WEIGHTED-COVARIANCE FACTOR DECOMPOSITION OF VARMA MODELS APPLIED TO FORECASTING QUARTERLY U.S. GDP AT MONTHLY INTERVALS*

Peter A. Zadrozny**
Bureau of Labor Statistics
2 Massachusetts Ave., NE, Room 3105
Washington, DC 20212
e-mail: zadrozny.peter@bls.gov

Baoline Chen
Bureau of Economic Analysis
1441 L Street, NW
Washington, DC 20230
e-mail: baoline.chen@bea.gov

May 18, 2008

Key words: principal-components-type decomposition of time-series models

JEL codes: C33, C53

******* INCOMPLETE: DO NOT CIRCULATE OR QUOTE WITHOUT PERMISSION *******

ABSTRACT

We develop and apply a method, called weighted-covariance factor decomposition (WCD), for reducing large estimated vector autoregressive moving-average (VARMA) data models of many "important" and "unimportant" variables to smaller VARMA-factor models of "important" variables and significant factors. WCD has four particularly notable features, compared to frequently used principal components decomposition, for developing parsimonious dynamic models: (1) WCD reduces larger VARMA-data models of "important" and "unimportant" variables to smaller VARMA-factor models of "important" variables, while still accounting for all significant covariances between "important" and "unimportant" variables; (2) WCD allows any mixture of stationary and nonstationary variables; (3) WCD produces factors, which can be used to estimate VARMA-factor models, but more directly reduces VARMA-data models to VARMA-factor models; and, (4) WCD leads to a model-based asymptotic statistical test for the number of significant factors. We illustrate WCD with U.S. monthly indicators (4 coincident, 10 leading) and quarterly real GDP. We estimate 4 monthly VARMA-data models of 5 and 11 variables, in log and percentage-growth form; we apply WCD to the 4 data models; we test each data model for the number of significant factors; we reduce each data model to a significant-factor model; and, we use the data and factor models to compute out-of-sample monthly GDP forecasts and evaluate their accuracy. The application's main conclusion is that WCD can reduce moderately large VARMA-data models of "important" GDP and up to 10 "unimportant" indicators to small univariate-ARMA-factor models of GDP which forecast GDP almost as accurately as the larger data models.

---

*Earlier versions of the material here were presented at the Society for Computational Economics (Aix-en-Provence, June 2002, and Seattle, July 2003), the NSF-NBER Time Series Meeting (Philadelphia, September 2002), the Workshop on Forecasting Techniques of the European Central Bank (Frankfurt, December 2002), and the Econometric Society (Evanston, June 2003). The paper represents the authors' views and does not represent any official positions of the Bureau of Economic Analysis or the Bureau of Labor Statistics. We thank Ataman Ozyildirim for providing data from the Conference Board.

**Affiliated as Research Fellow with the Center for Financial Studies (CFS), Goethe University, Frankfurt, Germany, and with the Center for Economic Studies and Ifo Institute for Economic Research (CESifo), Munich, Germany.

1. Introduction.

Parsimony or minimizing the "size" of a model, without compromising its ability to fit data, is a major goal when estimating a model. Parsimony is often interpreted as seeking a model with the fewest number of estimated parameters and this objective is often furthered by choosing a model with a lowest information criterion (Akaike, 1973; Schwarz, 1978; Hannan and Quinn, 1979). Accordingly, because models with fewer variables tend to have fewer parameters, preferring models with fewer variables furthers parsimony. In vector autoregressive moving-average (VARMA) modelling of time-series data, parsimony means dropping insignificant coefficients of lagged variables or of lagged disturbances or zeroing out insignificant disturbance covariances. In state-space modelling of time-series data, parsimony means choosing models with the fewest state variables, which is called minimum realization (MR) and was ported from engineering to econometrics (Aoki, 1983, 1987; Mittnik, 1990a-b, 1992; Kapetanios and Marcellino, 2004).

Usually a model's subject matter suggests which variables should be included in it. However, often the subject matter is unclear about which variables are essential or "important" and which variables are inessential or "unimportant". To avoid the risk of excluding possibly essential variables, investigators often include inessential variables. Thus, models often have too many variables and there is no general agreement on a systematic subject-matter-free method for reducing surely too many variables to only essential ones. The present weighted-covariance factor decomposition (WCD) method offers one possible solution to this problem by weighting variables in a model according to their "importance". The idea is to start with a so-called "data model" of "equally weighted" subject-matter-given variables, estimate the model using some conventional estimation method, and, then, use WCD to reduce the data model to a "factor model" of "important" weighted variables and significant factors. By classifying given variables as "important" or "unimportant" and by reducing a data model to a factor model, WCD, in effect, removes redundant variables.

In the paper, we derive WCD, describe a method for computing WCD, state an asymptotic statistical test based on WCD for the number of significant factors, derive a method for reducing a VARMA-data model to a VARMA-significant-factor model, and apply these steps to VARMA-data models estimated using actual (nonsimulated) U.S. monthly-indicator and quarterly-GDP data from January 1959 to June 2002. Using Kalman-filtering-based maximum likelihood estimation (MLE; Zadrozny, 1990), we estimate 4 monthly VARMA-data models of log and percentage-growth forms of 4 monthly coincident indicators, 10 monthly leading indicators, and quarterly real GDP. Although we also report decompositions of estimated data models based on equal weights on all variables, to evaluate WCD in terms of its ability to help develop models for forecasting GDP, in the application we focus primarily on decompositions of estimated data models based on GDP receiving unit weight and all the other indicator variables receiving zero weight.

Using the estimated data models and the derived factor models, we compute out-of-estimation-sample monthly forecasts of GDP and evaluate the forecasts' accuracies in terms of root mean-squared errors (RMSE). We produce and evaluate GDP forecasts mainly as a natural way of evaluating WCD and advocate WCD as a general tool for analyzing multivariate time-series models, not just as a tool for producing better forecasts. The application has the particularly interesting result that small high-frequency (monthly) factor models derived from larger high-frequency estimated data models, estimated using the monthly-quarterly data in their given mixed-frequency form, can produce high-frequency forecasts of "important" low-frequency GDP which are almost as accurate as forecasts made directly with the larger data models.

Hotelling (1933) introduced the principal components decomposition (PCD; Anderson, 1984, ch. 11). Let $C_k = E(y_t-\mu)(y_{t-k}-\mu)^T$, for k = 0, 1, 2, ..., denote the true covariance matrix of a vector of variables, y_t, and it's k-period lag, y_{t-k}, and let $\hat{C}_k$ denote a data-based estimate of C_k. If PCD uses only $\hat{C}_0$ and not also $\{\hat{C}_k\}_{k=1}^K$, as is often the case, then, PCD is static. Angelini, Henry, and Mestre (2001), Bai and Ng (2002), and Stock and Watson (1989, 1998, 2002) used only $\hat{C}_0$ and, in this sense, produced static PCDs. Forni and Reichlin (1998) and Forni, Hallin, Lippi, and Reichlin (2000) used Fourier transformed $\{\hat{C}_k\}_{k=0}^K$ and, thus, produced dynamic PCDs. Similarly, here we use $\{\hat{C}_0\}_{k=0}^K$ and produce dynamic decompositions.

Earlier estimation of dynamic factor models (Sargent and Sims, 1977; Geweke and Singleton, 1981) was restricted to MLE. Because MLE is computationally very demanding, the models had few variables, few factors, and few estimated parameters and the work remained outside the econometric mainstream (Connor and Korajczyk, 1993). However, by the early 1990s, huge reductions in computing speed, cost, and effort and a shift of interest to much more easily computed PCDs, caused PCDs to be used to develop parsimonious estimated dynamic macroeconomic and financial forecasting models. PCD began to be used to reduce from several hundred real (i.e., observed) financial and macroeconomic variables (Stock and Watson, 1989, 2002a-b; Forni et al., 1998, 2000; Angelini et al., 2001) to several thousand simulated variables (Bai and Ng, 2002), to a few ($\leq$ 5) factors.

PCD is unrestrictive because just to compute it requires no assumptions on a model, just that the sample covariance matrix being decomposed is positive definite, which occurs in practice with probability one if the sample has more sampling periods than variables and none of the variables exactly satisfy any linear equations. However, to support consistency of PCD for estimating the so-called approximate factor model (AFM), it has become customary to make relatively weak assumptions on a background AFM (Bai and Ng, 2002; Stock and Watson, 2002a-b), although this does not affect PCD computations. These assumptions usually restrict the AFM when the number of variables and sampling periods go to infinity. Although the classical approach to PCD (Anderson, 1963; 1984, ch. 11), which we follow here, also makes relatively weak assumptions about what happens when the number of sampling periods goes to infinity, it considers the number of variables fixed. In this regard, we adapt Anderson's asymptotic statistical test for the number of significant factors to the present setting. Because the adapted test is similarly based on relatively weak assumptions, it should be widely applicable.

WCD has three additional notable features, aside from having importance weights, being dynamic, and suggesting a test for significant factors.

1. Stationarity and invertibility. WCD applies to a data model whether it is stationary or nonstationary in any way, with a single or multiple trends, or is invertible or noninvertible (borderline or unit-root noninvertible, because a nonborderline noninvertible model can always be converted to an invertible model). This is because WCD uses a model's impulse-response or Wold-infinite-moving-average coefficients from 0 to h-1, where h denotes the finite number of chosen forecast periods, and these coefficients can be computed equally well regardless whether the data model is stationary, nonstationary, invertible, or noninvertible.

However, if some AR roots are nonstationary or near-nonstationary and h is sufficiently large, then, WCD will tend to "latch onto" the largest AR root and account for nearly 100% of weighted forecast-error covariances with one factor, a result which is often uninformative. In such cases, we can filter out nonstationary or near-nonstationary AR roots before computing WCD. However, the AR roots would be filtered out only to compute WCD and would not be removed from the data model before its reduction to a factor model, so that, depending on variable weightings, the factor model would inherit relevant nonstationary or near-nonstationary roots from the data model.

2. Factors and Factor Models. PCD reduces data to factors but does not reduce a data model to a factor model. PCD factors are often (Stock and Watson, 2002a-b) used to estimate VAR equations, called a dynamic factor model, for forecasting the variables of primary interest such as GDP. However, WCD does this more directly and in a more unified fashion, by reducing an estimated dynamic data model to a dynamic factor model.

3. Time Perspective. PCD depends on C_0 existing, which in turn depends on the data generating process being stationary. Thus, PCD should be applied only to stationary data. For example, Stock and Watson (2002b) first-difference time-series data to stationary form if their originally given values are nonstationary. Because C_0 is also the infinite-horizon forecast-error covariance matrix of stationary variables, PCD could be said to take an infinite-horizon time perspective. By contrast, because WCD decomposes the (weighted) covariance matrix of h-period-ahead forecast errors, WCD could be said to take a finite h-period-ahead time perspective.

Although monthly-quarterly frequencies are an interesting aspect of the present application and mixed frequencies occur almost always in applications which use both financial and macroeconomic time-series data, mixed frequencies are not inherent to WCD. Kalman-filtering-based MLE (Zadrozny, 1990) can handle mixed frequencies but, because of computational constraints, can be applied effectively only to models with limited numbers of variables and lags. The extended Yule-Walker (XYW) method (Chen and Zadrozny, 1998) is more likely to be successful in estimating a large VAR model with mixed-frequency data. XYW is a 2-step linear optimal instrumental variable method (Hansen, 1982; Stoica and Söderström, 1983) which efficiently estimates a VAR model when data have missing values, including mixed frequencies. However, unlike maximum likelihood estimation of a VARMA model using the Kalman filter, XYW cannot estimate a model with an MA part.

WCD also provides a purely-data-based and economic-theory-free variance decomposition, but this aspect will not be emphasized here. Sims (1980a-b) advocated estimating VAR models using loose smoothness restrictions on parameters instead of tight economic restrictions and computing variance decompositions of the estimated models in order to judge the explanatory power of one variable on another. Initially, Sims favored basing variance decompositions on Cholesky decomposition, a purely-data-based decomposition often having tenuous economic justification. Following Cooley and Leroy's (1985) critique, Bernanke (1986), Sims (1986), and most others now base variance decompositions on structural identifications. Being based on a presumably well-fitting data model, yet being economic-theory-free, WCD could be used as an exploratory data-based variance decomposition, prior to a more conclusive structural decomposition.

Although the literature cited above and the application below are based on sampling estimation methods, there are factor decompositions of time series based on Bayesian estimation methods (Otrok and Whiteman, 1998; Kim and Nelson, 1999; Aguilar and West, 2000). The statistics literature contains numerous factor decompositions of time series under the rubrics canonical analysis and reduced rank regression (Box and Tiao, 1977; Ahn and Reinsel, 1988; Deistler and Hamman, 2005). In sum, numerous other methods for decomposing time-series data to factors and data models to factor models have been considered.

The paper proceeds as follows. Section 2 defines WCD. Section 3 discusses two eigenvalue-decomposition methods for computing WCD. Section 4 adapts to the present setting Anderson's (1984, pp. 473-475) asymptotic statistical test for the number of significant factors. Section 5 discusses computing factors, reducing a data model to a significant-factor model in explicit and implicit forms, and computing forecasts made using a factor model. Section 6 applies WCD to ML estimated VARMA-data models of monthly indicators and quarterly GDP. Section 7 concludes the paper.

2. Defining WCD.

Let y_t denote an n×1 vector of sample-mean-adjusted variables observed in periods t = 1, ..., T and presumed to be generated by the VARMA data model

$$(2.1) \qquad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \xi_t + B_1 \xi_{t-1} + \dots + B_q \xi_{t-q},$$

where ξ_t is an n×1 vector of unobserved innovations, distributed normally, identically, independently, with zero means and positive definite covariance matrix Σ_ξ (> 0) or $\xi_t \sim NIID(0, \Sigma_\xi)$. $\Sigma_\xi > 0$ is necessary for data model (2.1) to have a unique WCD; the other two conditions are stated at the end of this section. Model (2.1) is stated more concisely in terms of lag operator L as $A(L)y_t = B(L)\xi_t$, where $A(L) = I_n - A_1L - \dots - A_pL^p$, $B(L) = I_n + B_1L + \dots + B_qL^q$, and I_n denotes the n×n identity matrix. Although equation (2.1) is a reduced form, whose parameters might be restricted further in terms of fewer structural parameters, to compute WCD we need only the reduced form.

In general, y_t = vector of data means + a vector of deviations from the means. The means are the nonstochastic and (unconditionally) predictable parts of y_t and the deviations from means are the stochastic and unpredictable parts. In general, the means could be nonconstant and could include time trends, calendar effects, and other regression effects. We abstract from the means because WCD pertains only to the stochastic deviations from means. In the application, for simplicity, we assume the means are constant, estimate them as sample means, and remove them at the start.

To ensure accurate computations, model (2.1) should first be normalized as follows. There are two cases to consider: (i) data on y_t are available and model (2.1) is estimated; (ii) data on y_t are unavailable but an estimate of model (2.1) is available. Let S denote the m×m diagonal matrix defined by $S = \mathrm{diag}[s_1, \ldots, s_n]$, where the s_i are positive normalizing values. In case (i), data should be normalized as (a) $y_t := S^{-1}y_t$ prior to estimating the data model, where := denotes assignment and s_i = sample standard deviation of variable i. In case (ii), estimated parameters should be normalized as (b) $A(L) := S^{-1}A(L)S$, $B(L) := S^{-1}B(L)S$, and $\Sigma_\xi := S^{-1}\Sigma_\xi S^{-1}$, where s_i = estimated standard deviation of disturbance i. The normalizations promote accurate computations, because they eliminate large differences in magnitudes of numbers due to large differences in units of measurement of different variables. In particular, normalization (b) puts disturbances in common units of standard deviations and makes Σ_ξ the correlation matrix of ξ_t.

We define characteristic AR and MA roots as follows. Model (2.1) is stationary if and only if (iff) the absolute characteristic AR roots are < 1, namely, iff $\det[I_n\lambda^p - A_1\lambda^{p-1} - \ldots - A_{p-1}\lambda - A_p] = 0$ implies $|\lambda| < 1$. Model (2.1) is invertible iff the absolute characteristic MA roots are < 1, namely, iff $\det[I_n\lambda^q + B_1\lambda^{q-1} + \ldots + B_{q-1}\lambda + B_q] = 0$ implies $|\lambda| < 1$. In econometrics, characteristic roots are usually defined as reciprocals of the present roots, so that a VARMA model is stationary and invertible iff its absolute characteristic AR and MA roots are > 1.

If data model (2.1) is stationary, it has the Wold infinite moving-average representation

$$y_t = \Psi(L)\xi_t = \left(\textstyle\sum_{i=0}^{\infty} \Psi_i L^i\right)\xi_t = \textstyle\sum_{i=0}^{\infty} \Psi_i \xi_{t-i}, \tag{2.2}$$

where $\Psi(L) = A(L)^{-1}B(L)$. The Wold coefficients, $\{\Psi_i\}_{i=0}^{\infty}$, are also the infinite impulse-response sequence of y_t: column j of Ψ_i measures the response of y_{t+i} to a one-period one-standard-deviation increase in disturbance j of ξ_t. Regardless whether the model is stationary or invertible, the finite impulse-response sequence, $\{\Psi_i\}_{i=0}^{h-1}$, can be computed by iterating on

$$(2.3)\qquad \Psi_i = \sum_{j=1}^{\min(i,p)} A_j \Psi_{i-j} + B_i ,$$

for i = 1, ..., h-1, starting with $\Psi_0 = I_n$, such that $B_i = 0$ for i > q.

In nonstationary or near-nonstationary cases, WCD could latch onto the largest nonstationary or near-nonstationary AR root (the latter being defined, say, as $\lambda_i > .99$), especially if h is large, and account for nearly 100% of weighted covariances with one factor, a result which is often uninformative. In such case, we could filter out nonstationary or near-nonstationary AR roots before computing WCD. Let $\{\lambda_i\}_{i=1}^{\upsilon}$ denote the ν nonstationary or near-nonstationary AR roots of data model (2.1). We would multiply the Wold representation by $\lambda(L) = (1-\lambda_1 L)\cdots(1-\lambda_\nu L)$ and obtain the filtered Wold representation, $\tilde{y}_t = \tilde{\Psi}(L)\xi_t$, where $\tilde{y}_t = \lambda(L)y_t$ and $\tilde{\Psi}(L) = \lambda(L)\Psi(L)$. Then, we would compute WCD for $\tilde{\Psi}(L)$ and Σ_ξ. The roots would be filtered out only to compute WCD and would not be removed from the data or the factor models.

Let $\eta_{ht} = y_{t+h} - E_t y_{t+h}$ denote the n×1 vector of errors from forecasting y_{t+h} at time t, for forecast $E_t y_{t+h}$ and finite forecast horizon h = 1, 2, ... In terms of innovations, the forecast errors are

$$(2.4)\qquad \eta_{ht} = \sum_{i=0}^{h-1} \Psi_i \xi_{t+h-i} ,$$

and have the covariance matrix

$$(2.5)\qquad \Gamma_h = E\,\eta_{ht}\eta_{ht}^{T} = \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^{T} ,$$

where superscript T denotes transposition. We propose evaluating WCD in terms of the forecast accuracy of the r×1 vector of "important" weighted variables, $w_t = Wy_t$, where W is an r×n matrix of chosen weights and r = rank(W) ≤ n. Then, $\Omega = W^T W$ is an n×n positive semi-definite ($\Omega \geq 0$) symmetric weighting matrix and v =

$E\eta_{ht}^{T}\Omega\eta_{ht} = tr[\Omega\Gamma_h]$ is the expected weighted h-period-ahead squared forecast error, where tr[·] denotes the trace of a matrix. Inserting equation (2.5) into v, we obtain

$$v = tr[\Omega \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]. \tag{2.6}$$

Usually, Ω is positive semi-definite and singular. WCD has a unique solution and can be computed using either method 1 or 2 described in section 3.2. However, preferred method 2 could fail if Ω is singular. If so, we may replace Ω with $\Omega + \beta I_n$, where β is a small positive scalar, large enough so that Ω is numerically positive definite, but small enough so that WCD is practically unaffected.

The following three examples illustrate singular and nonsingular weightings:

1. Minimum Portfolio Variance. Let $W = (w_1, \ldots, w_n)$ denote a 1×n vector of portfolio allocation weights. Every variable in a portfolio receives a positive or negative weight, $w_i \neq 0$, according to whether it is an asset or liability. Variables outside the portfolio that help to forecast it are included in the data model but have zero weight in W. $\Omega = W^T W$ has rank 1 and is singular for $n > 1$. WCD decomposes portfolio forecast-error variance. If $\Sigma_\xi > 0$, then, $\Gamma_h > 0$ and $W = (\bar{e}^T \Gamma_h^{-1} \bar{e})^{-1} \bar{e}^T \Gamma_h^{-1}$ minimizes portfolio variance with respect to W, subject to $W^T \bar{e} = 1$, where $\bar{e} = (1, \ldots, 1)^T$ is the n×1 vector of ones.

2. Forecasting. Let $W = e_i^T = (0, \ldots, 1, \ldots, 0)$ denote an n×1 "importance weight" vector with 1 in position i and 0 elsewhere. As in example one, $\Omega = W^T W$ has rank 1 and is nonsingular iff $n = 1$. Only the "important" variable i receives nonzero weight in WCD and is forecast. The remaining "unimportant" variables ensure that the data model accounts for all significant interactions among "important" and "unimportant" variables. More generally, $\Omega = \sum_{j=1}^{J} e_{i_j} e_{i_j}^T$, where $\{i_j\}_{j=1}^{J}$ denotes a nonrepeating subset of J integers from $\{1, \ldots, n\}$. In this case, only the "important" variables $\{i_j\}_{j=1}^{J}$ receive nonzero weight in WCD and are forecast. Because the scale of Ω is irrelevant in WCD, the importance weights could be any nonzero numbers.

3. Principal Components. Let $\Omega = \sum_{i=1}^{n} e_i e_i^T = I_n$, the nonsingular n×n identity matrix. In addition, if all variables are stationary and the forecast

horizon h is large (strictly, $h = \infty$), then, v = sum of variances of all variables is decomposed, WCD corresponds to PCD, and WCD provides insight about how the variables are jointly generated by factors.

WCD produces the $n \times n$ decomposition matrix R, such that $RR^T = \Sigma_\xi$, analogous to the factor-loading matrix in PCD. For such an R, $\varepsilon_t = R^{-1}\xi_t$ is distributed orthonormally, namely, $E\varepsilon_t\varepsilon_t^T = I_n$. If Σ_ξ is positive definite and nonsingular, R^{-1} exists. Disturbances ε_t are usually called "structural" because they vary mutually independently, but consistently with the innovation covariance matrix, Σ_ξ. Let r_i denote column i of R and v_i denote weighted h-step-ahead forecast-error covariances accounted for by structural disturbance i. Then, equation (2.6) implies $v = \sum_{i=1}^{n} v_i$, where

$$v_i = r_i^T Q r_i, \tag{2.7}$$

for i = 1, ..., n, and

$$Q = \sum_{i=0}^{h-1} \Psi_i^T \Omega \Psi_i \; . \tag{2.8}$$

We define WCD recursively. For i = 1, ..., n, let $X_i = [r_1, \ldots, r_i]$, let $\Sigma_i = \Sigma_\xi - X_{i-1}X_{i-1}^T$, starting with $X_0 = 0_{n\times n}$, and let $Y_i = [r_{i+1}, \ldots, r_n]$ be a matrix of slack variables which ensures that $r_i r_i^T + Y_i Y_i^T = \Sigma_i$. For each i = 1, ..., n, we want to maximize $\varphi_i = r_i^T Q r_i / v$. However, because the denominator v is independent of r_i, it suffices to maximize v_i.

First, given $\Sigma_1 = \Sigma_\xi$ (and Q), we maximize v_1 with respect to r_1 and Y_1, subject to $r_1 r_1^T + Y_1 Y_1^T = \Sigma_1$, by eliminating Y_1 and solving for r_1 and v_1. Then, given $\Sigma_2 = \Sigma_\xi - X_1 X_1^T$, for $X_1 = r_1$, we maximize v_2 with respect to r_2 and Y_2, subject to $r_2 r_2^T + Y_2 Y_2^T = \Sigma_2$, by eliminating Y_2 and solving for r_2 and v_2. Then, given $\Sigma_3 = \Sigma_\xi - X_2 X_2^T$, for $X_2 = [r_1, r_2]$, we maximize v_3 with respect to r_3 and Y_3, subject to $r_3 r_3^T + Y_3 Y_3^T = \Sigma_3$, by eliminating Y_3 and solving for r_3 and v_3. Continuing like this, we determine $X_{n-1} = [r_1, \ldots, r_{n-1}]$. Finally, given $\Sigma_n = \Sigma_\xi - X_{n-1}X_{n-1}^T$, for $X_{n-1} = [r_1, \ldots, r_{n-1}]$, we determine r_n such that $r_n r_n^T = \Sigma_n$. In this n-th case, $Y_n = 0_{n\times n}$. The result is the decomposition matrix $R = [r_1, \ldots, r_n]$, such that $RR^T = \Sigma_\xi$, $v_1 \geq \ldots \geq v_n \geq 0$ and $v_1 + \ldots + v_n = v$.

A unique WCD exists if $\Sigma_\xi Q$ is positive definite (hence, nonsingular) and has distinct eigenvalues. Positive definite means that $\Sigma_\xi Q$ has real and positive eigenvalues and real eigenvectors. If, as usual, $\Sigma_\xi Q$ is asymmetric, then, the eigenvectors are generally not orthogonal. A unique WCD could exist, but could fail to be computed, if $\Sigma_\xi Q$ is positive semi-definite and singular or has repeated eigenvalues. Therefore, we assume (i) Σ_ξ is positive definite, (ii) $Q = \sum_{i=0}^{h-1} \Psi_i^T \Omega \Psi_i$ is positive definite, and (iii) $\Sigma_\xi Q$ has distinct eigenvalues.

In practice, assumptions (i) to (iii) are not very restrictive. If the data have no identities, the estimated Σ_ξ will be positive definite. As discussed above, replacing Ω with $\Omega + \delta I_n$, where δ is small and positive, makes Ω and Q positive definite. We can show that assumptions (i) and (ii) imply that $\Sigma_\xi Q$ is positive definite. For positive definite $\Sigma_\xi Q$, we usually obtain WCD close to double precision accuracy ($\cong 10^{-14}$). Computed eigenvalues of $\Sigma_\xi Q$ have been distinct to within the same accuracy. Finally, $\Sigma_\xi Q$ is diagonalizable, if its eigenvalues are distinct.

If the data model is stationary, $\sum_{i=0}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T = \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T + \sum_{i=h}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T$ or the unconditional data covariances = the conditional h-period-ahead forecast-error covariances + the unconditional h-period-ahead forecast covariances, such that all of the sums are finite. WCD decomposes $\mathrm{tr}[\Omega \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]$ or the conditional forecast-error covariances (weighted by Ω, for any h, any VARMA model, subject to $RR^T = \Sigma_\xi$) and Box and Tiao (1977) decompose $\mathrm{tr}\{[\sum_{i=0}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T]^{-1} \times [\sum_{i=1}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T]\}$ or the unconditional forecast covariances (normalized, unweighted, for h = 1, any VAR model, not subject to any side restrictions). Although these objectives might seem at odds, WCD applied to the unnormalized and unweighted objectives $\mathrm{tr}[\sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]$ and $\mathrm{tr}[\sum_{i=1}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]$, for $h > 1$, often produces a similar number of significant factors and similar decomposition or factor-loading matrices, R. Thus, although decomposing $\mathrm{tr}[\Omega \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]$ makes sense for conditional h-period-ahead forecasting, the econometrics and statistics literatures contain numerous alternative related objectives.

3. Computing WCD.

In this section, following the definition of WCD in section 2, we state first-order conditions for WCD and two methods for computing WCD. The first-order conditions and the two computational methods have two corresponding parts: first, for i = 1, ..., n-1 and, then, for i = n. Method 2 simplifies the first part of method 1.

3.1 First-Order Conditions.

For i = 1, ..., n-1, the Lagrangian for maximizing v_i with respect to r_i and Y_i, subject to $r_i r_i^T + Y_i Y_i^T = \Sigma_i$, for given Q and Σ_i, is

$$(3.1) \qquad \mathit{L}_i = r_i^T Q r_i + \mathrm{tr}\{\Xi_i[\Sigma_i - r_i r_i^T - Y_i Y_i^T]\},$$

where Ξ_i is an n×n symmetric matrix of Lagrange multipliers. Then, the first-order necessary conditions, obtained by differentiating L_i with respect to r_i, Y_i, and Ξ_i and setting the derivatives to zero are

$$(3.2) \qquad (Q - \Xi_i) r_i = 0_{n\times 1},$$

$$(3.3) \qquad \Xi_i Y_i = 0_{n\times(n-i)},$$

$$(3.4) \qquad r_i r_i^T + Y_i Y_i^T = \Sigma_i,$$

for i = 1, ..., n-1, where $0_{k\times\ell}$ denotes the $k\times\ell$ zero matrix. We eliminate Y_i from equation (3.4) by postmultiplying the equation by Ξ_i and obtain

$$(3.5) \qquad (\Sigma_i - r_i r_i^T)\Xi_i = 0_{n\times n},$$

for i = 1, ..., n-1. Then, to eliminate Ξ_i from equation (3.5), we postmultiply the equation by r_i, use equation (3.2) to eliminate $\Xi_i r_i$, use $v_i = r_i^T Q r_i$, and obtain

$$(3.6) \qquad (\Sigma_i Q - v_i I_n) r_i = 0_{n\times 1},$$

for $i = 1, \ldots, n-1$. Equation (3.6) indicates n-1 eigenvalue problems, with eigenvalues v_i and eigenvectors r_i, which we solve for as follows.

3.2. Computational Methods 1 and 2.

Because $\Sigma_i Q$ is positive semi-definite (despite being generally asymmetric), its eigenvalues are real and nonnegative and, because its rank is n-i+1, its largest n-i+1 eigenvalues are positive and its smallest i-1 eigenvalues are zero. Because v_i is being maximized, for $i = 1, \ldots, n-1$, we set

$$(3.7) \qquad v_i = \lambda_i,$$

where λ_i denotes the largest positive eigenvalue of $\Sigma_i Q$ and set r_i to its rescaled eigenvector, z_i, as

$$(3.8) \qquad r_i = \sqrt{\frac{\lambda_i}{z_i^T Q z_i}}\, z_i,$$

where $z_i^T z_i = 1$.

We could include $i = n$ in the above first step, although this case would have to be treated differently anyway, because the eigenvector, r_n, must satisfy the covariance constraint, $r_n r_n^T = \Sigma_n$, with no remaining slack, $Y_n = 0$. To compute r_n, we postmultiply $r_n r_n^T = \Sigma_n$ by r_n and obtain the eigenvalue problem

$$(3.9) \qquad (\Sigma_n - v_n I_n) r_n = 0,$$

where $v_n = r_n^T r_n$. Because v_n is being maximized, we set

$$(3.10) \qquad v_n = \lambda_n,$$

where λ_n denotes the largest (and only positive) eigenvalue of Σ_n and set r_n to its rescaled eigenvector, z_n, as

$$(3.11) \qquad r_n = (\sqrt{\lambda_n})\, z_n,$$

where $z_n^T z_n = 1$.

Thus, we have method 1 for computing WCD: first, for i = 1, ..., n-1, set v_i to the largest eigenvalue λ_i of $\Sigma_i Q$ and set r_i to its rescaled eigenvector z_i, according to equations (3.7)-(3.8); then, set v_n to the largest eigenvalue, λ_n, of Σ_n and set r_n to its rescaled eigenvector, z_n, according to equations (3.10)-(3.11). The following method 2 simplifies the first part of method 1. It can be shown that, in theory, methods 1 and 2 compute identical WCDs. Method 2 for computing WCD is: first, for i = 1, ..., n-1, set the v_i's to the n-1 largest eigenvalues, λ_i, of $\Sigma_\xi Q$ and set the r_i's to their rescaled eigenvectors, z_i, according to equations (3.7)-(3.8); then, set v_n to the largest eigenvalue λ_n of Σ_n and set r_n to its rescaled eigenvector z_n, according to equations (3.10)-(3.11). Both methods require about the same amount of computational work as PCD. We used method 2 in the application in section 6.

4. Testing for Significant Factors.

For a given tolerance, $\rho \in (0,1)$, let H_m denote the hypothesis that the first m factors are significant in the sense that they account for at least $1-\rho$ of weighted covariances, $v = v_1 + \ldots + v_n$, or equivalently, that the last n-m factors account for no more than ρ of v. Concisely, H_m says that

$$\text{(4.1)} \qquad \delta_m = \bar{\rho}^T \bar{v} = -\rho \sum_{i=1}^{m} v_i + (1-\rho) \sum_{i=m+1}^{n} v_i \le 0,$$

where $\bar{\rho} = (-\rho, \ldots, -\rho, 1-\rho, \ldots, 1-\rho)^T$ and $\bar{v} = (v_1, \ldots, v_n)^T$ are n×1 vectors.

Consider the following testing sequence for the number of significant factors. Start with m = 1. For m = 1, test H_m; if H_m is not rejected, accept m = 1 as the number of significant factors and stop testing. Otherwise, for m = 2, test H_m; if H_m is not rejected, accept m = 2 as the number of significant factors and stop testing. Otherwise, continue like this until possibly reaching m = n-1; if H_m is not rejected for m = n-1, accept m = n-1 as the number of significant factors and stop testing. Otherwise, accept m = n as the number of significant factors. Because all of v is accounted for when m reaches n, the testing sequence is always conclusive.

WCD is, in effect, a nonlinear differentiable function from ϕ to $\bar{v}$, denoted by $\bar{v} = K(\phi)$. If ϕ is considered certain, say, because it is being assumed for a hypothetical model, then, H_m is not rejected iff $\delta_m \le 0$. However, in practice, ϕ is estimated and uncertain, so that H_m must be tested probabilistically. We now extent $\delta_m \le 0$ to probabilistic test (4.3), which, essentially, is Anderson's (1963; 1984, pp. 473-475) asymptotic statistical test adapted to the present setting.

If VARMA data model (2.1) is stationary and invertible and some other conditions hold (Hosoya and Taniguchi, 1982), then $\sqrt{T}(\hat{\phi} - \phi_0) \sim AN(0, \hat{S})$, where vector ϕ collects parameters of the VARMA data model, hat (^) denotes a quantity evaluated at the MLE of ϕ, subscript zero denotes a true value, ~ AN denotes asymptotic normal distribution as $T \to \infty$, and $\hat{S}$ denotes the asymptotic covariance matrix of $\sqrt{T}(\hat{\phi} - \phi_0)$. If $\sqrt{T}(\hat{\phi} - \phi_0) \sim AN(0, \hat{S})$ and further conditions hold (Serfling, 1982, pp. 122-124), then, $\sqrt{T}(\hat{\overline{v}} - \overline{v}_0) \sim AN(0, \nabla\hat{K}\,\hat{S}\,\nabla\hat{K}^T)$, where $\nabla\hat{K}$ denotes the Jacobian matrix of first-partial derivatives of $K(\phi)$ evaluated at $\hat{\phi}$, so that, because $\overline{\rho}$ is constant,

$$(4.2) \qquad \sqrt{T}(\hat{\delta}_m - \delta_{m0}) \sim AN(0, \hat{\sigma}^2_{\delta_m}),$$

where $\hat{\sigma}^2_{\delta_m} = \overline{\rho}^T \nabla\hat{K}\,\hat{S}\,\nabla\hat{K}^T\overline{\rho}$. For given $\hat{\phi}$ and $\hat{S}$, applying the method of matrix differentiation used by Mittnik and Zadrozny (1993), we obtain equations for computing $\nabla\hat{K}$, hence $\hat{\sigma}^2_{\delta_m}$, which are stated in the appendix.

Using equations (4.1)-(4.2) and following usual sampling-theory testing, H_m may be tested as follows. Let $\alpha \in (0,1)$ denote a chosen significance level and let c_α denote a critical value defined by $\mathrm{Prob}[z \le c_\alpha] = 1 - \alpha$, where $z \sim N(0,1)$. Then, for given m, ρ, and α, H_m is not rejected iff

$$(4.3) \qquad \tau_m = \frac{\hat{\delta}_m}{\hat{\sigma}_{\delta_m}} \le c_\alpha.$$

Even if $\hat{\phi}$ is not exactly asymptotically normally distributed, (4.2) can be considered a first-order normal approximation.

5. Computing Factors, Factor Models, and Forecasts.

We now discuss computing factors and reducing a VARMA data model of n variables to a smaller VARMA factor model of r (<< n) weighted variables of interest. In subsection 5.1, we discuss computing factors for an "in-sample" period used to estimate an n-variable data model. The computed factors could be used as observed variables to estimate a smaller (m+r)-variable model of m significant factors and r weighted variables of interest, such that m+r << n, in order to forecast the weighted variables, where << means "much less than." Stock and Watson (1989, 1998, 2002), Angelini et al. (2001), Bai and Ng (2002), and others did this using PCD. In subsection 5.2, we discuss two steps for reducing an n-variable data model to a smaller r-variable factor model. Step one reduces the n-variable VARMA model to an r-variable state-space form, in which state variables are explicit linear combinations of significant factors. Step two eliminates the factors and reduces the explicit state-space form for r weighted variables to an implicit VARMA form. Thus, the factor models in the two steps are called explicit and implicit. Section 6 illustrates reductions of data models of n = 5 and n = 11 variables to implicit factor models of r = 1.

5.1. Computing Factors.

Let f_t denote an n×1 vector of factors defined by $y_t = Rf_t$, where the computation of n×n matrix R is discussed in section 3. In addition to being the WCD matrix based on an estimated data model, R is also the factor loading matrix. Factors are unobserved but can be estimated as follows. Let $\{y_t\}_{t=1}^{N}$ denote the N-period in-sample data. First, if no observations in $\{y_t\}_{t=1}^{N}$ are missing, we can estimate all factors as $\{f_t\}_{t=1}^{N} = \{R^{-1}y_t\}_{t=1}^{N}$, because R is nonsingular (because $RR^T = \Sigma_\xi$ is positive definite). For a particular period t, if any element of y_t is unobserved or missing for any reason (mixed frequencies, missing starting or ending values, randomly missing inside-sample values), then, generally, no element of f_t can be computed as $f_t = R^{-1}y_t$ because generally R is a full matrix with no zero-valued elements. Thus, generally, f_t can be computed as $R^{-1}y_t$ only if every element of y_t is observed. For example, if $\{y_t\}_{t=1}^{N}$ contains monthly and quarterly observations, such that monthly variables are observed every period and quarterly variables are observed only at end-of-quarter months, but there are no missing values for other reasons, then, $\{f_t\}_{t=1}^{N}$ can be estimated as $\{R^{-1}y_t\}_{t=1}^{N}$ only for end-of-quarter periods. Missing values in $\{y_t\}_{t=1}^{N}$ could be estimated before computing $\{f_t\}_{t=1}^{N} = \{R^{-1}y_t\}_{t=1}^{N}$, for example, using a spline smoother. Although such smoothing methods are widely used, numerically reliable, and available in many software packages or easily programmed, they have the disadvantage of being unrelated to the estimated data model, which is avoided when computing f_t by using a Kalman filter or smoother based on the estimated data model.

Let x_t denote the state vector of a state-space representation of estimated data model (2.1), analogous to x_t in state-space representation (5.5)-(5.6), and let $\{x_{t|N}\}_{t=1}^{N} = \{E[x_t|\{y_\tau\}_{\tau=1}^{N}]\}_{t=1}^{N}$ denote Kalman-smoothed estimates of x_t based on estimated data model (2.1) and the in-sample data. Thus, we could estimate all in-sample factors as the smoothed estimates $\{f_{t|N}\}_{t=1}^{N} = \{R^{-1}H_t x_{t|N}\}_{t=1}^{N}$, where H_t is an observation matrix as in observation equation (5.5). The Kalman smoother is a direct extension of the Kalman filter (Anderson and Moore, 1979, ch. 7). Zadrozny (1990) and Mittnik and Zadrozny (2004) used the Kalman filter to estimate monthly VARMA models with monthly and quarterly data and to compute monthly forecasts of the quarterly variables based on the estimated models.

5.2. Computing Explicit and Implicit Significant-Factor Models.

The WCD algorithm in section 3 and the above definition of f_t imply that, like the columns of R, the elements of f_t are ordered from most to least significant. For now, we assume that m factors are significant. In section 5, we describe an asymptotic test of the number of significant factors. In the present subsection, we describe methods for deriving significant-factor models in explicit state-space and implicit VARMA forms. If m of the n factors are significant, we partition $f_t = (f_{1t}^T,\ f_{2t}^T)^T$, where f_{1t} and f_{2t} denote m×1 and (n-m)×1 vectors of significant and insignificant factors, and conformably partition $\varepsilon_t = (\varepsilon_{1t}^T,\ \varepsilon_{2t}^T)^T$, $R = [R_1, R_2]$, and $R^{-1} = [R_1^{-T},\ R_2^{-T}]^T$, where R_1 and R_2 are left n×m and right n×(n-m) blocks of R, R_1^{-1} and R_2^{-1} are top m×n and bottom (n-m)×n blocks of R^{-1}, and superscript -T denotes a transposed inverse. Generally, R_i^{-1} is not an inverse of R_i, only if $R = R_1$ or R_2.

Because R is nonsingular, we can restate data model (2.1) as the partitioned VARMA n-factor model

$$(5.1) \qquad \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = \begin{bmatrix} R_1^{-1}A_1R_1 & R_1^{-1}A_1R_2 \\ R_2^{-1}A_1R_1 & R_2^{-1}A_1R_2 \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \end{bmatrix} + \ldots + \begin{bmatrix} R_1^{-1}A_pR_1 & R_1^{-1}A_pR_2 \\ R_2^{-1}A_pR_1 & R_2^{-1}A_pR_2 \end{bmatrix} \begin{bmatrix} f_{1,t-p} \\ f_{2,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$+\begin{bmatrix} R_1^{-1}B_1R_1 & R_1^{-1}B_1R_2 \\ R_2^{-1}B_1R_1 & R_2^{-1}B_1R_2 \end{bmatrix}\begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \ldots + \begin{bmatrix} R_1^{-1}B_qR_1 & R_1^{-1}B_qR_2 \\ R_2^{-1}B_qR_1 & R_2^{-1}B_qR_2 \end{bmatrix}\begin{bmatrix} \varepsilon_{1,t-q} \\ \varepsilon_{2,t-q} \end{bmatrix}.$$

To see the consequences of treating f_{2t} as insignificant in equation (5.1), first, $\Sigma_\xi = R_1R_1^T + R_2R_2^T$ implies that WCD objective (2.6) partitions as

$$\text{(5.2)} \qquad tr[\Omega\sum_{i=0}^{h-1}\Psi_i\Sigma_\xi\Psi_i^T] = tr[\Omega\sum_{i=0}^{h-1}\Psi_iR_1R_1^T\Psi_i^T] + tr[\Omega\sum_{i=0}^{h-1}\Psi_iR_2R_2^T\Psi_i^T],$$

which is equivalent to $v = \sum_{i=1}^{m} v_i + \sum_{i=m+1}^{n} v_i$. In section 4, we describe an asymptotic statistical test for the number of significant factors, based on equation (5.2). In the test, f_{2t} and ε_{2t} are deemed to insignificantly affect the importance weighted variables, $w_t = Wy_t$, iff the variance of w_t accounted for by ε_{2t} is deemed sufficiently small relative to the whole variance of w_t. If so, we set $R_2 = 0_{n\times(n-m)}$ in the top $m\times1$ block of equation (5.1) for f_{1t}, ignore the $(n-m)\times1$ bottom block of equation (5.1) for f_{2t}, and in observation equation (5.5) treat R_2f_{2t} as a vector of serially-uncorrelated observation errors with negligible variances.

Thus, if f_{2t} and ε_{2t} are considered insignificant, we reduce VARMA n-factor model (5.1) to the VARMA m-factor model

$$\text{(5.3)} \qquad f_{1t} = A_1^*f_{1,t-1} + \ldots + A_p^*f_{1,t-p} + \varepsilon_{1t} + B_1^*\varepsilon_{1,t-1} + \ldots + B_q^*\varepsilon_{1,t-q},$$

where $A_i^* = R_1^{-1}A_iR_1$ for $i = 1, \ldots, p$, $B_j^* = R_1^{-1}B_jR_1$ for $j = 1, \ldots, q$, and $\varepsilon_{1t} \sim NIID(0,I_m)$. The reduction is not an exact mathematical implication, but reflects a decision, based on a statistical test, about the number of significant factors.

Significant-factor model (5.3) is stationary iff $\det[I_n\lambda^p - A_1^*\lambda^{p-1} - \ldots - A_{p-1}^*\lambda - A_p^*] = 0$ implies $|\lambda| < 1$ or, equivalently, iff the eigenvalues of transition matrix F of state equation (5.6) are less than one in absolute value. Generally, unless all factors are significant, the data and (significant) factor models will have different AR and MA roots, the factor model could be stationary even if the data model is nonstationary, but the factor model is nonstationary only if the data model is nonstationary.

To forecast the vector of importance-weighted variables, $w_t = Wy_t$, we consider the equation which defines the factors, $y_t = R_1f_{1t} + R_2f_{2t}$, as an observation equation in the significant factors, f_{1t},

$$w_t = WR_1f_{1t} + u_t, \tag{5.4}$$

so that $u_t = WR_2f_{2t} = W\sum_{i=0}^{\infty} \Psi_i R_2 \varepsilon_{2,t-i}$ is treated as a vector of serially-uncorrelated observation errors. If f_{2t} and ε_{2t} insignificantly affect w_t, then, Σ_u, the covariance matrix of u_t is small. On the other hand, to ensure that a Kalman filter applied with observation equation (5.4) is accurate, Σ_u should be treated as at least slightly positive definite. If $r > n-m$, then, $WR_2E(f_{2t}f_{2t}^T)R_2^TW^T$ is singular and inappropriate as a choice for Σ_u. Therefore, we treat u_t as distributed $NIID(0,\delta I_r)$, where δ is a small and positive scalar.

We now write VARMA process (5.3) and observation equation (5.4) as a state representation of w_t in the state vector x_t. A state representation consists of an observation equation and a state equation. We write observation equation (5.4) in terms of the state vector x_t as

$$w_t = Hx_t + u_t, \tag{5.5}$$

where $H = [W^TR_1, 0_{r\times m}, \ldots, 0_{r\times m}]$ is an $r\times s$ observation matrix, $x_t = (x_{1t}^T, \ldots, x_{\ell t}^T)^T$ is an $s\times 1$ state vector, $s = m\ell$, $\ell = \max(p,q+1)$, and $u_t \sim NIID(0,\delta I_r)$. Following Ansley and Kohn (1983), a state equation implied by VARMA process (5.3) is

$$x_t = Fx_{t-1} + G\varepsilon_{1t}, \quad F = \begin{bmatrix} A_1^* & I_m & 0_{m\times m} & \cdots & 0_{m\times m} \\ . & 0_{m\times m} & . & \cdots & . \\ . & . & . & \cdots & 0_{m\times m} \\ . & . & . & \cdots & I_m \\ A_\ell^* & 0_{m\times m} & . & \cdots & 0_{m\times m} \end{bmatrix}, \quad G = \begin{bmatrix} I_m \\ B_1^* \\ . \\ . \\ B_\ell^* \end{bmatrix}, \tag{5.6}$$

where A_i^*, B_i^*, and ε_{1t} are defined as in equation (5.3), $A_i^* = 0_{m\times m}$ for $i > p$, and $B_i^* = 0_{m\times m}$ for $i > q$. Recursive substitution shows that, in state equation (5.6), x_t consists of linear combinations of lagged f_1's and ε_1's. H and F are called the observation and transition matrices.

Dropping presumably insignificant u_t from observation equation (5.5) and combining it with the state equation, we obtain the transfer-function representation

$$(5.7) \qquad w_t = H\Phi(L)^{-1}G\varepsilon_{1t},$$

where $\Phi(L) = I_s - FL$. Let λ denote a nonzero complex-valued scalar sufficiently close to zero so that $\Phi(\lambda)$ is nonsingular; let $\phi(\lambda) = \det[\Phi(\lambda)] = 1 + \phi_1\lambda + \ldots + \phi_s\lambda^s$ denote the determinant of $\Phi(\lambda)$; and, let $\Phi^*(\lambda) = \phi(\lambda)\Phi(\lambda)^{-1} = I_s + \Phi_1^*\lambda + \ldots + \Phi_{s-1}^*\lambda^{s-1}$ denote the adjoint matrix of $\Phi(\lambda)$. Fadeev's algorithm (Gantmacher, 1959, pp. 87-89) easily computes the real-valued scalar coefficients of $\phi(\lambda)$ by iterating on

$$(5.8) \qquad \phi_i = -\mathrm{tr}(\Phi_{i-1}^*F)/i,$$

for i = 1, ..., s, and

$$(5.9) \qquad \Phi_i^* = \Phi_{i-1}^*F + \phi_i I_s,$$

for i = 1, ..., s-1, starting with $\Phi_0^* = I_s$.

Thus, transfer function (5.7) implies VARMA(s,s-1) process

$$(5.10) \qquad \phi(L)w_t = \chi_t,$$

where $\chi_t = H\Phi^*(L)G\varepsilon_{1t} = HG\varepsilon_{1t} + H\Phi_1^*G\varepsilon_{1,t-1} + \ldots + H\Phi_{s-1}^*G\varepsilon_{1,t-s+1}$ and the $H\Phi_i^*G$ are r×m coefficient matrices. Equation (5.10) is problemmatical if it has more disturbances than variables, m > r, which is avoided by recasting the equation in the standard VARMA form

$$(5.11) \qquad \phi(L)w_t = \Theta(L)\zeta_t = \Theta_0\zeta_t + \Theta_1\zeta_{t-1} + \ldots + \Theta_{s-1}\zeta_{t-s+1},$$

where the Θ_i are real-valued r×r coefficient matrices, $\Theta_0 = I_r$, and ζ_t is an r×1 disturbance vector ~ NIID$(0,\Sigma_\zeta)$. Generally, (5.11) is useful iff it is invertible, namely, iff $\det[\lambda^{s-1}\Theta(\lambda)] = 0$ implies $|\lambda| < 1$.

The following matrix spectral factorization algorithm computes an invertible $\Theta(L)\zeta_t$. If $\Gamma(\lambda) = (\sum_{k=-\infty}^{+\infty}\Gamma_k\lambda^k)/2\pi$ denotes the covariance generating function of $\Theta(L)\zeta_t$, where $\Gamma_k = \Theta(L)E\zeta_t\zeta_{t-k}^T\Theta(L)^T$, then, $\Gamma_k = H\Phi_k^* GG^T H^T + \ldots + H\Phi_{s-1}^* GG^T\Phi_{s-k-1}^{*} H^T$ for k = 0, ..., s-1, $\Gamma_k = \Gamma_{-k}^T$ for k = -1, ..., -s+1, $\Gamma_k = 0_{r\times r}$ for k = ±s, ±(s+1), ..., and

$$(5.12)\qquad \Gamma(\lambda) = \Theta(\lambda)\Sigma_\zeta\Theta(\lambda^{-1})^T/2\pi,$$

where $\Theta(\lambda) = \sum_{k=0}^{s-1}\Theta_k\lambda^k$. Equation (5.12) is valid whether or not $\Theta(L)\zeta_t$ is invertible, but an invertible $\Theta(L)\zeta_t$ can be computed iff $\det[H\Phi^*(\lambda)G] = 0$ has no unit roots. An invertible $\Theta(L)\zeta_t$ can be computed, for example, following Zadrozny (1998, sec. 4, pp. 1366-1368). For such an invertible $\Theta(L)\zeta_t$, $\det[\Theta(1)] \neq 0$, so that

$$(5.13)\qquad \Sigma_\zeta = 2\pi\Theta(1)^{-1}\Gamma(1)\Theta(1)^{-T}.$$

Thus, we have described computing an implicit factor model in the form of an invertible VARMA(s,s-1) process of the r importance-weighted variables of interest,

$$(5.14)\qquad w_t = \phi_1 w_{t-1} + \ldots + \phi_s w_{t-s} + \zeta_t + \Theta_1\zeta_{t-1} + \ldots + \Theta_{s-1}\zeta_{t-s+1}$$

where $\zeta_t \sim NIID(0,\Sigma_\zeta)$. Equation (5.14) is "implicit" because it implicitly incorporates the dynamics of the significant factors. Zellner and Palm (1974) considered a similar so-called "final" equation, but only for a data model, not for a smaller factor model.

5.3. Computing Forecasts.

We now consider h period-ahead forecasts of w_t made for out-of-sample periods N+1, ..., N+M. State representation (5.5)-(5.6) implies that h-period-ahead forecasts of w_t made in periods t = N+1-h, ..., N+M-h are given by

$$(5.15)\qquad w_{t+h|t} = HF^h x_{t|t},$$

where $x_{t|t}$ denotes the filtered estimate of x_t made in period t, based on the in-sample-estimated model and the accumulated data, $\overline{y}_{1,t} = (y_1^T, \ldots, y_t^T)^T$. Although the MA part of a model is not apparent in equation (5.15), it is part of the computation of $x_{t|t}$. The forecast error is $\eta^*_{ht} = w_{t+h} - w_{t+h|t}$ and has the covariance matrix

$$(5.16) \qquad \Gamma_h^* = \sum_{i=0}^{h-1} W\Psi_i R_1 R_1^T \Psi_i^T W^T + \Sigma_u,$$

which is positive definite, because $\Sigma_u = \delta I_r$ is positive definite. In section 6, we consider only nonrecursive forecasts, based on a constant model estimated with in-sample data (in recursive forecasting, the model is reestimated as more data becomes available, as time moves forward, and forecast-error covariances are updated correspondingly).

Reconstructibility means that the state vector can be estimated in terms of the model and the data (Kwakernaak and Sivan, 1972). Reconstructibility is relevant because filtered or smoothed estimates of the state vector can be computed iff reconstructibility holds. Data model (2.1) has a state representation analogous to equations (5.5)-(5.6), with observation matrix H and state-transition matrix F. If no data are missing for $t \geq \tau \geq 1$, so that all of $\overline{y}_{\tau,t} = (y_\tau^T, \ldots, y_t^T)^T$ is observed, then, $\overline{y}_{\tau,t} = R_{t-\tau}x_\tau$ + disturbances, where

$$(5.17) \qquad R_{t-\tau} = [H^T, (HF)^T, \ldots, (HF^{t-\tau})^T]^T.$$

The Cayley-Hamilton theorem tells us that, for $t-\tau \geq s-1$, $R_{t-\tau}$ has full rank s, where s is the dimension of the state vector, iff the reconstructibility matrix, $R_{s-1} = [H^T, (HF)^T, \ldots, (HF^{s-1})^T]^T$, has full rank s. R_{s-1} is called the reconstructibility matrix because, for $t-\tau \geq s-1$, x_τ can be estimated or "reconstructed" in terms of the model and $\overline{y}_{\tau,t}$ iff R_{s-1} has full rank. For example, abstracting from disturbances, $\overline{y}_{\tau,t} = R_{t-\tau}x_\tau$ implies that $(R_{t-\tau}^T R_{t-\tau})^{-1} R_{t-\tau}^T \overline{y}_{\tau,t}$ is a regression-type estimate of x_τ, for $t-\tau \geq s-1$, iff R_{s-1} has full rank.

If the model is time varying or the data have missing values for whatever reasons, reconstructibility can be defined more generally in terms of time-varying observation and state-transition matrices, H_t and F_t. In our experience (Zadrozny, 1990; Chen and Zadrozny, 1998), time-invariant VARMA models estimated using monthly-quarterly data have reconstructible state vectors according to the more general definition.

5.4. Comments.

If we replace data model (2.1) with significant-factor model (5.5)-(5.6), such that the factor model's parameters, A_1^*, ..., B_q^*, R_1, are considered free-standing and underived from data model (2.1), then, we always decrease the number of ARMA-coefficient parameters by $(n^2-m^2)(p+q)$ and often (iff $n > 2m-1$) decrease the number of disturbance-covariance parameters by $n(n+1)/2 - nm$. However, if R is derived from data model (2.1) by WCD, then, the number of free disturbance-covariance parameters stays unchanged at $n(n+1)/2$ and using factor model (5.5)-(5.6) saves $(n^2-m^2)(p+q)$ ARMA-coefficient parameters.

If going from data model (2.1) to factor model (5.5)-(5.6) reduces the number of parameters, then, the reduction in the number of parameters equals the number of overidentifying restrictions on data model (2.1), which may be tested using a likelihood ratio test. In the application in section 6, we estimate data model (2.1) as an unrestricted model and use WCD to derive factor model (5.5)-(5.6) from the data model. Alternatively, we could estimate factor model (5.5)-(5.6) as a free-standing model, which might reduce uncertainty about parameter estimates and improve forecast accuracy (Mariano and Murasawa, 2003). We could also estimate, test, and forecast with a more restricted, orthogonal, m-factor model, with stochastically independent factors, hence, with diagonal ARMA coefficient matrices. PCD factors are always orthogonal (Anderson, 1984, chs. 11 and 14), but WCD factors are generally orthogonal only if $\Sigma_\xi Q$ is symmetric. In practice, we want the best-forecasting factor model but practically cannot evaluate all alternatives. WCD should help by reducing an unrestricted n-variable data model to an m-factor model, generally for $m << n$. In the application in section 6, the $m = 1$ and 2 significant-factor models forecast GDP almost as accurately as the $n = 5$ and 11 unrestricted data models.

WCD and minimum realization (MR; Aoki, 1983, 1987; Mittnik, 1990a-b, 1992; Kapetanios and Marcellino, 2004), respectively, can remove redundant variables and lags from data models. In section 6, we use MLE to estimate parameters of data models. Although data models often have redundant variables and lags, MLE is statistically efficient if a model is correctly specified. We could estimate data models using MR and remove redundant lags at the start. If we did this, the estimated model should have no remaining lag redundancies. In theory, MR exactly removes redundancies manifested by zero singular values, but, in practice, MR is approximate, because sampling noises make truly zero singular values slightly positive. WCD is approximate in both theory and practice, because it uses empirical tests to determine the number of significant factors. Because WCD operates on assumed and estimated input matrices, Ω and $\Gamma_h = \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T$, and MR operates on estimated matrices, H, F, and G, and implied matrices $\Psi_i \Sigma_\xi \Psi_i^T = HF^iGG^T(F^i)^TH^T$, MR could precede WCD.

6. Application to Forecasting Quarterly GDP at Monthly Intervals.

6.1. Description of Data.

The data used in this application were all obtained from the Conference Board (2002). The variables depicted in Figure 1 are quarterly real GDP, monthly composite coincident index, and four monthly coincident indicators (employment, personal income, industrial production, and real manufacturing sales), all in standardized logs, with outliers greater than four standard deviations converted to missing values. Observed quarterly GDP is assigned to the third month of a quarter and is treated as missing in the first two months of each quarter. Data missing for any reason are not graphed. Thus, because horizontal time axes are indexed monthly in all figures, GDP should appear as an interrupted line, which is apparent in percentage-growth Figures 3 and 4, but not in log-level Figures 1 and 2. The monthly composite index and indicators are graphed for 522 months from 1959:1 to 2002:6 and quarterly GDP is graphed for 174 quarters from 1959:1 to 2002:2. Figure 3 differs from Figure 1 in that GDP and the indicators are in standardized first differences of logs, the first months and quarters are lost to differencing, GDP is differenced quarterly, and the indicators are differenced monthly.

The variables in Figure 2 are quarterly real GDP, monthly composite leading index, and ten leading indicators (average weekly hours, unemployment claims, new orders of capital goods, vendor performance, new orders of consumption goods, building permits, Standard and Poor's 500 stock index, the money stock in M2 definition, interest on 3-month Treasury bills minus interest on 1-year Treasury notes, and the Conference Board's index of consumer expectations), all in standardized logs, with outliers (seen in Figures 3-4) greater than 3.0 standard deviations converted to missing values. As in Figure 1, missing data are not graphed, GDP is treated as observed in the third month of a quarter, the indicators are graphed for 522 months from 1959:1 to 2002:6

and GDP is graphed for 174 quarters from 1959:1 to 2002:2. Figure 4 differs from Figure 2 in that GDP and the indicators are in standardized first differences of logs, the first months and quarters are lost to differencing, GDP is differenced quarterly, and the indicators are differenced monthly.

6.2. Data Models and Forecasts.

The paper applies WCD to VARMA-data models estimated using exact Kalman-filtering-based ML (Zadrozny, 1990) and U.S. monthly-indicator (4 coincident, 10 leading) and quarterly-real-GDP data from February 1959 to June 1992. The models were estimated using the monthly-quarterly data in their given mixed-frequency form in order to include at least the mixed-frequency aspect of "real time" analysis, where "real time" means taking explicit account of mixed frequencies, data delays, or data revisions. In developing forecasting models, Stock and Watson (2002b) accounted explicitly for data delays and Schumacher and Breitung (2006) accounted explictly for data delays and mixed frequencies. We emphasize, however, that using mixed-frequency data in the application has nothing per se to do with WCD. We note again that we produce and evaluate GDP forecasts mainly as a natural way of evaluating WCD and advocate WCD as a general tool for analyzing multivariate time-series models, not just as a tool for producing better forecasts.

Models 1-4 in Tables 1-4 are estimated using data from the "in-sample" period of 401 months, from 1959:2 to 1992:6. The "out-of-sample" period for forecasting and evaluating forecast accuracy is the 120 months from 1992:7 to 2002:6. In the tables, R_e^2 denotes the "estimation" or in-sample unadjusted R^2; NRMSE denotes normalized root mean-squared forecast error (RMSE) ÷ out-of-sample standard deviations of forecasted GDP; R_f^2 denotes the "forecasting" or out-of-sample unadjusted $R^2 = 1 - \text{NRMSE}^2$; Q denotes Ljung-Box Q statistics for testing significance of residual autocorrelations at lags 1-24 for monthly indicators and at lags 1-8 for quarterly GDP; and, p denotes the Q statistics' marginal significance levels or p values. Being based on fewer degrees of freedom, p values of GDP are higher per value of Q. |AR| and |MA| denote absolute AR and MA roots, with repetitions in parentheses.

[to be completed]

6.3. Factor Models and Forecasts.

"WCD 1a" means that W = (1,...,1) (all variables are weighted equally) and no AR roots are filtered out; "WCD 1b" means that W = (1,...,1) and AR roots with |AR| > .98 are filtered out; "WCD 2a" means that W = (0,...,0,1) (only GDP is weighted positively) and no AR roots are filtered out; and, "WCD 2b" means that W = (0,...,0,1) and AR roots with |AR| > .98 are filtered out. Finally, m denotes the number of factors determined to be significant, according to the statistical test in szection 5, for $\rho = \alpha = .05$.

We evalute WCD in terms of the forecast accuracy of WCD-reduced models, because a model is often considered "best" if it produces the most accurate out-of-sample forecasts. We are not trying to challenge other forecasting methods in a "horse race" for the best forecasting method. Thus, for simplicity, in section 6.2 we consider only nonrecursive forecasts in which a data model is estimated using in-sample data, for t = 1, ..., N, and is held fixed when computing and evaluating forecasts in out-of-sample periods, t = N+1, ..., N+M. By contrast, in forecasting horseraces, recursive forecasts are more commonly evaluated, in which the forecasting model is reestimated in every out-of-sample period as new data accrue.

The application's main conclusion is that WCD can reduce moderately large VARMA-data models of "important" GDP and up to 10 "unimportant" indicators to small univariate ARMA-factor models of "important" GDP, which forecast GDP almost as accurately as the larger VARMA-data models. It has been known since Nelson (1972) that small univariate ARMA models can forecast macroeconomic variables much more accurately than large econometric models. It is especially interesting here that the small monthly univariate-ARMA-factor models produced GDP forecasts almost as accurate as the VARMA-data models, yet are difficult to estimate directly using the quarterly GDP data in monthly form (e.g., with GDP observations considered missing in the first two months of a quarter and assigned to the third month of a quarter). We tried to do this using Kalman-filtering-based MLE but failed.

[to be completed]

7. Conclusion.

The paper develops and applies a method, called weighted-covariance factor decomposition (WCD) for reducing larger estimated VARMA-data models of "important" variables of primary interest and "unimportant" variables, included to help account for variations in the "important" variables, to smaller VARMA-factor models of the "important" variables and significant factors. WCD, first, reduces an estimated data model to a significant-factor model in state-space form, in which significant factors are explicit determinants of state variables, and, then, transforms the state-space form to an implicit-factor VARMA form in the "important" variables and their innovations, in which the significant factors have been eliminated but determine the VARMA structure.

[to be completed]

Appendix: Computing $\nabla\hat{K}$ to Test for the Number of Significant Factors.

[to be completed]

REFERENCES

Aguilar, O. and M. West (2000), "Bayesian Dynamic Factor Models and Variance Matrix Discounting for Portfolio Allocation," Journal of Business and Economic Statistics 18: 338-357.

Ahn, S.K. and G.C. Reinsel (1988), "Nested Reduced-Rank Autoregressive Models for Multiple Time Series," Journal of the American Statistical Association 83: 849-856.

Akaike, H. (1973), "Information Theory and and Extension of the Maximum Likelihood Principle," pp. 267-281 in Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), Budapest, Hungary: Akademia Kiado.

Anderson, T.W. (1963), "Asymptotic Theory for Principal Component Analysis," Annals of Mathematical Statistics 34: 122-148.

Anderson, T.W. (1984), An Introduction to Multivariate Statistical Analysis, 2nd. ed., New York, NY: John Wiley.

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

Angelini, E., J. Henry, and R. Mestre (2001), "A Multi Country Trend Indicator for Euro Area Inflation: Computation and Properties," Working Paper No. 60, European Central Bank, Frankfurt, Germany.

Ansley, C.F. and R. Kohn (1983), "Exact Likelihood of Vector Autoregressive Moving-Average Process with Missing or Aggregated Data," Biometrika 70: 275-278.

Aoki, M. (1983), Notes on Economic Time Series Analysis: System Theoretic Perspectives, Berlin, Germany: Springer Verlag.

Aoki, M. (1987), State Space Modeling of Time Series, Berlin, Germany: Springer Verlag.

Bai, J. and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models," Econometrica 70: 191-221.

Bernanke, B. (1986), "Alternative Explanations of the Money-Income Correlation," Carnegie-Rochester Conference Series on Public Policy 2: 49-99.

Box, G.E.P. and G. Tiao (1977), "A Canonical Analysis of Multiple Time Series," Biometrika 6: 355-365.

Chen, B. and P.A. Zadrozny (1998), "An Extended Yule-Walker Method for Estimating Vector Autoregressive Models with Mixed-Frequency Data," pp. 47-73 in Advances in Econometrics: Messy Data--Missing Observations, Outliers, and Mixed-Frequency Data, Vol. 13, T.B. Fomby and R.C. Hill (eds.), Greenwich, CT: JAI Press.

Chen, B. and P.A. Zadrozny (2003), "Higher Moments in Perturbation Solution of the Linear-Quadratic Exponential Gaussian Optimal Control Problem," Computational Economics 21: 45-64.

Cooley, T.F. and S.F. LeRoy (1985), "Atheoretical Macroeconomics: A Critique," Journal of Monetary Economics 1: 283-308.

Conference Board (2002), coincident and leading indicator data downloaded from http://www.conference-board.org, New York, NY: Conference Board.

Connor, G. and R.A. Korajczyk (1993), "A Test for the Number of Factors in an Approximate Factor Model," Journal of Finance 48: 1263-1291.

Deistler, M. and E. Hamman (2005), "Identification of Factor Models for Forecasting Returns," Journal of Financial Econometrics 3: 256-281.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," Review of Economics and Statistics 82: 540-554.

Forni, M. and L. Reichlin (1998), "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," Review of Economic Studies 65: 453-473.

Gantmacher, F.R. (1959), The Theory of Matrices, Vol. 1, Providence, RI: AMS Chelsea.

Geweke, J.F. and K.J. Singleton (1981), "Maximum Likelihood 'Confirmatory' Factor Analysis of Economic Time Series," International Economic Review 2: 37-54.

Hannan, E.J. and B.G. Quinn (1979), "The Determination of the Order of an Autoregression," Journal of the Royal Statistical Society Series B 41: 190-195.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50: 1029-1054.

Hosoya, Y. and M. Taniguchi (1982), "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," Annals of Statistics 10: 132-153.

Hotelling, H. (1933), "Analysis of a Complex of Statistical Variables into Principal Components, Journal of Educational Psychology 24: 417-441.

Kapetanios, G. and M. Marcellino (2004), "A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions," Working Paper No. 489, Department of Economics, Queen Mary University of London, London, UK.

Kim, C.J. and C.R. Nelson (1999), State Space Models with Regime Switching, Cambridge, MA: MIT Press.

Kollo, T. and H. Neudecker (1993), "Asymptotics of Eigenvalues and Unit-Length Eigenvectors of Sample Variance and Correlation Matrices," Journal of Multivariate Analysis 47: 283-300.

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

Magnus, J.R. and H. Neudecker (1988), Matrix Differential Calculus with Applications in Statistics and Econometrics, New York, NY: John Wiley & Sons.

Mariano, R.S. and Y. Murasawa (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," Journal of Applied Econometrics 18: 427-443.

Mittnik, S. (1990a), "Macroeconomic Forecasting Experience with Balanced State Space Models," International Journal of Forecasting 6: 337-348.

Mittnik, S. (1990b), "Forecasting with Balanced State Space Representations of Multivariate Distributed Lag Models," Journal of Forecasting 9: 207-218.

Mittnik, S. (1992), "Forecasting International Growth Rates with Leading Indicators: A System-Theoretic Approach," Computers and Mathematics with Applications 24: 31-41.

Mittnik, S. and P.A. Zadrozny (1993), "Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Dynamic Models," Econometrica 61: 857-870.

Mittnik, S. and P.A. Zadrozny (2005), "Forecasting Quarterly German GDP at Monthly Intervals Using Monthly Ifo Business Conditions Data," pp. 19-48 in J.E. Sturm and T. Wollmershäuser (eds.), Ifo Survey Data in Business Cycle and Monetary Policy Analysis, Heidelberg, Germany: Physica Verlag.

Nelson, C.R. (1972), "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," American Economic Review 62: 902-917.

Otrok, C. and C.H. Whiteman (1998), "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," International Economic Review 39: 997-1014.

Sargent, T.J. and C.A. Sims (1977), "Business Cycle Modeling without Pretending to Have Too Much A Priori Economic Theory," pp. 45-109 in C.A. Sims (ed.), New

Methods in Business Cycle Research, Minneapolis, MN: Federal Reserve Bank of Minneapolis.

Schumacher, C. and J. Breitung (2008), "Real-Time Forecasting of GDP Based on a Large Factor Model with Monthly and Quarterly Data," Discussion Paper No. 33/2006, Deutsche Bundesbank, Frankfurt, Germany.

Schwarz, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics 8: 147-164.

Serfling, R. (1980), Approximation Theorems of Mathematical Statistics, New York: John Wiley.

Sims, C.A. (1980a), "Macroeconomics and Reality," Econometrica 48: 1-48.

Sims, C.A. (1980b), "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," American Economic Review 70: 250-257.

Sims, C.A. (1986), "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis Quarterly Review 10 (1): 2-16.

Stock, J.H. and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators", pp. 351-394 in O.J. Blanchard and S. Fischer (eds.), NBER Macroeconomics Annual 1989, Cambridge, MA: M.I.T. Press.

Stock, J.H. and M.W. Watson (2002a), "Forecasting Using Principal Components from a Large Number of Predictors," Journal of the American Statistical Association 97: 1167-1179.

Stock, J.H. and M.W. Watson (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," Journal of Business and Economic Statistics 20: 147-162.

Stoica, P. and T. Söderström (1983), "Optimal Instrumental Variable Estimation and Approximate Implementations," IEEE Transactions on Automatic Control 28: 757-772.

Zadrozny, P.A. (1990), "Estimating a Multivariate ARMA Model with Mixed-Frequency Data: An Application to Forecasting U.S. GNP at Monthly Intervals," Working Paper No. 90-6, Research Department, Federal Reserve Bank of Atlanta, Atlanta, GA.

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.

Zellner, A. and F. Palm (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," Journal of Econometrics 2: 17-54.

Figure 1: Logs of Quarterly Real GDP, Monthly Coincident Index, and 4 Monthly Coincident Indicators, January 1959 to June 1992

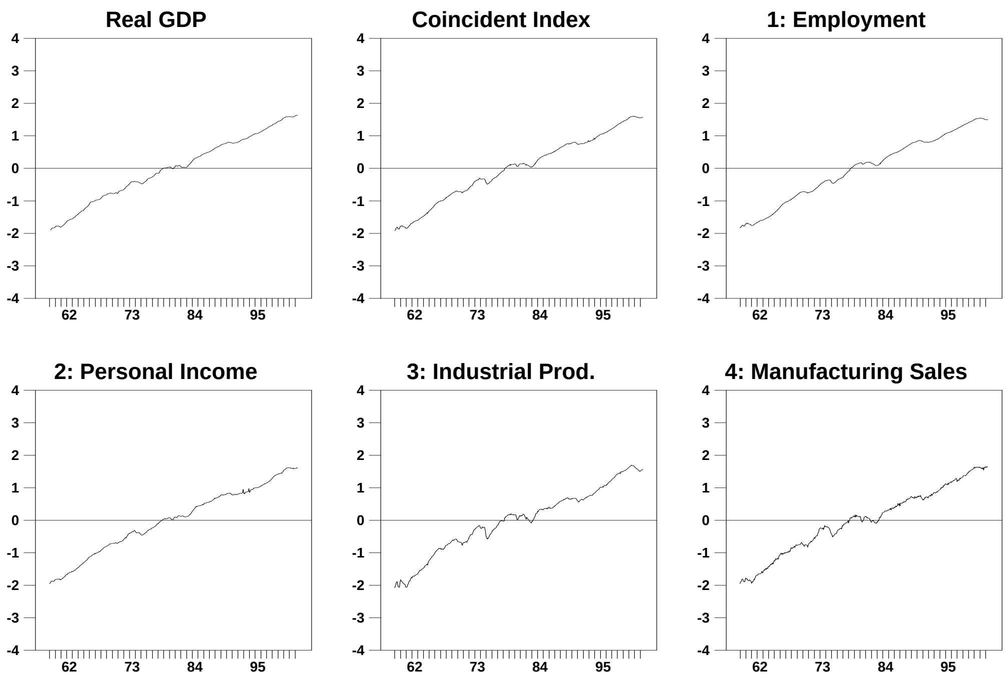

Figure 2: Percentage Growth of Quarterly Real GDP, Monthly Coincident Index, and 4 Monthly Coincident Indicators, January 1959 to June 2002

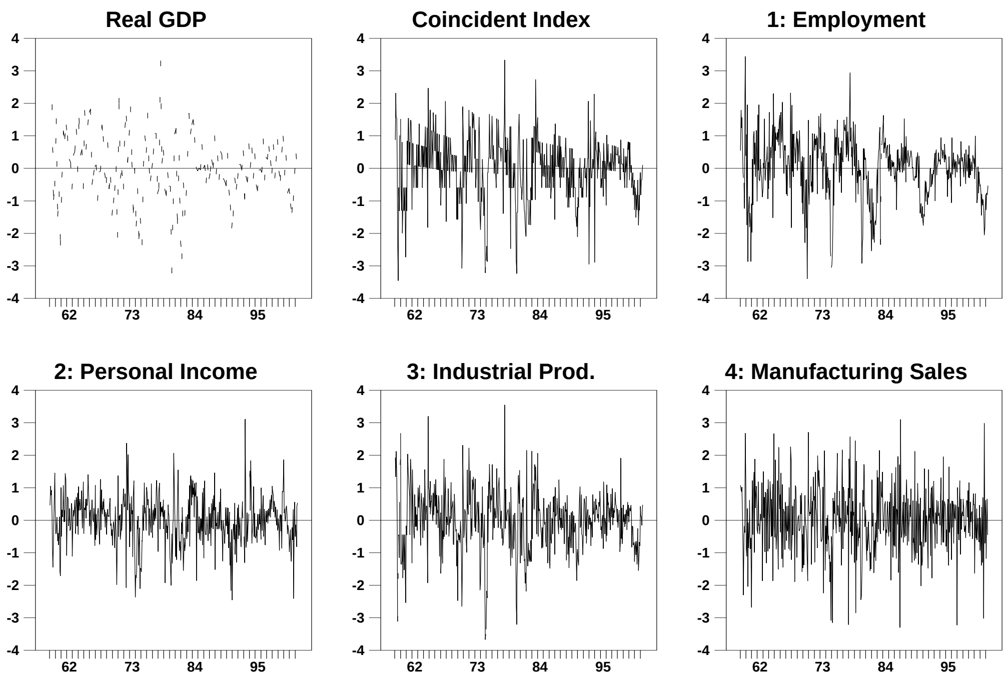

Figure 3: Logs of Quarterly Real GDP, Monthly Leading Index, and 10 Monthly Leading Indicators, January 1959 to June 2002

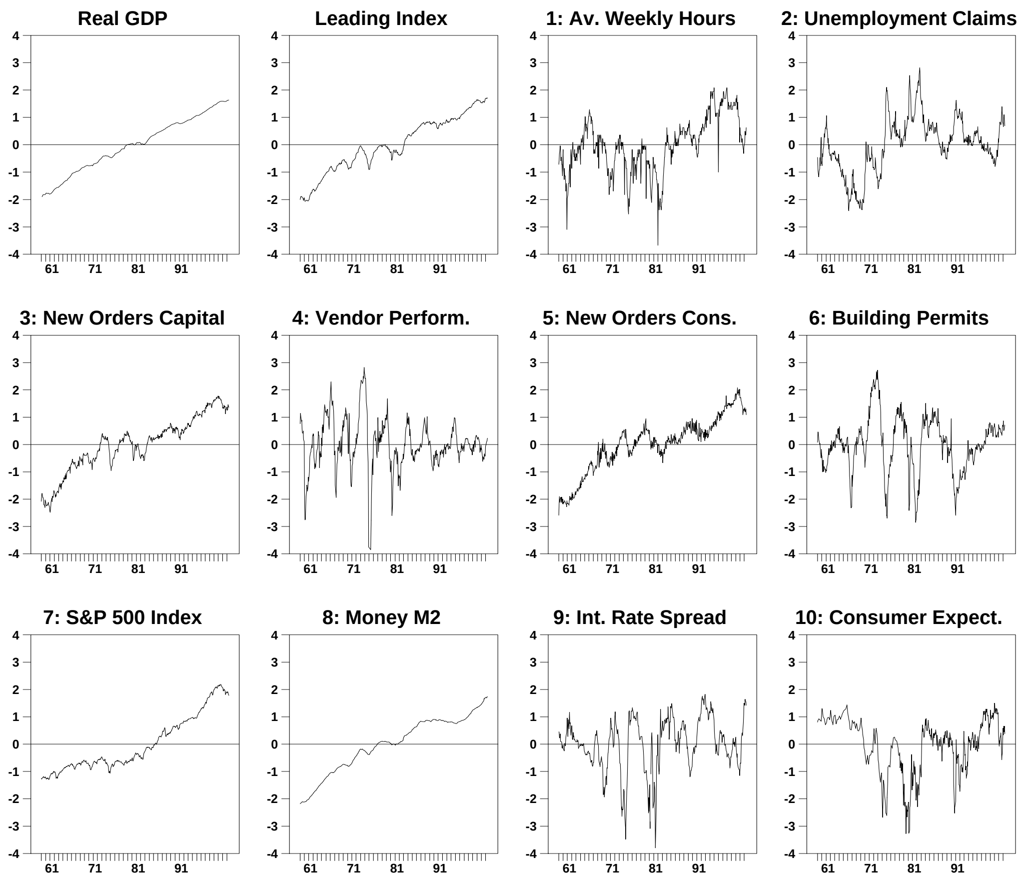

Figure 4: Percentage Growth of Quarterly Real GDP, Monthly Leading Index, and 10 Monthly Leading Indicators, January 1959 to June 2002

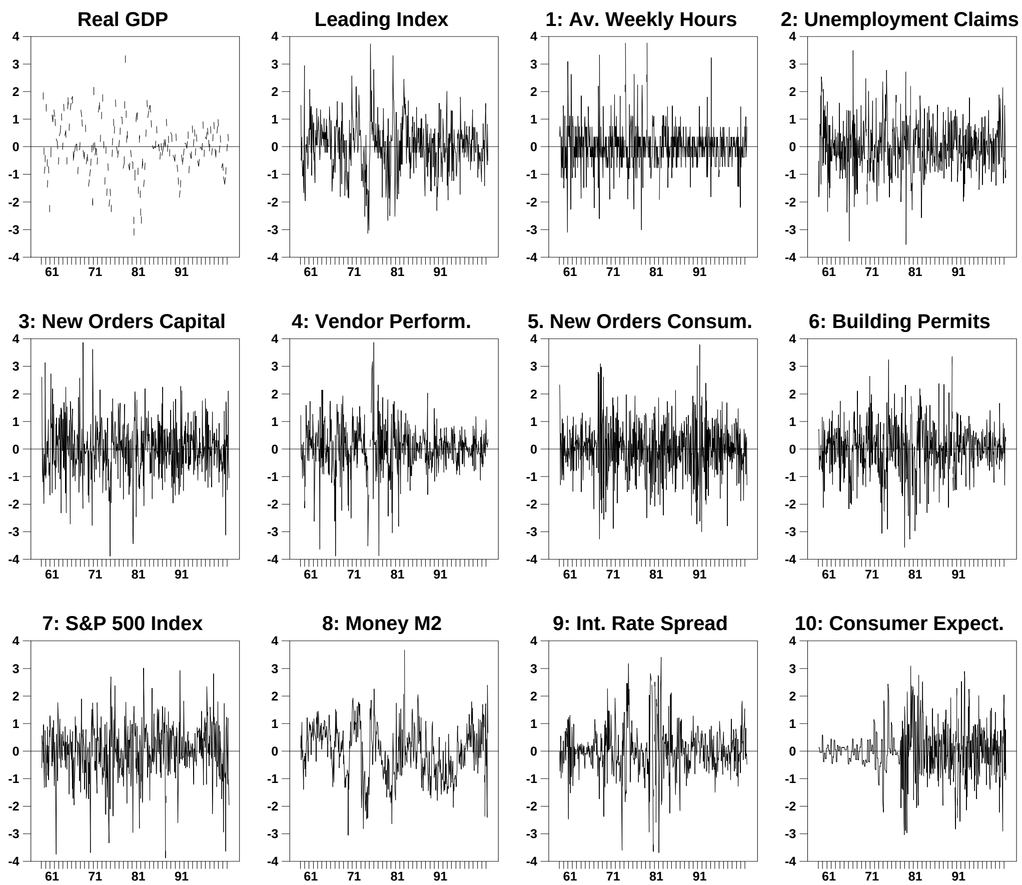

Table 1a-b: Logs of 4 Monthly Coincident Indicators and Quarterly GDP: Estimated Monthly Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly VIMA(1,1) Data Model 1
Using Data for 1959:1 - 1992:6

Variable	CI1	CI2	CI3	CI4	GDP
R_e^2	.992	.991	.988	.989	.973
Q	170.	26.1	25.7	41.3	17.5
P	.000	.350	.367	.016	.025

|AR| = 1.00(5); |MA| = 1.00, .234, .178, .108, .034

b. Data Model 1 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6

GDP std. devs.: in-sample .792, out-of-sample .252

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.009	.037	.999	.495
2	.015	.061	.996	.814
3	.018	.071	.995	.950
6	.034	.134	.982	.956
12	.066	.262	.931	.970
24	.130	.517	.732	.990
Average 1-24	.070	.279	.922	.900

N = in-sample degrees of freedom = no. of in-sample periods - no. of estimated parameters = 401 - 40 = 361 months.

M = no. of out-of-sample periods = 120 months.

Table 1c-d: Logs of 4 Monthly Coincident Indicators and Quarterly GDP: Implied Monthly GDP ARMA Model and Its Forecast Accuracy

c. Data Model 1 Decompositions for h = 12 Months

WCD	φ_1	φ_2	φ_3	φ_4	φ_5	m
1a	.776	.103	.067	.040	.014	4
1b	.461	.299	.110	.085	.045	4
2a	.991	.009	.000	.000	.000	1
2b	.874	.126	.000	.000	.000	2

d. Monthly GDP IMA(1,1) Factor Model 1-2b
and Its Monthly Forecast Accuracy for 1992:7 - 2002:6

$$\ln GDP_t = \ln GDP_{t-1} + \zeta_t + .131 \cdot \zeta_{t-1}$$

$$\sigma_\zeta = 1.18, \quad |AR| = 1.00, \quad |MA| = .131$$

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.018	.073	.995	.974
2	.018	.073	.995	.974
3	.018	.073	.995	.974
6	.035	.139	.981	.985
12	.068	.271	.927	.992
24	.130	.528	.721	.996
Average 1-24	.074	.302	.909	.990

Table 2a-b: Percentage Growth of 4 Monthly Coincident Indicators and Quarterly GDP: Estimated Monthly Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly VAR(1) Data Model
Using Data for 1959:1 - 1992:6

Variable	CI1	CI2	CI3	CI4	GDP
R_e^2	.349	.195	.154	.156	.554
Q	23.5	23.4	30.4	57.6	15.3
p	.487	.499	.173	.000	.054

|AR| = .805, .331, .145, .136, .088

b. Data Model 2 Monthly GDP Forecast Accuracy
for 1992:7 to 2002:6
GDP std. devs.: in-sample 1.09, out-of-sample .581

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.557	.960	.078	.696
2	.569	.980	.040	.711
3	.635	1.09	-.188	.794
6	.636	1.10	-.210	.817
12	.640	1.10	-.210	.739
24	.635	1.09	-.188	.679
Average 1-24	.630	1.09	-.188	.710

N = in-sample degrees of freedom = no. of in-sample periods - no. of estimated parameters = 401 - 40 = 361 months.

M = no. of out-of-sample periods = 120 months.

Table 2c-d: Percentage Growth of 4 Monthly Coincident Indicators and Quarterly GDP: Implied Monthly GDP ARMA Model and Its Forecast Accuracy

c. Data Model 2 Decompositions for h = 12 Months

WCD	φ_1	φ_2	φ_3	φ_4	φ_5	m
1a	.624	.164	.090	.065	.057	5
2a	.735	.256	.008	.000	.000	2

d. Monthly GDP ARIMA(2,1) Factor Model 2-2a and Its Forecast Accuracy for 1992:7 - 2002:6

$$\Delta \ln GDP_t = .866 \cdot \Delta \ln GDP_{t-1} - .037 \cdot \Delta \ln GDP_{t-2} + \zeta_t - .328 \cdot \zeta_{t-1}$$

$$\sigma_\zeta = 1.32, \quad |AR| = .907, \; .041, \quad |MA| = .328$$

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.633	1.09	-.188	.790
2	.633	1.09	-.188	.790
3	.633	1.09	-.188	.790
6	.618	1.06	-.124	.793
12	.639	1.10	-.210	.738
24	.635	1.09	-.188	.679
Average 1-24	.635	1.09	-.188	.716

Table 3a-b: Logs of 10 Monthly Leading Indicators and Quarterly GDP: Estimated Monthly Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly VIMA(1,1) Data Model 3
Using Data for 1959:1 - 1992:6

Var.	LI1	LI2	LI3	LI4	LI5	LI6	LI7	LI8	LI9	LI10	GDP
R_e^2	.841	.964	.975	.883	.950	.935	.990	.989	.908	.932	.968
Q	31.5	20.0	27.1	25.8	38.3	27.2	36.1	27.2	52.4	38.0	3.65
P	.139	.697	.298	.365	.032	.294	.053	.294	.001	.035	.888

$|AR|$ = 1.00(11)

$|MA|$ = .612(2), .538(2), .447, .436, .238, .237, .120, .145(2)

b. Data Model 3 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6
GDP std. devs.: in-sample .792, out-of-sample .252

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.028	.112	.987	1.49
2	.040	.160	.974	2.14
3	.037	.145	.979	1.93
6	.049	.195	.962	1.39
12	.077	.304	.908	1.13
24	.139	.552	.695	1.06
Average 1-24	.083	.328	.892	1.19

N = in-sample degrees of freedom = no. of in-sample periods - no. of estimated parameters = 401 - 187 = 214 months.

M = no. of out-of-sample periods = 120 months.

Table 3c-d: Logs of 10 Monthly Leading Indicators and Quarterly GDP: Implied Monthly GDP ARMA Model and Its Forecast Accuracy

c. Data Model 3 Decompositions for h = 12 Months

WCD	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}	φ_{11}	m
1a	.45	.19	.11	.08	.06	.03	.02	.01	.00	.00	.00	6
1b	.39	.22	.11	.08	.05	.05	.03	.02	.01	.00	.00	7
2a	.94	.05	.00	.00	.00	.00	.00	.00	.00	.00	.00	2
2b	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1

d. Monthly GDP IMA(1,1) Factor Model 3-2b and Its Monthly Forecast Accuracy for 1992:7 - 2002:6

$$\ln GDP_t = \ln GDP_{t-1} + \zeta_t - .679 \cdot \zeta_{t-1}$$

$$\sigma_\zeta = .996, \quad |AR| = 1.00, \quad |MA| = .679$$

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.035	.141	.980	1.87
2	.035	.141	.980	1.87
3	.035	.141	.980	1.87
6	.052	.206	.958	1.47
12	.085	.335	.888	1.24
24	.148	.586	.657	1.12
Average 1-24	.092	.365	.866	1.32

Table 4a-b: **Percentage Growth of 10 Monthly Leading Indicators and Quarterly GDP: Estimated Monthly Data Model and Its GDP Forecast Accuracy**

a. Estimated Monthly VAR(1) Data Model 4
Using Data for 1959:1 - 1992:6

Var.	LI1	LI2	LI3	LI4	LI5	LI6	LI7	LI8	LI9	LI10	GDP
R_e^2	.132	.106	.172	.135	.202	.159	.078	.555	.179	.101	.443
Q	66.6	27.3	51.1	50.4	83.4	36.9	24.5	40.3	41.0	60.9	29.0
p	.000	.293	.001	.001	.000	.045	.431	.020	.017	.000	.000

$|AR|$ = .728(2), .396, .342, .320(2), .220(2), .204, .135(2)

b. Data Model 4 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6
GDP std. devs.: **in-sample 1.09, out-of-sample .581**

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.596	1.03	-.061	.744
2	.664	1.14	-.232	.829
3	.725	1.25	-.558	.905
6	.649	1.12	-.248	.834
12	.623	1.07	-.150	.720
24	.636	1.09	-.198	.680
Average 1-24	.638	1.10	-.206	.720

N = in-sample degrees of freedom = no. of in-sample periods - no. of estimated parameters = 401 - 187 = 214 months.

M = no. of out-of-sample periods = 120 months.

Table 4c-d: Percentage Growth of 10 Monthly Leading Indicators and Quarterly GDP: Implied Monthly GDP ARMA Model 4-2b and Its Forecast Accuracy

c. Data Model 4 Decompositions for h = 12-Months

WCD	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}	φ_{11}	m
1a	.258	.163	.104	.080	.074	.065	.062	.057	.055	.047	.035	10
2a	.766	.174	.047	.012	.001	.000	.000	.000	.000	.000	.000	2

d. Monthly GDP ARIMA(2,1) Factor Model 4-2a and Its Monthly Forecast Accuracy for 1992:7 - 2002:6

$$\Delta \ln GDP_t = .993 \cdot \Delta \ln GDP_{t-1} - .191 \cdot \Delta \ln GDP_{t-2} + \zeta_t - .279 \cdot \zeta_{t-1}$$

$$\sigma_\zeta = 1.18, \quad |AR| = .732, \quad .261, \quad |MA| = .279$$

Months Ahead	RMSE	NRMSE	R_f^2	Theil U
1	.632	1.09	-.188	.790
2	.632	1.09	-.188	.790
3	.632	1.09	-.188	.790
6	.619	1.07	-.145	.795
12	.636	1.10	-.210	.735
24	.636	1.10	-.210	.680
Average 1-24	.633	1.09	-.188	.714