k + \hat{V}_k \).

It will be convenient to express equation (2.9) as \( m \) \( m \times 1 \) dimensional equations,

\[
[\hat{\epsilon}_1, \ldots, \hat{\epsilon}_m] = \underbrace{r_p[\hat{s}_1, \ldots, \hat{s}_m]}_B + [\hat{\eta}_1, \ldots, \hat{\eta}_m],
\tag{2.10}
\]

where, for \( j = 1, \ldots, m \), \( \hat{\epsilon}_j \) and \( \hat{\eta}_j \) are \( m \times 1 \) vectors of covariances and errors, \( r_p \) is the \( m \times mp \) system matrix of covariances, and \( B_j = (a_{1,j1}, \ldots, a_{1,jm}, \ldots, a_{p,j1}, \ldots, a_{p,jm})^T \) is the \( mp \times 1 \) vector of the coefficients of equation \( j \) of the VAR process. We add subscript \( p \) to \( r_p \) in order to emphasize that
it has p block columns in order to distinguish it later from its extension to the square and symmetric matrix \( r_x \), with r block columns. Similarly, we express equation (2.5), the population counterpart of equation (2.4), as

\[
\begin{pmatrix}
(1, \ldots, m)
\end{pmatrix} = \begin{pmatrix}
1_p & \ldots & 1_p
\end{pmatrix} \begin{pmatrix}
\ldots & \ldots & \ldots
\end{pmatrix},
\]

(2.11)

where, under stationarity and ergodicity, \( \begin{pmatrix}
1, \ldots, m
\end{pmatrix} = \text{plim}_{N \to \infty} \begin{pmatrix}
1, \ldots, m
\end{pmatrix} \) and \( 1_p = \text{plim}_{N \to \infty} r_p \).

A representative \( mr \times 1 \) column of equation (2.10) is

\[
\begin{pmatrix}
f_j
\end{pmatrix} = r_p \begin{pmatrix}
\hat{s}_j
\end{pmatrix} + \begin{pmatrix}
\hat{\beta}_j
\end{pmatrix},
\]

(2.12)

for \( j = 1, \ldots, m \). The first step in the XYW estimation method is to solve \( \begin{pmatrix}
f_j
\end{pmatrix} = r_p \begin{pmatrix}
\hat{s}_j
\end{pmatrix} + \begin{pmatrix}
\hat{\beta}_j
\end{pmatrix} \) for \( \hat{\beta}_j \) for each \( j = 1, \ldots, m \). We shall denote the solution of \( \begin{pmatrix}
f_j
\end{pmatrix} = r_p \begin{pmatrix}
\hat{s}_j
\end{pmatrix} + \begin{pmatrix}
\hat{\beta}_j
\end{pmatrix} \) by \( \hat{\beta}_j \). However, because different variables have different sampling frequencies, not all covariances in \( \begin{pmatrix}
f_j & r_p
\end{pmatrix} \) are computable. Therefore, before solving \( \begin{pmatrix}
f_j
\end{pmatrix} = r_p \begin{pmatrix}
\hat{s}_j
\end{pmatrix} + \begin{pmatrix}
\hat{\beta}_j
\end{pmatrix} \), we have to check \( \begin{pmatrix}
f_j & r_p
\end{pmatrix} \) and delete from it, hence, from equation (2.12), any rows with incomputable, i.e., missing, elements. As will become clear at the end of this section, in order to compute the asymptotically efficient XYW estimator of \( \beta \) and its asymptotic covariance matrix, we also have to check \( \begin{pmatrix}
f_j & r_{x-p}
\end{pmatrix} \) and delete missing elements from it. If \( r > p \), so that \( r_{x-p} \) is non-null, it is more efficient to combine the two operations and check \( \begin{pmatrix}
f_j & r_x
\end{pmatrix} \), for \( j = 1, \ldots, m \), and delete any rows with missing elements.
To see in detail how this step works, consider \[ \{ \hat{t}_1, ..., \hat{t}_m, \hat{r} \} \] written out in full scalar detail for the case \( m = 2, p = 1 \) and \( r = 3 \). These dimensions and the sampling scheme in table 2.1 imply the 6 x 8 matrix

\[
\begin{bmatrix}
\hat{c}_{1,11} & \hat{c}_{1,21} & \hat{c}_{0,11} & \hat{c}_{0,12} & \hat{c}_{1,11} & \hat{c}_{1,12} & \hat{c}_{2,11} & \hat{c}_{2,12} \\
\hat{c}_{1,12} & 0 & \hat{c}_{0,21} & \hat{c}_{0,22} & \hat{c}_{1,21} & 0 & \hat{c}_{2,21} & \hat{c}_{2,22} \\
\hat{c}_{2,11} & \hat{c}_{2,21} & \hat{c}_{1,11} & \hat{c}_{1,21} & \hat{c}_{0,11} & \hat{c}_{0,12} & \hat{c}_{1,11} & \hat{c}_{1,12} \\
\hat{c}_{2,12} & \hat{c}_{2,22} & \hat{c}_{1,12} & 0 & \hat{c}_{0,21} & \hat{c}_{0,22} & \hat{c}_{1,21} & 0 \\
\hat{c}_{3,11} & \hat{c}_{3,21} & \hat{c}_{2,11} & \hat{c}_{2,21} & \hat{c}_{1,11} & \hat{c}_{1,21} & \hat{c}_{0,11} & \hat{c}_{0,12} \\
\hat{c}_{3,12} & 0 & \hat{c}_{2,12} & \hat{c}_{2,22} & \hat{c}_{1,12} & 0 & \hat{c}_{0,21} & \hat{c}_{0,22}
\end{bmatrix}, \tag{2.13}
\]

where, as in table 2.1, a "0" indicates a missing value. Rows 2, 4, and 6 in matrix (2.13) have missing values and must be deleted. In general, different rows will be deleted from \[ \{ \hat{t}_j, \hat{r} \} \] for different values of \( j \). For each \( j \), the deletions may be described algebraically by premultiplying equation (2.12) by \( q_j \times mr \) selection matrix \( H_j \) comprising zeros and ones, where \( q_j (\leq mr) \) denotes the number of retained computable equations in (2.12). Each \( H_j \) matrix is obtained by deleting rows of the \( mr \times mr \) identity matrix corresponding to the incomputable rows of \[ \{ \hat{t}_j, \hat{r} \} \]. In this example,

\[
H_1 = H_2 = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}. \tag{2.14}
\]

In general, the computable reductions of equations (2.12) are

\[
*_{j} = \hat{D}_{j,p} \hat{B}_{j} + \hat{O}_{j}, \tag{2.15}
\]

for \( j = 1, ..., m \), where \( *_{j} = H_j \hat{t}_j \) is \( q_j \times 1 \), \( \hat{D}_{j,p} = H_j \hat{c}_{p} \) is \( q_j \times mp \), and \( \hat{O}_j = H_j \hat{r} \) is \( q_j \times 1 \). In the
above example, \([ [\hat{J}_j, \hat{D}_{j,p}, \hat{O}_j] \) has the same 3 rows for \( j = 1 \) as for \( j = 2 \), but, in general, it will have different rows for different values of \( j \). It will now be convenient to write the \( m \) equations (2.15) as the \( m^2 q \times 1 \) vector equation,

\[
\hat{\beta} = \hat{D}_{p} \hat{\psi} + \hat{O},
\]

(2.16)

where \( q = q_1 + \ldots + q_m (\leq m^2 r) \), \( \hat{\beta} = (\hat{\beta}_1^T, \ldots, \hat{\beta}_m^T)^T \) is \( q \times 1 \), \( \hat{D}_p \) is the \( q \times m^2 p \) block diagonal matrix defined by \( \hat{D}_p = \text{diag}[\hat{D}_{1,p}, \ldots, \hat{D}_{m,p}] \), \( \hat{\beta} = (\hat{\beta}_1^T, \ldots, \hat{\beta}_m^T)^T \) is \( m^2 p \times 1 \), and \( \hat{O} = (\hat{O}_1^T, \ldots, \hat{O}_m^T)^T \) is \( q \times 1 \).

Alternately, \( \hat{D}_p = H[I_m \otimes \hat{\psi}_p] \), \( \hat{\beta} = H \hat{\beta} \), and \( \hat{O} = H \hat{O} \), where \( H \) is the \( q \times m^2 r \) block diagonal matrix defined by \( H = \text{diag}[H_1, \ldots, H_m] \), \( \hat{\beta} = (\hat{\beta}_1^T, \ldots, \hat{\beta}_m^T)^T \) and \( \hat{O} = (\hat{O}_1^T, \ldots, \hat{O}_m^T)^T \) are \( m^2 r \times 1 \), and \( \otimes \) denotes the Kronecker product.

Analytically determining the value of \( q \) in the general case of any mixed-frequency sampling scheme is a complicated combinatorial exercise. By construction \( q \leq m^2 r \), and identifying condition (2.17) implies \( q \geq m^2 p \). Because equation (2.16) is in the form of a regression equation, the standard rank condition for identifying a coefficient vector applies here in analogous form: \( \hat{D}_p \) has full column rank equal to \( m^2 p \) for any finite sample of size \( N \) and in the limit as \( N \to \infty \), i.e.,

\[
| \hat{D}_p^T \hat{D}_p | > 0,
\]

(2.17)

almost surely for finite \( N \geq r + 2 \) and in the limit as \( N \to \infty \).

The identifying conditions on the \( m^2 p \) coefficients in \( \beta \) are the \( q \) scalar equations in \( \hat{D}_p \hat{\psi} = \hat{\beta} \).

To obtain more insight, condition (2.17) may be expanded as follows. For \( j = 1, \ldots, m \), (i) \( \beta_j \) is
identified or determined if (column) \( \text{rank} [\hat{\mathbf{D}}_{j,p}] = mp \); (ii) \( \beta_j \) is exactly identified or uniquely determined if \( \text{rank} [\hat{\mathbf{D}}_{j,p}] = mp = \text{rank} [\hat{\mathbf{D}}_{j,p}^*, \hat{\mathbf{D}}_{j,p}^*] \); and (iii) \( \beta_j \) is overidentified or nonuniquely determined if \( \text{rank} [\hat{\mathbf{D}}_{j,p}] = mp < \text{rank} [\hat{\mathbf{D}}_{j,p}^*, \hat{\mathbf{D}}_{j,p}^*] = mp + 1 \). In theory, only the rows of the population equation \( \hat{\delta}_j = \hat{D}_{j,p} \beta_j \), which are beyond row \( mp \) and are linearly independent of the first \( mp \) rows and of each other, contribute overidentifying restrictions. However, in practice, every row of the sample equation (2.15) will tend to contribute a separate overidentifying restriction on \( \beta_j \). Therefore, in practice, the number of overidentifying restrictions on \( \beta_j \), the coefficients of equation \( j \) in the model, will tend to be \( s_j = q_j - mp \), so that the total number of overidentifying restrictions on all of the coefficients in the model, i.e., on \( \beta \), will tend to be \( s = \sum_{j=1}^{m} s_j \).

Consider estimating \( \beta \) with the generalized method of moments (GMM) estimator, as follows. Let \( Q \) be a \( q \times q \) symmetric positive definite symmetric (PDS) weighting matrix. The objective is to minimize the error quadratic form

\[
\mathbf{f} = Q \mathbf{0} = (\mathbf{\delta} - \hat{\mathbf{D}}_p \beta)^T Q (\mathbf{\delta} - \hat{\mathbf{D}}_p \beta), \tag{2.18}
\]

with respect to \( \beta \) and \( Q \). The quadratic form equals the sum of weighted products of the errors in \( \mathbf{0} \).

First, consider \( Q \) fixed and minimize \( f \) with respect to \( \beta \). To do this, differentiate \( f \) with respect to \( \beta \), to obtain the differential form \( df = -2d\beta^T \hat{\mathbf{D}}_p^T Q \mathbf{\delta} + 2d\beta^T \hat{\mathbf{D}}_p^T Q \hat{\mathbf{D}}_p \beta \). Then, set \( df = 0 \), to obtain the first-order condition

\[
(\hat{\mathbf{D}}_p^T Q \hat{\mathbf{D}}_p) \$ = \hat{\mathbf{D}}_p^T Q \mathbf{\delta}, \tag{2.19}
\]

If identifying condition (2.17) holds, then, equation (2.19) can be solved for \$ as
\[ \hat{s} = (\hat{\mathbf{D}}_p^T \hat{Q} \hat{\mathbf{D}}_p)^{-1} \hat{\mathbf{D}}_p^T \hat{Q} \hat{s}. \]  

(2.20)

The consistency of \( \hat{s} \), for any PDS weighting matrix \( Q \), is quickly proved, as follows. Under stationarity and ergodicity, \( \text{plim}_{N \to \infty} \hat{\mathbf{D}}_p = \mathbf{D}_p \) and \( \text{plim}_{N \to \infty} \hat{s} = \delta \), where \( \mathbf{D}_p = H[I_m \otimes \mathbf{1}_p] \) and \( \delta = H(\ldots) \).

Also, under these conditions, \( \text{plim}_{N \to \infty} \{ \hat{s}_1, \ldots, \hat{s}_n \} = 0 \), which immediately implies \( \text{plim}_{N \to \infty} \mathbf{0} = 0 \).

Equations (2.16) and (2.19) imply \( (\hat{\mathbf{D}}_p^T \hat{Q} \hat{\mathbf{D}}_p)(\hat{s} - \beta) = \hat{\mathbf{D}}_p Q \mathbf{0} \). Slutzky’s theorem says that the probability limit of a continuous transformation of a random variable is the transformation of the probability limit of the random variable. Therefore, the above results and Slutzky’s theorem imply \( (D_p^T Q D_p)(\text{plim}_{N \to \infty} \hat{s} - \beta) = 0 \). Identifying condition (2.17) and the PDS of \( Q \) imply \( (D_p^T Q D_p) \) is nonsingular. Therefore, \( \text{plim}_{N \to \infty} \hat{s} = \beta \).

The asymptotic covariance matrix of \( \hat{s} \), defined by \( \text{AV}(\hat{s}) = \frac{1}{N} \text{plim}_{N \to \infty} N(\hat{s} - \beta)(\hat{s} - \beta)^T \), is derived as follows. Use equation (2.16) to eliminate \( \hat{s} \) from equation (2.20), use the definition \( \mathbf{0} = H \hat{\mathbf{D}}_p \) to eliminate \( \mathbf{0} \), and rearrange terms, to obtain \( \hat{s} - \beta = (D_p^T \hat{Q} D_p)^{-1} \hat{D}_p^T \hat{Q} H \hat{D}_p \), hence,

\[ (\hat{s} - \beta)(\hat{s} - \beta)^T = (D_p^T \hat{Q} D_p)^{-1} \hat{D}_p^T \hat{Q} H \hat{D}_p (D_p^T \hat{Q} D_p)^{-1}. \]  

(2.21)

Again invoking stationarity and ergodicity and applying Slutzky’s theorem, equation (2.21) implies

\[ \text{plim}_{N \to \infty} N(\hat{s} - \beta)(\hat{s} - \beta)^T = (D_p^T Q D_p)^{-1} D_p^T Q H E[N \hat{s} \hat{s}^T] H^T Q D_p (D_p^T Q D_p)^{-1}. \]  

(2.22)

because \( \text{plim}_{N \to \infty} N \hat{s} \hat{s}^T = E[N \hat{s} \hat{s}^T] \).
Next, we determine lower and upper bounds on $E[N \hat{\pi}^T]$ in terms of known quantities. Recall that $\hat{\pi} = \xi + \hat{\varphi}$, where $\xi = \text{vec}[U_1^T, ..., U_r^T]^T$ is the vector of "original" errors in equation (2.4) and $\hat{\varphi} = \text{vec}[\hat{\varphi}_1^T, ..., \hat{\varphi}_r^T]^T$ is the vector of "additional" errors in equation (2.9) caused by some computable covariances, $\hat{C}_{k, ij}$, being based on less than full sample observations, i.e., $N_{k, ij} < N$.

The inability to determine $E[N \hat{\pi}^T]$ exactly in terms of known quantities is caused by the dependence of $\hat{\varphi}$ on unknown elements of $\hat{\beta}$ which are being estimated. That is, under previously made standard assumptions, $\hat{\varphi}$ and $\xi$ approach zero at the same rate, so that unknown terms of $\hat{\pi}^T$ do not approach zero quicker than known terms.

Suppose we have a sample with missing data due to mixed frequencies or for any other reasons. Consider the full $N$-period sample obtained by filling in the missing data. Because every full sample has no "additional" errors, for the filled-in sample, $E[\hat{\pi}^T] = E[\xi^T]$, such that, as will be shown, $E[\hat{\pi}^T] = \frac{1}{N} \Omega$, where $\Omega$ is PDS and independent of $N$. Because using more data reduces error covariances, $E[\hat{\pi}^T]$ of the given missing-data sample satisfies $E[\hat{\pi}^T] - \frac{1}{N} \Omega \geq 0$, where the inequality means that the difference is PDS. Thus, $\frac{1}{N} \Omega$ is a lower bound of $E[\hat{\pi}^T]$.

Alternately, suppose every computed covariance is based on terms indexed in exactly the same sampling periods, i.e., $\hat{C}_{k, ij} = \frac{1}{M} \sum_{l=1}^{M} \#(y_{l+i}, y_{l+j})\#(y_{l+j}, y_{l+i+k})$, where $M = \min\{N_{k,ij} | k = 0, ..., r; i, j = 1, ..., m\}$. For example, in such case, table 2.1 implies $\{t_1, t_2, ..., t_M\} = \{t(r), t(r)+2, ..., N-2, N\}$,
where \( t(r) \) is the first even sampling date \( \geq r + 1 \), so that \( M = \frac{1}{2}[N-t(r)+1] \). Like the full-sample covariances, such covariances result in no "additional" errors and, therefore, imply \( \mathbb{E}[\hat{\Sigma}^{\hat{\Sigma}^T}] = \mathbb{E}[\Sigma^{\Sigma^T}] = \frac{1}{M} \Omega \), where \( \Omega \) is the same PDS matrix as above. However, because using less data increases error covariances, \( \mathbb{E}[\hat{\Sigma}^{\hat{\Sigma}^T}] \) of the given missing-data sample satisfies \( \frac{1}{M} \Omega - \mathbb{E}[\hat{\Sigma}^{\hat{\Sigma}^T}] \geq 0 \). Thus, we have the lower and upper bounds

\[
\Omega \leq \mathbb{E}[N \hat{\Sigma}^{\hat{\Sigma}^T}] \leq \left( \frac{N}{M} \right) \Omega,
\]

where, in a mixed-frequency sample, \( \frac{N}{M} \) is the largest relative periodicity. For example, in monthly-quarterly and quarterly-annual data, \( \frac{N}{M} \) is, respectively, 3 and 4. Although the spread, \( \frac{N}{M} - 1 \), introduces ambiguity about covariances of estimated coefficients, hence, ambiguity about tests of significance of and restrictions on coefficients, it does not affect the asymptotically efficient XYW estimator or the chi-squared test of overidentifying YW restrictions because, as is shown in the following derivations, these depend on the "shape" of \( \mathbb{E}[N \hat{\Sigma}^{\hat{\Sigma}^T}] \), i.e., on \( \Omega \), but not on its size.

For a full sample of \( N \) periods, \( \xi = \frac{1}{N} \sum_{t=1}^{N} \text{vec}(z_t, \hat{\epsilon}) = \frac{1}{N} \sum_{t=1}^{N} \epsilon_t \otimes z_t \), so that \( \mathbb{E}[N\xi^\xi^T] \)

\[
= \frac{1}{N} \sum_{s=1}^{N} \sum_{t=1}^{N} \mathbb{E}[\epsilon_s \otimes \epsilon_t] \otimes z_s \otimes \hat{\epsilon}_t \].
\]

Thus, using the independence of \( \epsilon_s \) from \( \epsilon_t \), for \( s \neq t \), and of \( \epsilon_t \) from lagged values of \( y_t \), we obtain

\[
\mathbb{E}[N\xi^\xi^T] = \Omega = \Sigma_{\epsilon} \otimes \Gamma',
\]

where
Combining equations (2.22) and (2.24), we obtain

$$AV(\hat{\$}) = \frac{1}{N} (D_p^T Q D_p)^{-1} D_p^T Q G r Q D_p (D_p^T Q D_p)^{-1},$$

(2.25)

where $G_r = H(\hat{E}_r \otimes \hat{r}) H^T$. The block diagonal structure of $H$ implies

$$G_r = \begin{bmatrix} F_{11} H_{11}^T & \ldots & F_{1m} H_{1m}^T \\ \vdots & \ddots & \vdots \\ F_{m1} H_{m1}^T & \ldots & F_{mm} H_{mm}^T \end{bmatrix},$$

(2.26)

where $\sigma_{ij}$ is element $(i,j)$ of $\Sigma_e$. The reason for extending the operation of deleting incomputable rows of equations (2.12) to that of deleting incomputable rows of $[\hat{\$}_j, r \hat{e}_j]$, when $r > p$, was to have available at this point selection matrices $H_1, \ldots, H_m$ for which $\hat{G}_x = H(\hat{E}_r \otimes \hat{r}) H^T$ is computable.

Hansen (1982) and Stoica and Soderstrom (1983) proved that $AV(\hat{\$})$ is minimized when

$$Q^* = G_r^{-1}.$$

(2.27)

It is straightforward to show that, under the standard assumptions of controllability and reconstructibility on the VAR parameters (Kwakernaak and Sivan, 1972; Zadrozny, 1988, 1990) $G_r$ is nonsingular. Equation (2.27) is derived as follows. Consider
Minimize \( \text{tr}[V] \) with respect to \( Q \), for given \( D_p \) and \( G_r \). Differentiate \( \text{tr}[V] \) with respect to \( Q \), set the result equal to zero, rearrange terms using the trace rules \( \text{tr}(A) = \text{tr}(A^T) \) and \( \text{tr}(AB) = \text{tr}(BA) \), to obtain

\[
\text{tr}\{dQ[-D_p(D_p Q G_r Q D_p)^{-1} D_p Q G_r Q D_p]^{-1} D_p Q G_r Q D_p]^{-2} D_p Q G_r Q D_p]^{-1} \} = 0. \tag{2.29}
\]

If \( \text{tr}(dA B) = 0 \) for all possible values of \( dA \), then, \( B = 0 \). Using this result, equation (2.29) implies

\[
G_r Q D_p (D_p Q G_r Q D_p)^{-2} D_p = D_p (D_p Q G_r Q D_p)^{-1} D_p Q G_r Q D_p (D_p Q G_r Q D_p)^{-1} D_p. \tag{2.30}
\]

One possible solution of equation (2.30), but generally not the only solution, is \( Q^* = G_r^{-1} \). The nonuniqueness is unimportant because we simply need one optimal weighting matrix, unique or not.

Therefore, setting \( Q = \hat{G}_r^{-1} \) in equations (2.20) and (2.28), results in the second step, asymptotically efficient, XYW estimate of \( \beta \),

\[
\hat{\beta}^* = (\hat{D}_p^T \hat{G}_r^{-1} \hat{D}_p^T)^{-1} \hat{D}_p^T \hat{G}_r^{-1} \hat{\beta}. \tag{2.31}
\]

Following equation (2.23), the asymptotic covariance matrix of \( \hat{\beta}^* \), \( AV(\hat{\beta}^*) = \frac{1}{N} \text{plim}_{N \to \infty} N(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)^T \), has the lower and upper bound estimates

\[
\frac{1}{M} (\hat{D}_p^T \hat{G}_r^{-1} \hat{D}_p^T)^{-1} \leq AV(\hat{\beta}^*) \leq \frac{1}{N} (\hat{D}_p^T \hat{G}_r^{-1} \hat{D}_p^T)^{-1}. \tag{2.32}
\]

Hansen (1982) proved \( \hat{\beta}^* \) is asymptotically normally distributed, if the data generating process is stationary.
Thus, we have derived the asymptotically efficient XYW/GMM estimate, $\hat{\beta}^*$, of $\beta$ and bounds on its asymptotic covariance matrix, $AV(\hat{\beta}^*)$. In practice, $\hat{\beta}^*$, is computed in three steps: (1) compute the consistent estimate $\hat{\beta}$ using equation (2.20), for any choice of PDS weighting matrix $Q$, e.g., $Q = I_q$, the $q \times q$ identity matrix; (2) given $\hat{\beta}$, compute a consistent estimate, $\hat{\Sigma}_e$, of $\Sigma_e$ according to section 3; (3) given $\hat{\Sigma}_e$, compute $\hat{\beta}^*$ using equation (2.31) and bounds on $AV(\hat{\beta}^*)$ using inequalities (2.32).

Following standard GMM estimation theory, the validity of the $s = m^3p$ overidentifying restrictions may be tested using the following Wald type statistic,

$$f^* = \theta^* \hat{G}_x^{-1} \theta^*,$$

(2.33)

where $\theta^* = \hat{\beta}^* - \hat{\beta}$. Under the standard assumptions, $f^*$ is asymptotically distributed chi-squared with $s$ degrees of freedom, i.e., $f^* \sim \chi^2(s)$. As usual, a large value of $f^*$ rejects the overidentifying restrictions. Alternate likelihood-ratio or Lagrange-multiplier type tests are discussed by Greene (1997, ch. 11, sec. 6).

Any subset of overidentifying restrictions can be tested by dropping elements of $\theta^*$ associated with untested restrictions and reducing the degrees of freedom of $f^*$ correspondingly, although doing this seems arbitrary. For simplicity, we have only considered the vector of instrumental variables, $z_t$, which includes all variables of the VAR process at all lags from 1 to $p$. Any subsets of these instruments could be considered and their implied overidentifying restrictions tested, subject only to the requirement that identifying condition (2.17) holds. A more natural way of varying
the overidentifying restrictions is to vary $z_t$ by deleting or adding lags of the variables or by dropping selected coefficients. Dropping coefficients leaves unchanged equations (2.31) and (2.32) for computing $\hat{\$}^*$ and $AV(\hat{\$}^*)$, although columns of $\hat{D}_p$ associated with dropped coefficients drop out; the $P^2$ test changes only in that degrees of freedom increase by the number of dropped coefficients.

Residual-based diagnostic tests to check adequacy of fit of the estimated model, for example, in terms of absence of residual autocorrelations (Ljung and Box, 1978) or minimality of an information criterion (Schwarz, 1978), should also be conducted. Although, as discussed at the beginning of section 3, residuals of low-frequency variables in a mixed-frequency sample generally cannot be computed directly, residuals of all variables can be computed for samples with any mixtures of frequencies by applying the Kalman filter to the estimated model (Zadrozny, 1990).
3. ESTIMATION OF DISTURBANCE COVARIANCES

In this section, we discuss computing a consistent estimate, $\hat{\Sigma}_e$, of the disturbance covariance matrix, $\Sigma_e$, a necessary ingredient for computing the XYW coefficient estimate $\hat{\beta}^*$, its asymptotic covariance matrix, $AV(\hat{\beta}^*)$, and test the statistic $f^*$, according to section 2.

If the sample of observations has no missing values, then, we could estimate $\hat{\Sigma}_e$ in the usual way as $\hat{\Sigma}_e = \frac{1}{N-p} \sum_{t=p+1}^{N} \hat{\epsilon}_t \hat{\epsilon}_t^T$ where $\hat{\epsilon}_t$ is the residual vector computed as $\hat{\epsilon}_t = y_t - \hat{A}_1 y_{t-1} - \cdots - \hat{A}_p y_{t-p}$. However, if the sample has missing values due to mixed frequencies, residuals are computable depending on whether certain coefficients are zero. We illustrate this point in the case of estimating a bivariate first-order process with the sampling scheme in table 2.1. In this case, the first residual, $\hat{\epsilon}_{1,t}$, is always computable for odd values of $t = p+1, \ldots, N$ and for all values of $t$ in this range if $a_{1,2} = 0$; the second residual, $\hat{\epsilon}_{2,t}$, is computable for even values of $t = p+1, \ldots, N$ if $a_{1,2} = 0$ and for no values of $t$ otherwise. Therefore, in estimation with mixed frequency data, $\hat{\Sigma}_e = \frac{1}{N-p} \sum_{t=p+1}^{N} \hat{\epsilon}_t \hat{\epsilon}_t^T$ is computable only in special cases.

We now derive a consistent estimate of $\Sigma_e$, which is computable for any VAR model estimated with any mixed-frequency data. The estimate takes as a starting point the initial consistent estimate $\hat{\Sigma} = vec[(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_p)^T]$ of section 2. If $\hat{\Sigma}$ represents a stationary model, then, that model has the Wold MA($\infty$) representation

$$y_t = R(L)e_t = (\sum_{i=0}^{\infty} R_i L^i)e_t = \sum_{i=0}^{\infty} R_i e_{t-i}, \quad (3.1)$$
where \( R(L) = \hat{A}(L)^{-1} \), \( A(L) = I_m - \hat{A}_1 L - \cdots - \hat{A}_p L^p \), \( I_m \) is the \( m \times m \) identity matrix, and \( L \) denotes the lag operator. The Wold coefficients are computed by iterating on

\[
R_i = \sum_{j=1}^{\min(i,p)} \hat{A}_j R_{i-j},
\]

for \( i = 1, 2, \ldots \), starting with \( R_0 = I_m \). The Wold representation implies

\[
\hat{C}_0 = \sum_{i=0}^{\infty} R_i \hat{E}_i R_i^T.
\]

where \( \hat{C}_0 \) denotes the computable sample estimate of \( C_0 = E \sigma_{\varepsilon} \sigma_{\varepsilon}^T \), defined in section 2.

Applying the familiar vectorization rule \( \text{vec}(ABC) = [C^T \otimes A] \text{vec}(B) \) to equation (3.3), we obtain

\[
\text{vec}(\hat{E}_i) = [\sum_{i=0}^{\infty} (R_i \otimes R_i)]^{-1} \text{vec}(\hat{C}_0).
\]

as a consistent, but incomputable, estimate of \( \Sigma_\varepsilon \). Below, we prove that \( \text{plim}_{N \to \infty} \hat{E}_i = \Sigma_\varepsilon \). For a finite \( K \),

\[
\text{vec}(\hat{E}_{i,K}) = [\sum_{i=0}^{K} (R_i \otimes R_i)]^{-1} \text{vec}(\hat{C}_0)
\]

is a computable approximation of \( \hat{E}_i \). For greater accuracy, \( \hat{E}_{i,K} \) should be computed by solving the reduced equation of the form \( \text{vech}(\hat{C}_0) = M \text{vech}(\hat{E}_{i,K}) \), obtained by removing redundancies in \( \hat{C}_0 \) and \( \hat{E}_i \) due to symmetry, where \( \text{vech}(\cdot) \) denotes the vectorization of the nonredundant lower triangular part of a symmetric square matrix.

We need to answer two questions: (1) under what conditions is \( \sum_{i=0}^{K} (R_i \otimes R_i) \) nonsingular and (ii) how large should \( K \) be to ensure that the relative difference between \( \hat{E}_{i,K} \)
and $\Sigma_\varepsilon$ is within a prescribed tolerance, $\tau$?

To answer question (i): First, because $\hat{E}_\varepsilon$ consistently estimates $\Sigma_\varepsilon$, if the sample size, $N$, is sufficiently large, the relative difference between $\hat{E}_\varepsilon$ and $\Sigma_\varepsilon$ is within $\tau$, with probability arbitrarily close to one. Therefore, we assume that $N$ is sufficiently large and consider $\hat{E}_\varepsilon$ and $\Sigma_\varepsilon$ to be identical.

Let $\alpha = \max\{ |\hat{\theta}_1|, \ldots, |\hat{\theta}_{mp}| \}$, where $\hat{\theta}_1, \ldots, \hat{\theta}_{mp}$ are the characteristic roots of the estimated model. For convenience, we assume the roots are distinct. Under this assumption, using an eigenvalue decomposition of $(R_j \otimes R_j)$, it is easy to prove that $\sum_{j=0}^{K} (R_j \otimes R_j)$ is nonsingular, for any $K = 0, 1, \ldots$, and in the limit as $K \to \infty$, if (but not only if) the estimated model is stationary.

To answer question (ii): Also using the eigenvalue decomposition of $(R_j \otimes R_j)$, it is easy to show that

$$
\| \text{vec}(\hat{E}_{\varepsilon, K} - \Sigma_\varepsilon) \|_2 / \| \text{vec}(\Sigma_\varepsilon) \|_2 \leq \frac{n^{2K+2}}{1 - n^{2K+2}},
$$

(3.6)

where $\|\bullet\|_2$ denotes the 2-norm of a vector (Golub and Van Loan, 1989, pp. 53-58). Therefore,

$$
K > K = \text{int} \left[ \frac{\log(J)}{\log(\tau)} - 1 \right],
$$

(3.7)

where int[•] is the integral part of a real number, ensures that relative accuracy is within tolerance, i.e., $\| \text{vec}(\hat{E}_{\varepsilon, K} - \Sigma_\varepsilon) \|_2 / \| \text{vec}(\Sigma_\varepsilon) \|_2 \leq \tau$. For example, when $\alpha = .90$ and $\tau = .0001$, $K = 43$.

The answers to questions (i) and (ii) suggest that to compute $\Sigma_\varepsilon$ accurately by evaluating equation (3.5), it suffices to use truly or nearly stationary data to obtain a stationary initial estimated
model and to set $K$ consistent with inequality (3.7). However, even if the data generating process is stationary, sampling variability may cause the initial estimated model to be nearly or exactly nonstationary, which may cause $\sum_{i=0}^{K} (R_{i} \otimes R_{i})$ to be singular for certain values of $K$ or may cause it to diverge or converge slowly as $K \to \infty$. In practice, numerical errors due to finite-precision computations cause $\| \text{vec}(E_{i,K} - \Sigma_{e}) \|_{2} / \| \text{vec}(\Sigma_{e}) \|_{2}$ to reach a minimum at some $K = \bar{K}$, which is often less than $K$, so that prescribed relative accuracy cannot be achieved. Therefore, inequality (3.7) indicates that $K$ should be set as large as possible in order to minimize theoretical errors, but $K$ should be set as small as possible in order to minimize numerical errors.

The following three-part pre-filtering step should reduce or eliminate this conflict: (i) filter some or all of the characteristic roots out of the data and the Wold coefficients, (ii) recompute $\hat{\mathcal{C}}_{0}$ using the filtered data, and (iii) evaluate equation (3.5) for the filtered Wold coefficients and the recomputed $\hat{\mathcal{C}}_{0}$. In the "partial" version, nonstationary and "large" stationary characteristic roots are filtered out; in the "full" version, all characteristic roots are filtered out. After partial filtering, $\sum_{i=0}^{K} (R_{i} \otimes R_{i})$ should be firmly nonsingular, for all finite $K = 0, 1, \ldots$ and in the limit as $K \to \infty$, and $\alpha$ should be acceptably below one. After full filtering, equations (3.4) and (3.5) are theoretically identical for $K = (m+s)(\rho-1)$, where $s$ and $\rho$ are defined next, and theoretical errors of using equation (3.5) disappear, but unavoidable numerical errors remain.

We assume the characteristic roots of the initial estimated model are ordered in decreasing size, as $|\hat{\theta}_{1}| \geq \ldots \geq |\hat{\theta}_{s}| > |\hat{\theta}_{s+1}| \geq \ldots \geq |\hat{\theta}_{mp}|$, such that the $s$ largest roots are filtered out.
Complex roots must be filtered out in conjugate pairs. Let $D$ denote the shortest sampling interval for which every variable is observed; e.g., monthly-quarterly, quarterly-annual, and monthly-annual or monthly-quarterly-annual data, respectively, imply $D = 3, 4, \text{ and } 12$ periods. We need to compute the filtered data vectors

$$\tilde{y}_t = \hat{f}(L^D)y_t,$$  \hspace{1cm} (3.8)

for $t = \rho s + 1, \ldots, N$, where $\hat{f}(L^D) = (1 - \delta_1 L^D) \cdots (1 - \delta_s L^D)$. Throughout this section a "~" denotes a filtered datum or estimate.

To see that $\hat{f}(L^D)$ indeed filters out the $s$ largest roots from $R(L)$, note that $R(L) = \hat{A}^*(L)/[\hat{f}(L) \hat{g}(L)]$, where $\hat{A}^*(L)$ is the adjoint matrix of $\hat{A}(L)$, $\hat{f}(L) = (1 - \delta_1 L) \cdots (1 - \delta_s L)$, and $\hat{g}(L) = (1 - \delta_{s+1} L) \cdots (1 - \delta_m L)$, and that $\hat{f}(L^D) = \hat{f}(L)\hat{h}(L)$, where $\hat{h}(L) = \prod_{i=1}^s \sum_{j=0}^{D-1} (\delta_i L)^j$. Then, we see that $R(L) = \hat{f}(L^D)R(L) = \hat{h}(L)\hat{A}^*(L)/\hat{g}(L)$ is filtered as desired, because the denominator, $\hat{g}(L)$, contains only the $mp$-s smallest roots. If all roots are filtered out, then, $R(L) = \hat{h}(L)\hat{A}^*(L)$ is a finite polynomial in $L$ of degree $(m - 1)p + mp(D - 1)$.

However, $s$ should be kept small because the filtering sacrifices $\rho s$ initial observations.

To implement $\hat{f}(L^D)$, we consider the iterations $\tilde{y}_t^{(\ell)} = (1 - \delta(L^D))\tilde{y}_t^{(\ell-1)}$, for $\ell = 1, \ldots, s$, starting with $\tilde{y}_t^{(0)} = y_t$. The iterations fall into two cases depending on whether $\delta_1$ is real or complex.

For $\ell = 1, \ldots, s$, if $\delta_1$ is real, we compute $\tilde{y}_t^{(\ell)}$ as
\[
\tilde{y}_t^{(l)} = \tilde{y}_t^{(l-1)} - \hat{\theta}_l \tilde{y}_{t-D}^{(l-1)},
\]
(3.9)

for \( t = \rho l + 1, \ldots, N \). If \( \hat{\theta}_l = \hat{\alpha}_l + \hat{\beta}_l \sqrt{-1} \) is complex, then, we assume the roots have been ordered so that \( \hat{\theta}_{t+1} = \hat{\alpha}_l - \hat{\beta}_l \sqrt{-1} \) is the complex conjugate of \( \hat{\theta}_l \), where \( \hat{\alpha}_l \) and \( \hat{\beta}_l \) are real numbers.

Therefore, for \( l = 1, \ldots, s \), if \( \hat{\theta}_l \) is complex, we skip computing \( \tilde{y}_t^{(l)} \) and compute \( \tilde{y}_t^{(l+1)} \) as

\[
\tilde{y}_t^{(l+1)} = \tilde{y}_t^{(l-1)} - N_{t1} \tilde{y}_{t-D}^{(l-1)} - N_{t2} \tilde{y}_{t-2D}^{(l-1)},
\]
(3.10)

for \( t = \rho(l+1) + 1, \ldots, N \), where \( N_{t1} = 2 \hat{\alpha}_l \) and \( N_{t2} = -\hat{\alpha}_l^2 - \hat{\beta}_l^2 \). Thus, we obtain \( \tilde{y}_t = \tilde{y}_t^{(s)} \), for \( t = \rho s + 1, \ldots, N \).

The Wold representation of \( \tilde{y}_t \) is

\[
\tilde{y}_t = R(L)e_t = \sum_{l=0}^{\infty} R_{l} \tilde{y}_{t-l},
\]
(3.11)

where \( R(L) = \hat{f}(L^D) R(L) \). We compute \( R(L) \) analogously to \( \tilde{y}_t \). For \( l = 1, \ldots, s \), whether \( \hat{\theta}_l \) is real or complex, \( R_{l}^{(l)} = R_{l} \), for \( i = 0, \ldots, D-1 \); if \( \hat{\theta}_l \) is real, we compute \( R_{l}^{(l)} \) as

\[
R_{l}^{(l)} = R_{l}^{(l-1)} - \hat{\theta}_l R_{l-D}^{(l-1)},
\]
(3.12)

for \( i = D, \ldots, K \); otherwise, if \( \hat{\theta}_l \) and \( \hat{\theta}_{l+1} \) are a complex conjugate pair, we skip computing \( R_{l}^{(l)} (L) \) and compute \( R_{l}^{(l+1)} (L) \) as

\[
R_{l}^{(l+1)} = R_{l}^{(l-1)} - N_{l1} R_{l-D}^{(l-1)},
\]
(3.13)

for \( i = \rho, \ldots, 2\rho-1 \), and as
\[ R^{(i+1)}_i = R^{(i-1)}_i - \hat{N}_1 R^{(i-1)}_{1-D} - \hat{N}_2 R^{(i-1)}_{1-2D}, \]  

(3.14) for \( i = 2, ..., K \). Thus, we obtain \( R_i = R^{(s)}_i \), for \( i = 0, ..., K \).

Corresponding to \( \hat{C}_0 \), in terms of filtered data, we define

\[ \tilde{C}_{0,ij} = \frac{1}{N_0,ij} \sum_{t=1}^N \#(\tilde{y}_{i,t}) \#(\tilde{y}_{j,t}). \]  

(3.15)

where, for \( i, j = 1, ..., m \), \( \tilde{C}_{0,ij} \) denotes element \((i,j)\) of \( \tilde{C}_0 \), \#(\tilde{y}_{i,t}) = \tilde{y}_{i,t} \) when \( \tilde{y}_{i,t} \) is observed and 0 otherwise, and \( N_{0,ij} \) denotes the number of observed terms in the sum. Therefore, corresponding to equation (3.3), we obtain

\[ \tilde{C}_0 = \sum_{i=0}^{m} R_i \hat{E} R_i^T. \]  

(3.16)

Finally, after truncating and vectorizing equation (3.16), we obtain the post-filtering estimate of \( \Sigma_e \),

\[ \text{vec} (\hat{E}_i) = [\sum_{j=0}^{K} (R_j \otimes R_j)]^{-1} \text{vec}(\tilde{C}_0). \]  

(3.17)

To conclude this section, we prove that \( \text{plim}_{N \to \infty} \hat{E}_i = \Sigma_e \). The proof that \( \text{plim}_{N \to \infty} \hat{E}_i = \Sigma_e \) is identical except for one extra detail which we note. Subtracting the population counterpart from equation (3.3), we obtain

\[ \hat{C}_0 - C_0 = \sum_{i=0}^{m} (R_i \hat{E} R_i^T - R_i \hat{E} R_i^T). \]  

(3.18)

In section 2, we proved that \( \text{plim}_{N \to \infty} \hat{\Phi} = \beta \), which, using Slutzky's theorem, implies \( \text{plim}_{N \to \infty} R(L) = \psi(L) \). A stationary data generating process implies \( \text{plim}_{N \to \infty} \hat{C}_0 = C_0 \). Then, vectorizing equation (3.18) and applying Slutzky's theorem, we obtain
\[ [\sum_{j=0}^{\infty} (R_j \otimes R_j)](\text{plim}_{N \to \infty} \mathbf{E}_j - \Sigma_c) = 0. \] (3.19)

However, because a stationary process implies \( \sum_{j=0}^{\infty} (R_j \otimes R_j) \) is nonsingular, equation (3.19) implies \( \text{plim}_{N \to \infty} \mathbf{E}_j = \Sigma_c \). The extra detail in the proof that \( \text{plim}_{N \to \infty} \mathbf{E}_j = \Sigma_c \) is that \( \text{plim}_{N \to \infty} \hat{\mathbf{f}} = \beta \) also implies \( \text{plim}_{N \to \infty} \hat{R}(L) = f(L) \), the true filter based on the \( s \) largest population characteristic roots, so that \( \text{plim}_{N \to \infty} R(L) = f(L)\psi(L) \).
4. APPLICATION TO U.S. MONTHLY EMPLOYMENT AND QUARTERLY GNP

In this section, we present a first application and test of extended Yule-Walker (XYW) estimation, in particular, compare its accuracy with maximum likelihood (ML) estimation. We implemented the ML estimation with monthly-quarterly data as in Zadrozny (1990), by computing the Gaussian likelihood function with the Kalman filter.

We used historical observations on U.S. monthly total employment and quarterly real GNP, spanning 492 months from January 1948 to December 1988. The same data were used in Zadrozny (1990) to estimate a monthly forecasting model. As before, we used the variables in the form of first-differences of (natural) logarithms. We used total employment in the form of month-to-month differences of logarithms, denoted DLEMP, and real GNP in the form of quarter-to-quarter differences of logarithms, denoted DLGNP. Like the pre-transformed observations, DLEMP is observed every period or month and DLGNP is observed every third period or quarter.

Because GNP observations represent accumulated output over a quarter, it is natural to assign DLGNP to the third month of a quarter. Also, consequently, GNP is said to be temporally aggregated. If, in the hypothetical contrasting case, GNP observations represented output in a particular month of a quarter, then, GNP would be not-temporally-aggregated or skip-sampled. Although treating GNP as temporally aggregated slightly improves model fit and out-of-sample forecasting (Zadrozny, 1990), this was not attempted here.

First, we applied ML estimation to the historical sample, to determine a basic model. Then, we used the basic model to generate 100 simulated samples of size 492. Finally, we applied ML and XYW estimation to the 100 samples, to reestimate the basic model. We computed bias, standard deviation, and root mean-squared error statistics to compare the accuracies of ML and XYW estimation.
estimates. 100 samples were sufficient for the statistics to stabilize within three decimal digits.

We restricted the search for the basic model to full VAR(p) models, for p = 1, ..., 4, where "full" means all variables in the model appear at all lags from 1 to p. We tested significance of residual autocorrelations at lags 1 to 36 with Ljung and Box (1978) Q statistics. The tests indicated residuals of models for p = 2, 3, 4 were insignificantly autocorrelated at the 1% level. Therefore, we considered the full VAR(p) models for p = 2, 3, 4 to be acceptable for the historical sample. Because the VAR(2) model had the lowest value of Schwarz's (1978) Bayesian information criterion, we chose it as the best, basic model.

Table 4.1 reports the results of ML estimation of the basic model using the monthly-quarterly historical sample. The top two rows of the table display estimated coefficients, as transpositions of column vectorizations of coefficient matrices. The middle two columns display R-squared statistics and marginal significance levels of Ljung and Box Q statistics of residuals of the estimated equations. The bottom four columns display characteristic roots, their moduli, and periods of cycles implied by the roots. Abstracting from disturbances, the model can be expressed as $y_t = \sum_{i=1}^{4} c_i \lambda_i^t$, for $t = 1, 2, ..., \text{where } c_i$ are 2x1 vectors determined by starting values, $y_0$ and $y_{-1}$, and $\lambda_i$ are characteristic roots. Each term represents a cycle with a period or length of $\rho_i$ months, whose amplitude declines at the exponential rate $|\lambda_i|$. If $\lambda_i$ is real and positive, $\rho_i = \infty$; if $\lambda_i$ is real and negative, $\rho_i = 2$ months; if $\lambda_i = a \pm b(-1)^{1/2}$ is complex, where a and b are real numbers, $\rho_i = 2\pi / \cos^{-1}[(a^2 + b^2)^{1/2}]$ months. Except in special cases, the coefficients provide no direct information about the model's dynamics.

Table 4.1: Maximum Likelihood Estimate of Basic Model Using Monthly-Quarterly Historical Sample.
### Coefficient Estimates

<table>
<thead>
<tr>
<th>vec((\hat{A}_1))</th>
<th>.1311</th>
<th>.2139</th>
<th>.2130</th>
<th>.1827</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec((\hat{A}_2))</td>
<td>.1109</td>
<td>.3417</td>
<td>.1631</td>
<td>.1431</td>
</tr>
</tbody>
</table>

### Model Fit Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>R(^2)</th>
<th>Q(36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLEMP</td>
<td>.2476</td>
<td>.0327</td>
</tr>
<tr>
<td>DLGNP</td>
<td>.4519</td>
<td>.7265</td>
</tr>
</tbody>
</table>

### Characteristic Roots

<table>
<thead>
<tr>
<th>Number</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8222</td>
<td>.0000</td>
<td>.8222</td>
<td>(\infty)</td>
</tr>
<tr>
<td>2</td>
<td>-.4365</td>
<td>.0000</td>
<td>.4365</td>
<td>2.000</td>
</tr>
<tr>
<td>3</td>
<td>-.0360</td>
<td>.3313</td>
<td>.3332</td>
<td>3.742</td>
</tr>
<tr>
<td>4</td>
<td>-.0360</td>
<td>-.3313</td>
<td>.3332</td>
<td>3.742</td>
</tr>
</tbody>
</table>

We generated 100 simulated samples with the basic model. For each sample, we computed
\(y_t = \hat{A}_1 y_{t-1} + \hat{A}_2 y_{t-2} + \epsilon_t\) for \(t = 1, \ldots, 542\), starting from \(y_0 = 0\) and \(y_{-1} = 0\), where \(\epsilon_t\) were pseudo-random draws from the \(N(0, \bar{E})\) distribution. We used the pseudo-random number generator RAN3 described in Press et al. (1992). To eliminate any possible influences of starting values, we dropped \(y_1, \ldots, y_{50}\) from each sample. Thus, we obtained 100 simulated full-monthly samples of size 492. As in the basic model, the second variable in each simulated sample represents DLGNP. Finally, we derived monthly-quarterly samples from the full samples, by replacing values of the second variables in periods 1, ..., 490, 491, with a missing value indicator.
Table 4.2: Maximum Likelihood Estimates of Basic Model Using Full-Monthly Simulated Samples

<table>
<thead>
<tr>
<th>Average Coefficient Estimates</th>
<th>vec((\hat{A}_1))^T</th>
<th>vec((\hat{A}_2))^T</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec((\hat{A}_1))^T</td>
<td>.1264</td>
<td>.1830</td>
</tr>
<tr>
<td>vec((\hat{A}_2))^T</td>
<td>.1185</td>
<td>.3051</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biases of Coefficient Estimates</th>
<th>vec((\hat{A}_1))^T</th>
<th>vec((\hat{A}_2))^T</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec((\hat{A}_1))^T</td>
<td>-.0048</td>
<td>-.0310</td>
</tr>
<tr>
<td>vec((\hat{A}_2))^T</td>
<td>.0076</td>
<td>-.0365</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviations of Coefficient Estimates</th>
<th>vec((\hat{A}_1))^T</th>
<th>vec((\hat{A}_2))^T</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec((\hat{A}_1))^T</td>
<td>.0437</td>
<td>.0385</td>
</tr>
<tr>
<td>vec((\hat{A}_2))^T</td>
<td>.0414</td>
<td>.0375</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Root Mean-Squared Errors of Coefficient Estimates</th>
<th>vec((\hat{A}_1))^T</th>
<th>vec((\hat{A}_2))^T</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec((\hat{A}_1))^T</td>
<td>.0440</td>
<td>.0494</td>
</tr>
<tr>
<td>vec((\hat{A}_2))^T</td>
<td>.0421</td>
<td>.0523</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic Roots of Average Estimates</th>
<th>Number</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.7839</td>
<td>.0000</td>
<td>.7839</td>
<td>∞</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.4399</td>
<td>.0000</td>
<td>.4399</td>
<td>2.000</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.0292</td>
<td>.3153</td>
<td>.3167</td>
<td>3.778</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.0292</td>
<td>-.3153</td>
<td>.3167</td>
<td>3.778</td>
</tr>
</tbody>
</table>
Table 4.3: Maximum Likelihood Estimates of Basic Model Using Monthly-Quarterly Simulated Samples.

| vec(\(\hat{A}_1\))^T | .1305 | .2043 | .2166 | .1517 |
| vec(\(\hat{A}_2\))^T | .1135 | .3266 | .1830 | .1215 |

| vec(\(\hat{A}_1\))^T | -.0006 | -.0096 | .0036 | -.0310 |
| vec(\(\hat{A}_2\))^T | .0026 | -.0151 | .0199 | -.0216 |

| vec(\(\hat{A}_1\))^T | .0518 | .0651 | .0614 | .0608 |
| vec(\(\hat{A}_2\))^T | .0461 | .0618 | .0677 | .0632 |

| vec(\(\hat{A}_1\))^T | .0518 | .0658 | .0615 | .0683 |
| vec(\(\hat{A}_2\))^T | .0462 | .0636 | .0706 | .0668 |

<table>
<thead>
<tr>
<th>Number</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8067</td>
<td>.0000</td>
<td>.8067</td>
<td>(\infty)</td>
</tr>
<tr>
<td>2</td>
<td>-.4446</td>
<td>.0000</td>
<td>.4446</td>
<td>2.000</td>
</tr>
<tr>
<td>3</td>
<td>-.0399</td>
<td>.3558</td>
<td>.3580</td>
<td>3.734</td>
</tr>
<tr>
<td>4</td>
<td>-.0399</td>
<td>-.3558</td>
<td>.3580</td>
<td>3.734</td>
</tr>
</tbody>
</table>

We reestimated the basic model in four ways, using the 100, full-monthly and monthly-quarterly, simulated samples of size 492: (i) we applied ML estimation to full-monthly samples; (ii) we applied ML estimation to monthly-quarterly samples; (iii) we applied XYW estimation to full-monthly samples; and, (iv) we applied XYW estimation to monthly-quarterly samples.
Tables 4.2 and 4.3 display average coefficient estimates and their biases, standard deviations, and root mean-squared errors (RMSEs). Biases are average estimates minus estimates in table 4.1, used to generate the samples. Tables 4.2 and 4.3 show that monthly-quarterly ML estimates are only slightly less accurate than full-monthly ML estimates.

In the XYW estimations, we determined identifying restrictions as in Section 2. We considered all variables in the model from 1 to r lags as instruments. For each r, we considered YW blocks 1 to r and deleted incomputable equations from the blocks, to obtain a total of q computable restrictions for identifying the coefficients. The bivariate VAR(2) model and monthly-quarterly data imply a total of \( s = 2r - 8 \) overidentifying restrictions. Thus, to identify all 8 coefficients, \( r \) must be \( \geq 4 \). When \( r = 4 \), the coefficients are exactly identified; each unit increase in \( r \) beyond 4 adds 2 overidentifying restrictions.
Figure 4.1: Auto- and Cross-Correlograms of the Historical Monthly-Quarterly DLEMP and DLGNP Observations.
Figure 4.1 provides some insights for setting \( r \). It displays own- and cross-correlograms of DLEMP and DLGNP in the historical monthly-quarterly sample. The own-DLEMP and cross correlograms, which can be computed for all monthly lags (cf., table 2.1), are depicted by continuous lines. The own-DLGNP correlogram, which can be computed only at quarterly lags, is depicted by a broken line. The dashed parallel lines denote approximate 95% confidence bounds. Beyond about 8 months the correlograms are insignificant or marginally significant, i.e., within or close to the 95% bounds. Thus, Figure 4.1 suggests setting \( r \) to at least 8 months.

We applied XYW, for \( r = 4, 8, 12, 16, 20, \) and 24, to reestimate the basic model with the full-monthly and monthly-quarterly, simulated samples. Table 4.4 reports average, minimum, and maximum RMSE statistics of these reestimations. XYW/\( k \) signifies XYW estimation with \( r = k \). In full-monthly estimation, XYW/4 produces the lowest average, minimum, and maximum RMSEs, and the RMSEs tend to increase slowly with \( r \). Because XYW/4 is basically ordinary least squares, ML and XYW/4 results are very similar. Tables 4.2 and 4.5 allow more detailed comparisons between full-monthly ML and XYW/4 results. Apparently, overidentifying XYW restrictions provide no benefits in full-sample estimation of a correctly specified model.

By contrast, in monthly-quarterly estimation, the 16 overidentifying restrictions introduced by increasing \( r \) from 4 to 12 cause average, minimum, and maximum RMSEs to decrease by 76% to 87%. Additional restrictions lead to small additional decreases in RMSEs. Tables 4.3 and 4.6 allow more detailed comparisons between monthly-quarterly ML and XYW/24 results. The tables show: only slightly higher XYW biases; with one exception, slightly higher XYW standard deviations and RMSEs for coefficients of DLEMP in columns 2 and 3; and, from 75% to 185% higher XYW standard deviations and RMSEs for coefficients of DLGNP in columns 4 and 5. However, average
XYW estimates are quite close to average ML estimates in terms of moduli and periods of characteristic roots.

Table 4.4: Summary RMSE Statistics of Maximum Likelihood and Extended Yule-Walker Estimates of Basic Model Using Full-Monthly and Monthly Quarterly Simulated Samples.

<table>
<thead>
<tr>
<th>Method</th>
<th>Fully-Monthly</th>
<th>Monthly-Quarterly</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>.0429</td>
<td>.0369</td>
</tr>
<tr>
<td>XYW/4</td>
<td>.0531</td>
<td>.0385</td>
</tr>
<tr>
<td>XYW/8</td>
<td>.0548</td>
<td>.0473</td>
</tr>
<tr>
<td>XYW/12</td>
<td>.0638</td>
<td>.0424</td>
</tr>
<tr>
<td>XYW/16</td>
<td>.0533</td>
<td>.0470</td>
</tr>
<tr>
<td>XYW/20</td>
<td>.0552</td>
<td>.0501</td>
</tr>
<tr>
<td>XYW/24</td>
<td>.0790</td>
<td>.0519</td>
</tr>
</tbody>
</table>
Table 4.5: Extended Yule-Walker Estimates of Basic Model
Using Full-Monthly Simulated Samples.

\( r = 4, \ q = 8, \ s = 0 \)

<table>
<thead>
<tr>
<th>Average Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec(( \hat{\Lambda}_1 ))( ^T )</td>
</tr>
<tr>
<td>vec(( \hat{\Lambda}_2 ))( ^T )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biases of Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec(( \hat{\Lambda}_1 ))( ^T )</td>
</tr>
<tr>
<td>vec(( \hat{\Lambda}_2 ))( ^T )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviations of Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec(( \hat{\Lambda}_1 ))( ^T )</td>
</tr>
<tr>
<td>vec(( \hat{\Lambda}_2 ))( ^T )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Root Mean-Squared Errors of Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>vec(( \hat{\Lambda}_1 ))( ^T )</td>
</tr>
<tr>
<td>vec(( \hat{\Lambda}_2 ))( ^T )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic Roots of Average Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
Table 4.6: Extended Yule-Walker Estimates of Basic Model Using Monthly-Quarterly Simulated Samples.

\( r = 24, \, q = 48, \, s = 40 \)

<table>
<thead>
<tr>
<th>( \hat{\text{vec}}(\hat{A}_1)^T )</th>
<th>( \hat{\text{vec}}(\hat{A}_2)^T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1613</td>
<td>.2090</td>
</tr>
<tr>
<td>.1459</td>
<td>.3118</td>
</tr>
</tbody>
</table>

| \( \text{Biases of Coefficient Estimates} \) |
|---|---|---|---|
| \( \hat{\text{vec}}(\hat{A}_1)^T \) | .0302 | -.0049 | -.0445 | .0013 |
| \( \hat{\text{vec}}(\hat{A}_2)^T \) | .0350 | -.0299 | -.0142 | -.0462 |

| \( \text{Standard Deviations of Coefficient Estimates} \) |
|---|---|---|---|
| \( \hat{\text{vec}}(\hat{A}_1)^T \) | .0624 | .0964 | .1077 | .1730 |
| \( \hat{\text{vec}}(\hat{A}_2)^T \) | .0563 | .0979 | .1072 | .1463 |

| \( \text{Root Mean-Squared Errors of Coefficient Estimates} \) |
|---|---|---|---|
| \( \hat{\text{vec}}(\hat{A}_1)^T \) | .0693 | .0965 | .1165 | .1730 |
| \( \hat{\text{vec}}(\hat{A}_2)^T \) | .0663 | .1024 | .1081 | .1534 |

| \( \text{Characteristic Roots of Average Estimates} \) |
|---|---|---|---|---|
| Number | Real | Imaginary | Modulus | Period |
| 1 | .7906 | .0000 | .7906 | \( \infty \) |
| 2 | -.4266 | .0000 | .4266 | 2.000 |
| 3 | -.0094 | .3092 | .3094 | 3.924 |
| 4 | -.0094 | -.3092 | .3094 | 3.924 |
This section has presented an initial application and test of XYW estimation of a bivariate VAR(2) model with full-monthly and monthly-quarterly samples. The findings are that, compared with ML estimates: (i) full-monthly XYW estimates have similarly low biases and standard deviations; (ii) monthly-quarterly XYW estimates have (a) marginally greater biases, (b) marginally greater standard deviations of coefficients of high-frequency, monthly variables; and, (c) significantly greater standard deviations of coefficients of low-frequency, quarterly variables. Compared with ML estimates, XYW estimates are accurate in terms of (i) to (iib). Recent findings (cf., Clark, 1996) that sampling variability of the optimal weighting matrix, \( Q^* \), can lead to inaccurate small-sample GMM estimates, may provide a clue for reducing standard deviations of XYW estimates of coefficients of low-frequency variable. However, the real test of XYW estimation will be its accuracy when ML estimation is infeasible.
5. CONCLUSION

The paper has derived and applied the XYW method for estimating a VAR model using a sample with mixed-frequency data. However, XYW estimation is applicable to any sample having missing values for any reasons, as long as identifying condition (2.17) holds. Because an XYW estimate is a GMM estimate specialized to VAR estimation with missing data, under standard GMM assumptions it is consistent, asymptotically efficient, and asymptotically normally distributed. Therefore, asymptotically justified statistical inferences with XYW can proceed as in GMM estimation.

The intuition behind XYW estimation is as follows. The usual estimation methods for VAR models such as OLS, GLS, and ML use minimal or exactly identifying sample covariance restrictions on parameters to estimate a model. XYW estimation uses higher-order covariance restrictions to compensate for missing values. As a GMM method, XYW estimation uses more than minimal or overidentifying restrictions on parameters. The YW covariance restrictions on parameters are essentially data determined, being a natural consequence of the form of the VAR model and the choice of lagged variables in the model as instruments.

The application of XYW estimation in section 4 to U.S. data on monthly employment and quarterly GNP shows that it is feasible and can produce estimates of similar accuracy to those produced by ML estimation. Although ML estimation is a reliable method of long standing, it is a nonlinear method which quickly becomes infeasible as the number of parameters increase. By contrast, XYW estimation is a linear method which should be applicable to models with many more parameters. The present application is a start for testing XYW estimation with a variety of data sets. The field is wide, especially if one considers models involving lower-frequency macroeconomic data.
and higher-frequency financial data.

XYW estimation should also be investigated for its small-sample properties because recent simulation studies have cast some doubts on the accuracy of small-sample GMM estimates. Results range from Fuhrer, Moore, and Schuh's (1995) rejection of GMM estimates of an inventory model as unacceptably inaccurate compared with ML estimates, to Clark's (1996) finding that GMM estimates of a multi-factor dynamic model are only slightly more biased and dispersed than ML estimates. Also, Cadzow's (1982) finding that overidentifying YW restrictions significantly help to improve a rational spectrum (or autoregressive moving-average) model's fit of data generated by an unrelated (sinusoid plus white-noise) process suggests it would be worth systematically investigating XYW estimation of misspecified models in full or missing-data samples.
REFERENCES


