WEIGHTED-COVARIANCE FACTOR DECOMPOSITION OF VARMA MODELS APPLIED TO U.S. MONTHLY-QUARTERLY MACROECONOMIC DATA*

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

June 26, 2011

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

JEL codes: C33, C53

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

---

*Previous versions of this work have been presented in Aix-en-Provence (6/02), Philadelphia (10/02), Frankfurt (12/02), Evanston (6/03), Seattle (7/03), Salt Lake City (8/07), San Francisco (4/08), Irvine (5/08), Paris (6/08), Denver (8/08), Washington (9/08), Vienna (6/09), Canberra (7/09), and Davis (9/09). The paper represents the authors' views and does not necessarily 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 and Tara Sinclair for comments.

**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.

ABSTRACT

We develop and apply a method, called weighted-covariance factor decomposition (WCFD), for reducing an estimated VARMA "data" model of observed variables being considered to a smaller VARMA "factor" model of a subset of observed variables of primary interest. Although WCFD is conceptually and computationally closely related to principal components decomposition (PCD), it has 3 notably different features: (1) whereas PCD strictly applies only to stationary data, WCFD applies without change to any mixture of stationary and nonstationary data and models; (2) whereas PCD implicitly takes a long time perspective, in WCFD the user sets a time perspective of any finite duration; (3) like PCD, WCFD can reduce data to factors, but, unlike PCD, also reduces "data" models to "factor" models. We illustrate WCFD with U.S. monthly indicators (4 coincident, 10 leading) and quarterly real GDP. We estimate 4 monthly models of 5 and 11 variables, in log and differenced-log forms, apply WCFD to the estimated models, determine the number of significant factors for each model; reduce each model to a univariate "implicit-factor" GDP model; and compare the estimated "data" and reduced "factor" models in terms of RMSEs of out-of-sample GDP forecasts. The application's main conclusion is that WCFD can reduce moderately-large monthly models of quarterly GDP and up to 10 monthly indicators to univariate monthly GDP models which can forecast out-of-sample GDP at monthly intervals about as accurately as their larger antecedent models.

1. Introduction.

Parsimony or minimizing the "size" of a model, without compromising its ability to fit data, is the 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 (e.g., Akaike, 1973). 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 (e.g., Aoki, 1983, 1987; Mittnik, 1990a-b).

Subject matter can often indicate what "primary" variables should be included in a model, but is often unclear about what other "secondary" variables should be included, which are not of primary interest but are presumed to significantly affect the primary variables. Models may include too many secondary variables and there is no general agreement on eliminating unnecessary ones. The present weighted-covariance factor decomposition (WCFD) method addresses this issue by weighting variables in a PC-like decomposition. The idea is to start with a set of observed variables, estimate a VARMA "data" model of them, and, then, use WCFD to reduce the data model to a smaller "factor" model involving only variables of "primary" interest.

In the paper, we discuss deriving WCFD, computing WCFD, adapting to WCFD Anderson's (1984, pp. 473-475) asymptotic test for the number of significant eigenvalues of a sample covariance matrix, reducing a data model to a factor model, and applying these steps to 4 estimated monthly data models of log and first-differenced-log forms of U.S. monthly-indicator and quarterly-GDP data from January 1959 to June 2002. Using the estimated data models and the derived factor models, we compute out-of-sample monthly forecasts of GDP and evaluate their accuracy in terms of root mean-squared errors (RMSE). We evaluate GDP forecasts as a way of evaluating WCFD, not as an attempt to forecast GDP per se. The application has the interesting result that the derived monthly univariate GDP-factor models forecast GDP at monthly intervals almost as accurately as the larger antecedent data models.

Hotelling (1933) introduced the principal components decomposition (Anderson, 1984, ch. 11). Let $\{\hat{C}_k\}_{k=0}^{\infty}$ denote population autocovariance matrices of a vector of variables in y_t and their k-period lags in y_{t-k} and let $\hat{C}_k$ denote a sample-based estimate of C_k. Classical PCD uses only $\hat{C}_0$ and, thus, is static (Bai and Ng, 2002; Stock and Watson, 2002a-b). 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 $\{\hat{C}_0\}_{k=0}^{K}$ to produce dynamic decompositions.

Earlier maximum likelihood estimates (MLE) of dynamic factor models (Sargent and Sims, 1977; Geweke and Singleton, 1981) were restricted to small models, with few variables and few parameters, because MLE is demanding computationally. By the 1990s, despite large reductions in computing costs, interest shifted to applying PCD to hundreds of observed financial and macroeconomic variables to produce also small dynamic factor models (Stock and Watson, 2002a-b; Forni et al., 1998, 2000).

PCD is computationally feasible if $\hat{C}_0$ is nonsingular, which happens with probability one if the data have more sampling periods than variables and no variables exactly satisfy linear equations. Additional assumptions in terms of the approximate factor model (AFM) are usually made so that PCD provides consistent estimates of AFM parameters, principally assumptions on eigenvalues as the sample size and the number of variables go to infinity (Bai and Ng, 2002; Stock and Watson, 2002a-b). By contrast, in this paper, we assume a fixed and finite set of variables.

WCFD has three notable features that differentiate it from PCD:

1. Stationarity. Strictly, PCD assumes that the data are stationary, so that C_0 exists. For example, Stock and Watson (2002b) first-difference time-series data to stationary form if their originally given values are nonstationary. By contrast, WCFD applies to a data model regardless whether it is stationary or nonstationary, because WCFD uses a model's transfer-function coefficients from 0 to h-1, where h denotes a finite number of chosen forecast periods and the finite number of transfer-function coefficients can be computed whether the model is stationary or nonstationary.

If some AR roots are nonstationary or near-nonstationary and h is sufficiently large, then, WCFD will tend to "latch onto" the largest AR characteristic 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 WCFD. The AR roots are filtered out only to compute WCFD and are not removed from the estimated data model before it is reduced to a factor 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. WCFD 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. Because, for stationary data, 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 WCFD decomposes the weighted covariance matrix of h-period-ahead forecast errors, WCFD could be said to take a finite h-period-ahead time perspective.

Although mixed frequencies have nothing per se to do with WCFD, they are used in the application to make it more interesting. Kalman-filtering-based MLE has been used to estimate VARMA models using mixed-frequency data (Zadrozny, 1990; Mittnik and Zadrozny, 2004) but computational constraints restrict MLE 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 which efficiently estimates a VAR model when data have missing values, including mixed frequencies. However, unlike Kalman-filter-based MLE, XYW can estimate a VAR model but not a VARMA model.

WCFD also provides a purely-data-based and economic-theory-free variance decomposition, but this aspect will not be emphasized here. Sims (1980a-b) advocated computing variance decompositions of estimated VAR models to judge explanatory power of one variable on another. Initially, Sims favored variance decompositions based on Cholesky decomposition, which is a purely numerical method, inherently with no particular economic meaning. Following Cooley and Leroy's (1985) critique, Bernanke (1986), Sims (1986), and most others now mostly base variance decompositions on structural identifications. Being based on a presumably well-fitting data model, yet being economic-theory-free, WCFD could be used as an exploratory data-based variance decomposition, prior to a more conclusive structural decomposition. The application in section 6 similarly follows a reduced-form or parameter-restriction-free approach, in contrast to the structural approaches of, say, Mariano and Murasawa (2003) or Evans (2005), with many detailed restrictions on parameters.

Although the literature cited above and the application below are based on sampling estimation methods, dynamic factor models have been estimated using Bayesian methods (Otrok and Whiteman, 1998; Kim and Nelson, 1999; Aguilar and West, 2000). The statistics literature has also considered dynamic factor models under the rubrics canonical analysis (Box and Tiao, 1977) and reduced rank regression (Ahn and Reinsel, 1988; Deistler and Hamman, 2005).

The paper proceeds as follows. Section 2 defines WCFD. Section 3 proposes two closely related methods for computing WCFD, also closely related to the usual method for computing PCD. Section 4 adapts to WCFD Anderson's (1984, pp. 473-475) asymptotic test for the number of significant eigenvalues of a sample covariance matrix. Section 5 discusses computing factors and reducing a data model to a significant-factor model in explicit and implicit forms. Section 6 applies WCFD to ML estimated VARMA-data models of monthly indicators and quarterly GDP. Section 7 concludes the paper.

2. Defining WCFD.

In what follows, all quantities are real valued except possibly characteristic roots, which may be complex valued. 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 a 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 or $\xi_t \sim \text{NIID}(0_{n\times 1},\Sigma_\xi)$, where $0_{n\times 1}$ denotes the n×1 zero vector. 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 WCFD we need only the reduced form.

In general, y_t is a vector of data means plus a vector of deviations from the means. The means are the nonstochastic and (unconditionally) predictable parts of y_t and the deviations from the 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 WCFD pertains only to the stochastic deviations from the 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 standardized 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 $S = \text{diag}[s_1, ..., s_n]$ denote the n×n diagonal matrix with positive standardizing values, s_i. In case (i), the data should be standardized as (a) $y_t := S^{-1}y_t$ prior to estimating the data model, where := denotes assignment and s_i denotes the sample standard deviation of variable i. In case (ii), estimated reduced-form parameters should be standardized 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 denotes the estimated standard deviation of disturbance i. The standardizations promote accurate computations, because they eliminate differences in magnitudes of numbers due to differences in units of measurement of different variables. In particular, standardization (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 the absolute characteristic AR roots are < 1, namely, if and only if $\det[I_n\lambda^p - A_1\lambda^{p-1} - ... - A_{p-1}\lambda - A_p] = 0$ implies $|\lambda| < 1$. Model (2.1) is invertible if and only if the absolute characteristic MA roots are < 1, namely, if and only if $\det[I_n\lambda^q + B_1\lambda^{q-1} + ... + B_{q-1}\lambda + B_q] = 0$ 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 if and only if its absolute characteristic AR and MA roots are > 1, although here it is more convenient to define the characteristic roots as above, so that they can be computed as the eigenvalues of a companion matrix (e.g., F in equation (5.7)).

If data model (2.1) is stationary, then, it has the transfer function

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

where $\Psi(L) = A(L)^{-1}B(L)$. Whether or not the model is stationary or invertible, the finite sequence $\{\Psi_i\}_{i=0}^{h-1}$ can be computed by iterating on

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

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, WCFD 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 WCFD. Let $\{\lambda_i\}_{i=1}^{\upsilon}$ denote the ν nonstationary or near-nonstationary AR roots of data model (2.1). We would filter transfer function (2.2) using $\lambda(L) = (1-\lambda_1 L)\cdots(1-\lambda_\nu L)$ and would obtain filtered function $\tilde{y}_t = \tilde{\Psi}(L)\xi_t$, where $\tilde{y}_t = \lambda(L)y_t$ and $\tilde{\Psi}(L) = \lambda(L)\Psi(L)$. Then, we would compute WCFD for $\tilde{\Psi}(L)$ and Σ_ξ. The roots would be filtered out from $\Psi(L)$ only to compute WCFD and would not be removed from the data models 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

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

and have the covariance matrix

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

where superscript T denotes transposition. We propose evaluating WCFD in terms of the forecast accuracy of the r×1 vector of primary weighted variables, $w_t = Wy_t$, where W is an r×n matrix of chosen weights and $1 \le r = \text{rank}(W) \le n$. Then, $\Omega = W^T W$ represents the weights as an n×n symmetric positive semi-definite matrix and $v = E\eta_{ht}^T \Omega \eta_{ht} = \text{tr}[\Omega\Gamma_h]$ 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

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

The following three examples illustrate singular and nonsingular weighting matrices:

1. Minimum Portfolio Variance. Let $W = (w_1, \ldots, w_n)$ denote a $1\times n$ vector of portfolio allocation weights. Every variable in a portfolio receives a nonzero weight according to whether it is an asset ($w_i > 0$) or a liability ($w_i < 0$). 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 $n > 1$. WCFD decomposes portfolio forecast-error variance. If Σ_ξ is positive definite, then, Γ_h is positive definite and $W = (\overline{e}^T \Gamma_h^{-1} \overline{e})^{-1} \overline{e}^T \Gamma_h^{-1}$ minimizes portfolio variance with respect to W, subject to $W^T \overline{e} = 1$, where $\overline{e} = (1, \ldots, 1)^T$ is the $n\times 1$ vector of ones.

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

3. Principal Components. Let $\Omega = \sum_{i=1}^{n} e_i e_i^T = I_n$, the nonsingular $n\times n$ identity matrix. In addition, if all variables are stationary and the forecast horizon, h, is large (strictly, $h = \infty$), then, v is the sum of the unweighted variances of all variables and WCFD corresponds to PCD.

WCFD produces $n\times n$ decomposition matrix R, which satisfies $RR^T = \Sigma_\xi$ and is analogous to the factor-loading matrix in PCD. If Σ_ξ is positive definite, as we assume in assumption (i) below, then, R is invertible and we can define structural disturbance vector $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})^T \equiv R^{-1}\xi_t$, usually called "structural" because its elements are uncorrelated. Let r_i denote column i of R and let v_i denote weighted h-step-ahead forecast-error covariances accounted for by structural disturbance i. Then, equation (2.6) can be written equivalently as $v = \sum_{i=1}^{n} v_i$, where

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

for $i = 1, \ldots, n$, and

(2.8) $Q \equiv \sum_{i=0}^{h-1} \Psi_i^T \Omega \Psi_i$.

Q is symmetric positive semidefinite by construction and may or may not be positive definite if the forecast horizon, h, is sufficiently large.

We define WCFD recursively. For i = 1, ..., n-1, let $\Sigma_i = \Sigma_\xi - X_i X_i^T$, $X_i = [r_1, ..., r_{i-1}]$, $X_1 = 0_{n\times 1}$, and $Y_i \equiv [r_{i+1}, ..., r_n]$, a matrix of slack variables which ensures that $r_i r_i^T + Y_i Y_i^T = \Sigma_i$ is satisfied. For i = 1, ..., n-1, we want to maximize $\varphi_i = r_i^T Q r_i / v$, with respect to r_i, so that $r_i r_i^T + Y_i Y_i^T = \Sigma_i$ is satisfied and, for i = n, we want r_n to satisfy $X_n X_n^T + r_n r_n^T = \Sigma_\xi$. Because 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 $X_2 = r_1$ and $\Sigma_2 = \Sigma_\xi - X_2 X_2^T$, 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 $X_3 = [r_1, r_2]$ and $\Sigma_3 = \Sigma_\xi - X_3 X_3^T$, 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 $X_n = [r_1, ..., r_{n-1}]$. Finally, given X_n and $\Sigma_n = \Sigma_\xi - X_n X_n^T$, we determine r_n which satisfies $r_n r_n^T = \Sigma_n$, so that $Y_n = 0_{n\times n}$, and, thus, determine $R = [r_1, ..., r_n]$.

We assume (i) that Σ_ξ is positive definite, (ii) that $W \neq 0_{r\times n}$, hence, that $\Sigma_\xi Q = \Sigma_\xi W^T W \neq 0_{n\times n}$ (have positive rank), and (iii) that $\Sigma_\xi Q$ has distinct nonzero (real and positive) eigenvalues. Section 3 shows that the significant part of WCFD, associated with significantly positive and distinct eigenvalues of $\Sigma_\xi Q$, exists and is unique if assumptions (i)-(iii) hold. In practice, assumptions (i) and (iii) can be expected to hold when variables are not subject to linear restrictions and assumption (ii) holds always by construction.

If the data model is stationary, then, $\sum_{i=0}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T = \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T + \sum_{i=h}^{\infty} \Psi_i \Sigma_\xi \Psi_i^T$ or unconditional data covariances = conditional h-period-ahead forecast-error covariances + unconditional h-period-ahead forecast covariances, such that all three terms are finite. WCFD decomposes $tr[\Omega \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T]$ or 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 unconditional forecast covariances (standardized, unweighted, for h = 1, any VAR model, not subject to any side restrictions). Although these objectives might seem significantly different, WCFD applied to unstandardized 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 produced similar results.

3. Computing WCFD.

We consider computing WCFD in two parts: for i = 1, ..., n-1 and for i = n. Following the definition of WCFD in section 2, first, we derive first-order conditions for i = 1, ..., n-1, discuss two methods for solving them, and, then, state the condition for i = n and discuss a method for solving it.

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

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

where Ξ_i is an n×n matrix of Lagrange multipliers. We obtain first-order conditions by differentiating $\mathcal{L}_i$ with respect to r_i, Y_i, and Ξ_i and setting the derivatives to zero,

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

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

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

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

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

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 the eigenvalue problems

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

for i = 1, ..., n-1.

We recall some linear algebra. A real positive semidefinite or positive definite square matrix, whether symmetric or not, has real, respectively, nonnegative and positive eigenvalues and real eigenvectors. Σ_ξ is symmetric positive semidefinite by construction and positive definite by assumption; Q is symmetric positive semidefinite by construction and may be positive definite if the forecast horizon, h, is sufficiently large. The product of two symmetric positive semidefinite matrices is positive semidefinite and generally asymmetric. If Σ_ξ has full rank n, then, Σ_i has rank n-i+1.

We now describe method 1 for solving first-order conditions (3.6). For i = 1, ..., n, let $\lambda_{i1} \geq \ldots \lambda_{in} \geq 0$ denote the real nonnegative eigenvalues of $\Sigma_i Q$, ordered in decreasing size and let z_{i1}, ..., z_{in} denote associated real eigenvectors. Then, for i = 1, ..., n-1, because $v_i = r_i^T Q r_i$ is being maximized, we set $v_i = \lambda_{i1}$, determine m, the number of significant v_i, and, for i = 1, ..., m, set

$$(3.7)\qquad r_i = \sqrt{\frac{\lambda_{i1}}{z_{i1}^T Q z_{i1}}}\, z_{i1},$$

where associated eigenvector z_{i1} is scaled as $z_{i1}^T z_{i1} = 1$ and $v_i = r_i^T Q r_i$, as desired. The eigenvalues of $\Sigma_i Q$ exist and are unique in any case. Assumption (iii), that nonzero eigenvalues of $\Sigma_\xi Q$ are distinct, implies that $[z_{11}, \ldots, z_{m1}]$ and, hence, that $R_1 = [r_1, \ldots, r_m]$ exist and are unique (Wilkinson, 1965, pp. 4-6).

If the last n-m eigenvalues of $\Sigma_\xi Q$ are sufficiently near zero to be numerically indistinct, then, $R_2 = [r_{m+1}, \ldots, r_n]$ generally cannot be computed accurately (Golub and Van Loan, 1996). But, because we need only significant R_1 for further WCFD computations, we can forego computing insignificant R_2.

We now describe method 2 for solving first-order conditions (3.6). Method 2 follows from method 1 because, for i, j = 1, ..., n and i ≠ j, $r_i^T Q r_j = 0$ and $r_i^T \Sigma_\xi^{-1} r_j$ (cf., Box and Tiao, 1977, pp. 356-357). Let $\mu_1 \geq \ldots \geq \mu_n \geq 0$ denote the real nonnegative eigenvalues of $\Sigma_\xi Q$ in decreasing size and let $x_1, \ldots, x_n$ denote associated real eigenvectors. Then, for i = 1, ..., n-1, we compute μ_i, set $v_i = \mu_i$, determine m = the number of significant v_i, and, for i = 1, ..., m, set

$$(3.8) \qquad r_i = \sqrt{\frac{\mu_1}{x_i^T Q x_i}}\, x_i,$$

where associated eigenvector x_i is scaled as $x_i^T x_i = 1$ and $v_i = r_i^T Q r_i$, as desired.

First-order conditions (3.6) could include part 2 for i = n, but, for greater computational accuracy, we solve it separately as follows. First, compute v_n residually as $v_n = tr(\Sigma_\xi Q) - \sum_{i=1}^{n-1} v_i$. If v_n is significant and r_n is needed for further WCFD computations, then, we recommend computing r_n by solving $r_n r_n^T = \Sigma_n$ for r_n. Postmultiplying $r_n r_n^T = \Sigma_n$ by r_n, we obtain the eigenvalue problem

$$(3.9) \qquad (\Sigma_n - \omega_i I_n)\chi_i = 0_{n\times 1},$$

for i = 1, ..., n, where $\omega_1 \geq \ldots \geq \omega_n$ denote the eigenvalues of Σ_n in decreasing size and $\chi_1, \ldots, \chi_n$ denote associated eigenvectors. Because Σ_n has rank one, $\omega_1 > 0$, $\omega_2 = \ldots = \omega_n = 0$, and we set

$$(3.10) \qquad r_n = \sqrt{\omega_1}\, \chi_1,$$

where associated eigenvector χ_1 is scaled as $\chi_1^T \chi_1 = 1$ and $\omega_1 = r_n^T r_n$ and $RR^T = \Sigma_\xi$, as desired.

Methods 1 and 2 share part 2. Method 2 is simpler to implement and was used in the application in section 6. Both methods require about the same number of computations as PCD (Anderson, 1984, ch. 11, pp. 451-460). Whereas PCD decomposes a covariance matrix ($\hat{C}_0$), WCFD decomposes decomposes the product of a covariance matrix (Σ_ξ) and a symmetric positive semi-definite weighting matrix (Q).

4. Testing for the Number of Significant Factors.

This section describes a strategy for testing for the number of significant v_i, i.e., for the number of significant "factors". For a given "threshold", $\rho \in (0,1)$, let H_m denote the hypothesis that, for i = 1, ..., m, v_i are significant in the sense of accounting for at least the fraction $1-\rho$ of weighted covariances. Thus, H_m is true if and only if

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

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

Consider the following testing sequence. Start with m = 1. For m = 1, test H_m; if H_m is accepted, accept m = 1 as the number of significant factors. Otherwise, for m = 2, test H_m; if H_m is accepted, accept m = 2 as the number of significant factors. Otherwise, continue like this until possibly reaching m = n-1; if H_m is accepted for m = n-1, accept m = n-1 as the number of significant factors. Otherwise, accept m = n as the number of significant factors. Because $v_i \ge 0$ and $\sum_{i=1}^{n} v_i = 1$, the testing sequence is always conclusive.

WCFD is in effect a nonlinear differentiable function from ϕ to $\bar{v}$, denoted $\bar{v} = \bar{v}(\phi)$. If ϕ is certain, say, because it is assumed for a hypothetical model, then, H_m is not rejected if and only if $\delta_m \le 0$. However, generally, ϕ is estimated and uncertain, so that H_m must be tested probabilistically. We can do this by expanding $\delta_m \le 0$ to a probabilistic statement similar to Anderson's (1984, ch. 11, pp. 473-475) asymptotic test for the number of significant PCD factors.

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

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

where $\hat{\sigma}^2_{\delta_m} \equiv \overline{\rho}^T\nabla\hat{\overline{v}}\,\hat{S}\,\nabla\hat{\overline{v}}^T\overline{\rho}$. For given $\hat{\phi}$ and $\hat{S}$, applying a matrix-differentiation method (Mittnik and Zadrozny, 1993; Chen and Zadrozny, 2003, appendix A), it is straightforward but tedious to obtain equations for computing $\nabla\hat{\overline{v}}$, hence $\hat{\sigma}^2_{\delta_m}$.

Using equations (4.1)-(4.2) and following standard 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 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 if and only if

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

If $\hat{\phi}$ is not asymptotically normally distributed but has an asymptotic distribution, then, inequality (4.3) becomes an approximate test. For example, if the data model has unit roots, then, c_α based on the normal distribution will tend to underestimate the true c_α.

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 significant-factor model of r ($\le$ n) primary weighted variables of interest. In subsection 5.1, we define factors and discuss computing significant factors for an in-sample period used to estimate an n-variable data model. The computed significant factors could be used as observed variables to estimate a smaller m+r-variable model of m significant factors and r primary variables (m+r $\le$ n) for forecasting the primary variables. Stock and Watson (2002a-b), 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 1 reduces the n-variable VARMA model to an "explicit" r-variable state-space form, in which state variables are explicit linear combinations of significant factors. Step 2 eliminates the factors and reduces the explicit state-space form for r primary variables to an implicit VARMA form, in which the factors do not appear explicitly but are implicit in the VARMA structure. Section 6 illustrates reductions of 5- and 11-variable data models to 1-variable GDP implicit-factor models.

5.1. Computing Factors.

We define the n×1 vector of factors analogously to the approximate factor model (Bai and Ng, 2002; Stock and Watson, 2002a-b) by $f_t = R^{-1}y_t$. If we compute R according to section 3 and test its columns for significance in terms of associated v_i, then, we obtain $Rf_t = R_1f_{1t} + R_2f_{2t}$, where $R = [R_1, R_2]$, $R_1 = [r_1, \ldots, r_m]$, $R_2 = [r_{m+1}, \ldots, r_n]$, $f_t = (f_{1t}^T, f_{2t}^T)^T$, and subscripts 1 and 2, respectively, denote significant and insignificant parts. Then,

$$(5.1) \qquad f_t = \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = R^{-1}y_t = \begin{bmatrix} (R^{-1})_1 \\ (R^{-1})_2 \end{bmatrix} y_t .$$

If y_t is observed in period t, then, equation (5.1) is feasible. Otherwise, if y_t has missing values for any reason, because R^{-1} is generally a full matrix with no zero elements, then, no element of f_t can generally be computed using equation (5.1). However, in such cases, equation (5.1) can be made feasible by replacing the missing values of y_t with estimates, made using a Kalman filter or smoother (Anderson and Moore, 1979) or some other missing-data estimation method.

5.2. Computing Explicit and Implicit Factor Models.

We now describe computation of an invertible VARMA representation for the weighted variables of primary interest in w_t. Zellner and Palm (1974) considered similar equations, which they called "final" equations, but only for what we are calling data models.

We now suppose that significant $v_1, \ldots, v_m$ and associated significant R_1 have been determined and describe a method for computing significant-factor models in explicit state-space and implicit VARMA forms. Let $\varepsilon_t = (\varepsilon_{1t}^T, \varepsilon_{2t}^T)^T$ be partitioned conformably with R and f_t into m×1 and (n-m)×1 subvectors of significant and insignificant disturbances. We can restate VARMA-data model (2.1) as VARMA-factor model

$$(5.2)\qquad \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = \begin{bmatrix} (R^{-1})_1 A_1 R_1 & (R^{-1})_1 A_1 R_2 \\ (R^{-1})_2 A_1 R_1 & (R^{-1})_2 A_1 R_2 \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \end{bmatrix} + \ldots + \begin{bmatrix} (R^{-1})_1 A_p R_1 & (R^{-1})_1 A_p R_2 \\ (R^{-1})_2 A_p R_1 & (R^{-1})_2 A_p R_2 \end{bmatrix} \begin{bmatrix} f_{1,t-p} \\ f_{2,t-p} \end{bmatrix}$$

$$+ \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} (R^{-1})_1 B_1 R_1 & (R^{-1})_1 B_1 R_2 \\ (R^{-1})_2 B_1 R_1 & (R^{-1})_2 B_1 R_2 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \ldots + \begin{bmatrix} (R^{-1})_1 B_q R_1 & (R^{-1})_1 B_q R_2 \\ (R^{-1})_2 B_q R_1 & (R^{-1})_2 B_q R_2 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-q} \\ \varepsilon_{2,t-q} \end{bmatrix}.$$

where $R^{-1} = [(R^{-1})_1^T, (R^{-1})_2^T)^T$.

To see the consequences of treating f_{2t} as insignificant in equation (5.2), first, $\Sigma_\xi = R_1R_1^T + R_2R_2^T$ implies that WCFD objective (2.6) partitions as

$$(5.3)\qquad tr[\Omega \textstyle\sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T] = tr[\Omega \sum_{i=0}^{h-1} \Psi_i R_1 R_1^T \Psi_i^T] + tr[\Omega \sum_{i=0}^{h-1} \Psi_i R_2 R_2^T \Psi_i^T],$$

which is equivalent to $v = \sum_{i=1}^{m} v_i + \sum_{i=m+1}^{n} v_i$. Because in equation (5.3) the second term is insignificant, in equation (5.2) we can set $R_2 = 0_{n\times(n-m)}$ and can discard the $(n-m)\times 1$ bottom subequation for f_{2t}, and in equation (5.5) can discard u_t as negligible white noise.

Thus, with f_{2t} and ε_{2t} insignificant and discarded, we reduce VARMA n-factor model (5.2) to VARMA m-factor model

$$(5.4)\qquad f_{1t} = A_1^* f_{1,t-1} + \ldots + A_p^* f_{1,t-p} + \varepsilon_{1t} + B_1^* \varepsilon_{1,t-1} + \ldots + B_q^* \varepsilon_{1,t-q},$$

where $A_i^* = (R^{-1})_1 A_i R_1$ for $i = 1, \ldots, p$, $B_j^* = (R^{-1})_1 B_j R_1$ for $j = 1, \ldots, q$, and $\varepsilon_{1t} \sim NIID(0, I_m)$. Process (5.4) is stationary if and only if $\det[I_n\lambda^p - A_1^*\lambda^{p-1} - \ldots - A_{p-1}^*\lambda - A_p^*] = 0$ implies $|\lambda| < 1$ or, equivalently, if and only if the eigenvalues of transition matrix F in state equation (5.6) are less than one in absolute value. Generally, if $m < n$ factors are significant, then, the data models and significant-factor models have different AR and MA roots.

We now write the w_t process in state-space and transfer-function forms. First, $w_t = W\sum_{i=0}^{\infty} \Psi_i R_1 \varepsilon_{1,t-i} + u_t$, where $u_t = W\sum_{i=0}^{\infty} \Psi_i R_2 \varepsilon_{2,t-i}$ is considered negligible white noise, namely, $u_t \sim NIID(0, \delta I_r)$, where δ is a small positive scalar. Setting δ to a positive number rather than to zero promotes accurate Kalman-filter computations. Thus, w_t has observation equation

$$(5.5) \qquad w_t = Hx_t + u_t,$$

in terms of s×1 state vector $x_t = (x_{1t}^T, \ldots, x_{\ell t}^T)^T$, where $H = [W^T R_1, 0_{r\times m}, \ldots, 0_{r\times m}]$ is r×s, $s = m\cdot\max(p,q+1)$, VARMA process (5.4) implies state equation

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

where $A_i^* = 0_{m\times m}$ for $i > p$, and $B_i^* = 0_{m\times m}$ for $i > q$ (Ansely and Kohn, 1983).

Combining (5.5) and (5.6) and discarding u_t yields transfer function

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

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) 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 $\Phi_0^* = I_s$.

Transfer function (5.7) implies VARMA(s,s-1) representation

(5.10) $\phi(L)w_t = H\Phi^*(L)G\varepsilon_{1t}$,

where $H\Phi^*(L)G\varepsilon_{1t} = HG\varepsilon_{1t} + H\Phi_1^*G\varepsilon_{1,t-1} + \ldots + H\Phi_{s-1}^*G\varepsilon_{1,t-s+1}$ and $H\Phi_i^*G$ are r×m, for $r \leq m$. We want a "standard" VARMA form which has the same number of disturbances as variables and whose MA part is invertible. Even if r = m, then, the computed MA part of (5.10) may be noninvertible. If there are no unit MA roots, then, using matrix spectral factorization (e.g., Zadrozny, 1998, sec. 4, pp. 1366-1368), we can recompute (5.10) to standard invertible VARMA form

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

where $\phi(L) = 1 - \phi_1 L - \ldots - \phi_s L^s$ is a scalar polynomial of degree s, $\Theta(L)\zeta_t = \zeta_t + \Theta_1\zeta_{t-1} + \ldots + \Theta_{s-1}\zeta_{t-s+1}$ is an invertible r×r matrix polynomial of degree s-1, and ζ_t is an r×1 disturbance vector ~ $NIID(0,\Sigma_\zeta)$.

6. Application to U.S. Monthly-Quarterly Macroeconomic Data.

6.1. Description of the Data.

U.S. macroeconomic data from January 1959 to June 2002 were obtained from the Conference Board (2002). The data in figures 1-2 are quarterly real GDP, the monthly composite coincident index, and 4 monthly coincident indicators (employment, personal income, industrial production, and real manufacturing sales). The data in figure 1 are standardized logs and the data in figure 2 are standardized first differences of logs. Observed quarterly GDP was assigned to month 3 of a quarter and treated as missing in months 1-2 of a quarter. Monthly variables were differenced over 1 month and GDP was differenced over 1 quarter or 3 months. Outliers greater than 3.5 standard deviations were converted to missing values. Data which were actually missing or treated as missing are not graphed. Thus, with monthly horizontal time axes, GDP is graphed as an interrupted line, although this is apparent only in figures 2 and 4. The data are graphed for the available 522 months from January 1959 to June 2002. The monthly coincident index is graphed for completeness, but was not used.

The data in figures 3-4 are quarterly real GDP, the monthly composite leading index, and 10 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). The data in figures 3-4 were processed and are displayed as in figures 1-2 and for the same period. The monthly leading index is graphed for completeness, but was not used.

6.2. Data Models and Forecasts.

The paper applies WCFD to models estimated using Kalman-filtering-based ML (Zadrozny, 1990) and the U.S. macroeconomic data. The models were estimated using the monthly-quarterly data at their given sampling rates. Although mixed frequencies have nothing per se to do with WCFD, using them makes the application more interesting. GDP forecasts are made and evaluated only as a way of comparing estimated data models with their WCFD reductions and not in order to develop GDP forecasts per se.

The data models were estimated using the 401-month in-sample period from January 1959 to June 1992. GDP forecasts were made and evaluated for the 120-month out-of-sample period from July 1992 to June 2002. In the tables, R_e^2 denotes unadjusted R^2; RMSE denotes root mean-squared forecast error; Q denotes Ljung-Box Q statistics for testing significant residual serial correlations at 1-24 monthly lags for the indicators and at 1-8 quarterly lags for GDP; and, p denotes p values of the Q statistics. Being based on fewer degrees of freedom, the GDP p values are higher per value of Q. |AR| and |MA| denote absolute AR and MA roots, with the number of repeated roots in parentheses.

6.3. Factor Models and Forecasts.

In tables c-d "WCFD 1a" means that W = (1,...,1) (all variables weighted equally) and no AR roots are filtered out; "WCFD 1b" means that W = (1,...,1) and AR roots with |AR| > .98 are filtered out; "WCFD 2a" means that W = (0,...,0,1) (only GDP gets nonzero weight) and no AR roots are filtered out; and, "WCFD 2b" means that W = (0,...,0,1) and AR roots with |AR| > .98 are filtered out. Finally, m denotes the number of significant v_i, hence, the number of significant "factors", according to a nonprobabilistic application of the test in section 4 (α = 1, c_α = 0) and threshold ρ = .05.

We evalute WCFD in terms of forecast accuracy because a "best" model should make the most accurate out-of-sample predictions. Because we are not in a forecasting competition, we consider only nonrecursive forecasts based on a fixed data model estimated using the in-sample data from January 1959 to June 1992. Recursive forecasts are based on repeatedly reestimated models in the out-of-sample period.

7. Conclusion.

The paper has developed and illustrated the weighted-covariance factor decomposition (WCFD) method for reducing an estimated VARMA-data model, of variables of primary interest and secondary variables presumed to affect the primary variables, to VARMA-implicit-factor models, of only the primary variables, which embody the significant effects of the secondary variables on the primary variables.

The application's main conclusion is that WCFD can reduce estimated VARMA-data models of primary GDP and up to 10 secondary indicators to univariate-ARMA implicit-factor GDP models which forecast GDP about as accurately as their antecedent VARMA-data models. This was demonstrated empirically by Nelson (1972), who showed that small univariate-ARMA models can forecast U.S. macroeconomic data as accurately or more accurately than large econometric models. It is interesting here that the monthly univariate-ARMA GDP models forecast GDP almost as accurately as the 5- to 11-variable monthly VARMA-data models, yet could not be estimated directly using only the quarterly GDP data in monthly form.

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. (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.

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.

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

Evans, M.D. (2005), "Where Are We Now? Real-Time Estimates of the Macro Economy," International Journal of Central Banking 1: 127-175.

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.

Golub, G.H. and C.F. Van Loan (1996), Matrix Computations, third edition, Baltimore, MD: Johns Hopkins University Press.

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.

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

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. 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.

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 (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.

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

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.

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 2002

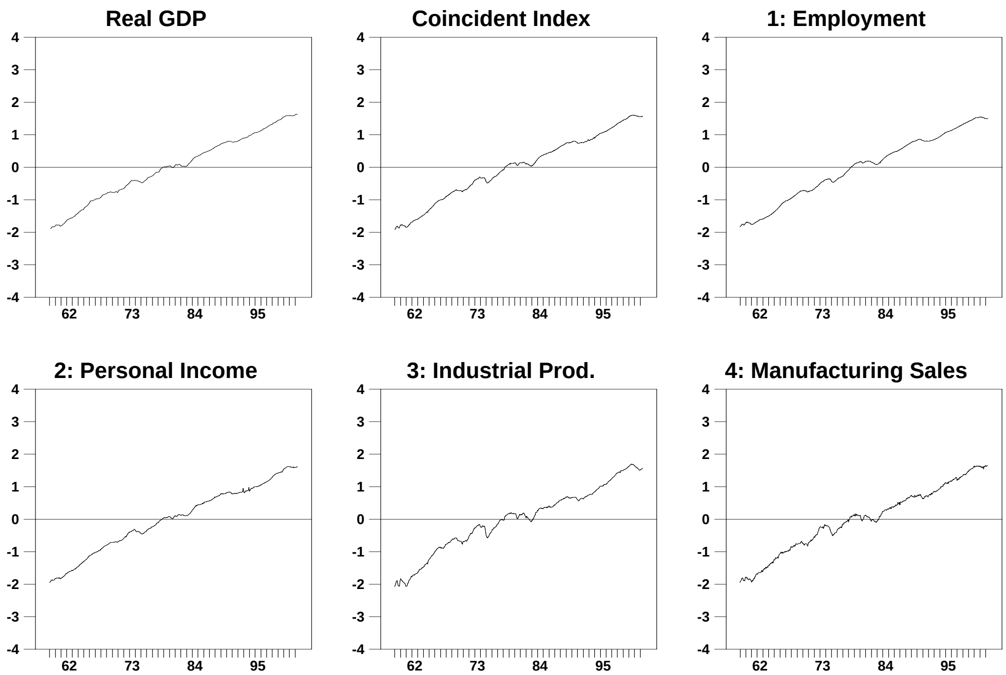

Figure 2:

Differenced-Logs of Quarterly Real GDP, the Monthly Coincident Index, and 4 Monthly Coincident Indicators, January 1959 to June 2002

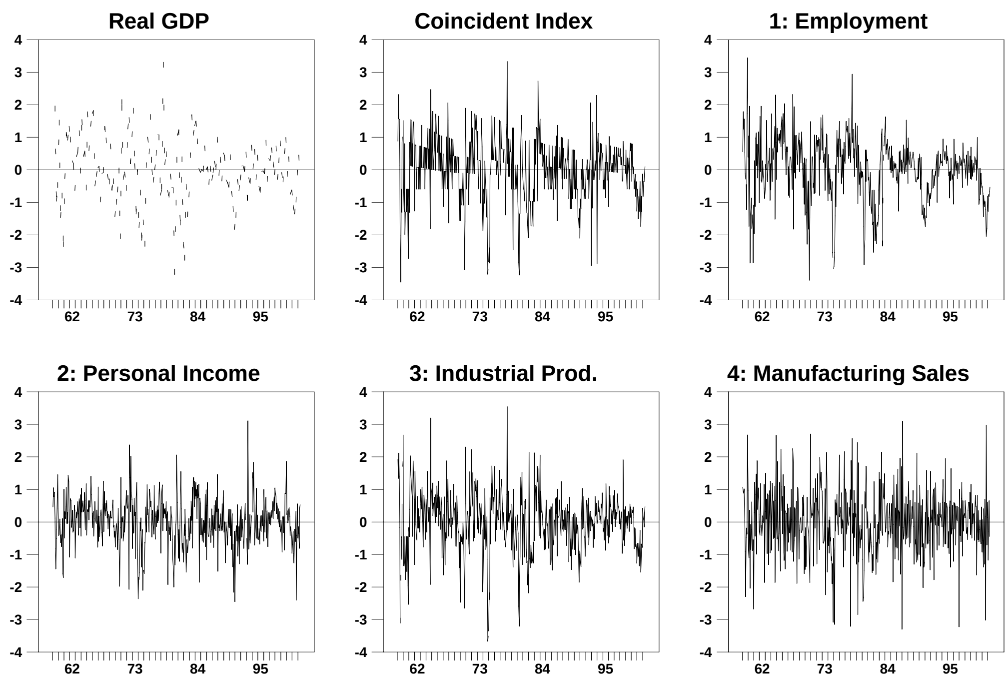

Figure 3:

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

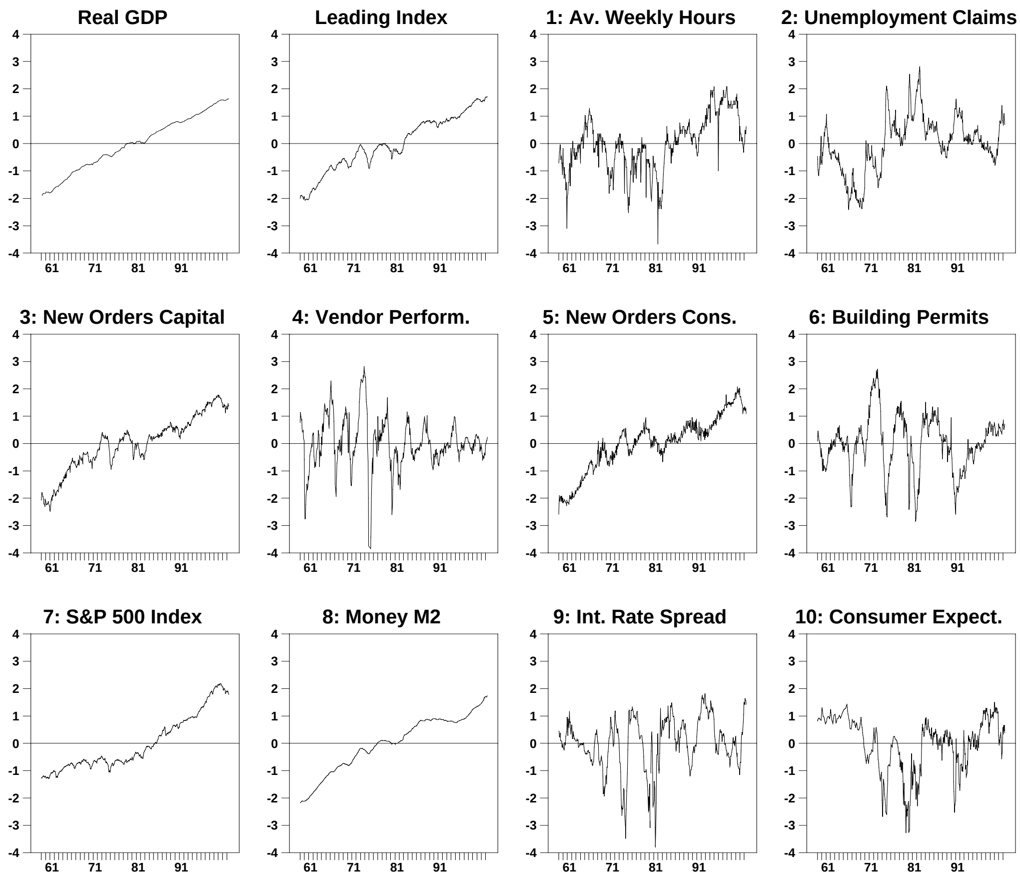

Figure 4:

Differenced-Logs of Quarterly Real GDP, the Monthly Leading Index, and 10 Monthly Leading Indicators, January 1959 to June 2002

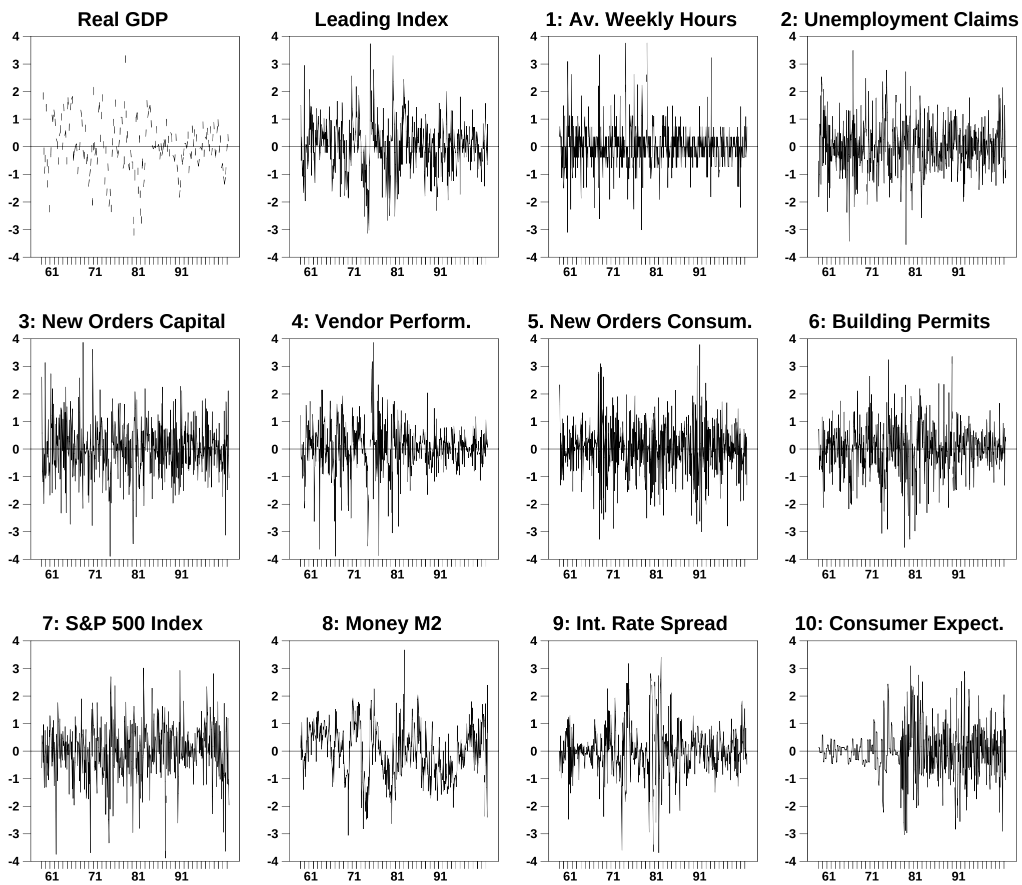

Table 1a-b:

Logs of 4 Monthly Coincident Indicators and Quarterly GDP: Estimated VIMA(1,1) Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly 5-Variable VIMA(1,1) Data Model
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 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6

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

Months Ahead of GDP Forecasts	RMSE
1	.009
2	.015
3	.018
6	.034
12	.066
18	.098
24	.130
Average 1-24	.070

401 in-sample months from January 1959 to June 1992.

120 out-of-sample months from July 1992 to December 2002.

Degrees of freedom = 401 - no. of est. params. = 401 - 40 = 361.

Table 1c-d:

Logs of 4 Monthly Coincident Indicators and Quarterly GDP:
WCFD Derived IMA(1,1) GDP Model and Its Forecast Accuracy

c. WCFD of 5-Variable VIMA(1,1) Data Model for h = 12 Months

WCFD	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 IMA(1,1) GDP Model Based on WCFD 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 of GDP Forecasts	RMSE
1	.018
2	.018
3	.018
6	.034
12	.067
18	.099
24	.130
Average 1-24	.074

WCFD cases:

1a: W = (1, ... , 1) and no AR roots are filtered out;

1b: W = (1, ..., 1) and |AR| AR roots > .98 are filtered out;

2a: W = (0, ..., 0, 1) and no AR roots are filtered out;

2b: W = (0, ..., 0, 1) and no |AR| roots > .98 are filtered out.

Table 2a-b:

Logs of 10 Monthly Leading Indicators and Quarterly GDP:
Estimated VIMA(1,1) Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly 11-Variable VIMA(1,1) Data Model
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 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6
GDP std. devs.: **in-sample .792, out-of-sample .252**

Months Ahead of GDP Forecasts	RMSE
1	.028
2	.040
3	.037
6	.049
12	.077
18	.109
24	.139
Average 1-24	.083

401 in-sample months from January 1959 to June 1992.

120 out-of-sample months from July 1992 to December 2002.

Degrees of freedom = 401 - no. of est. params. = 401 - 187 = 214.

Table 2c-d:

Logs of 10 Monthly Leading Indicators and Quarterly GDP: WCFD Derived IMA(1,1) GDP Model and Its Forecast Accuracy

c. WCFD of 11-Variable VIMA(1,1) Model for h = 12 Months

WCFD	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 IMA(1,1) GDP Model Based on WCFD 2a 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 of GDP Forecasts	RMSE
1	.035
2	.035
3	.035
6	.052
12	.085
18	.117
24	.148
Average 1-24	.092

WCFD cases:

1a: W = (1, ... , 1) and no AR roots are filtered out;

1b: W = (1, ..., 1) and |AR| AR roots > .98 are filtered out;

2a: W = (0, ..., 0, 1) and no AR roots are filtered out;

2b: W = (0, ..., 0, 1) and no |AR| roots > .98 are filtered out.

Table 3a-b:

Diff-Logs of 4 Monthly Coincident Indicators and Quarterly GDP:
Estimated VAR(1) Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly 5-Variable 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 Monthly GDP Forecast Accuracy
for 1992:7 to 2002:6
GDP std. devs.: in-sample 1.09, out-of-sample .581

Months Ahead of GDP Forecasts	RMSE
1	.557
2	.569
3	.635
6	.636
12	.640
18	.636
24	.635
Average 1-24	.630

401 in-sample months from January 1959 to June 1992.

120 out-of-sample months from July 1992 to December 2002.

Degrees of freedom = 401 - no. of est. params. = 401 - 40 = 361.

Table 3c-d:

Diff-Logs of 4 Monthly Coincident Indicators and Quarterly GDP:
WCFD Derived ARMA(2,1) GDP Model and Its Forecast Accuracy

c. WCFD of 5-Variable Data Model for h = 12 Months

WCFD	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 ARMA(2,1) GDP Model Based on WCFD 2a
and Its Forecast Accuracy for 1992:7 - 2002:6

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

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

Months Ahead of GDP Forecasts	RMSE
1	.633
2	.633
3	.633
6	.618
12	.639
18	.636
24	.635
Average 1-24	.635

WCFD cases:

1a: W = (1, ... , 1) and no AR roots are filtered out;

1b: W = (1, ..., 1) and |AR| AR roots > .98 are filtered out;

2a: W = (0, ..., 0, 1) and no AR roots are filtered out;

2b: W = (0, ..., 0, 1) and no |AR| roots > .98 are filtered out.

Table 4a-b:

Diff-Logs of 10 Monthly Leading Indicators and Quarterly GDP: Estimated VAR(1) Data Model and Its GDP Forecast Accuracy

a. Estimated Monthly 11-Variable VAR(1) Data Model
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 Monthly GDP Forecast Accuracy
for 1992:7 - 2002:6
GDP std. devs.: **in-sample 1.09, out-of-sample .581**

Months Ahead of GDP Forecasts	RMSE
1	.596
2	.664
3	.725
6	.649
12	.623
18	.636
24	.636
Average 1-24	.638

401 in-sample months from January 1959 to June 1992.

120 out-of-sample months from July 1992 to December 2002.

Degrees of freedom = 401 - no. of est. params. = 401 - 187 = 214.

Table 4c-d:

Diff-Logs of 10 Monthly Leading Indicators and Quarterly GDP: WCFD Derived ARMA(2,1) GDP Model and Its Forecast Accuracy

c. WCFD of 11-Variable VAR(1) Data Model for h = 12 Months

WCFD	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 ARMA(2,1) GDP Model Based on WCFD 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 of GDP Forecasts	RMSE
1	.632
2	.632
3	.632
6	.619
12	.636
18	.636
24	.636
Average 1-24	.633

WCFD cases:

1a: W = (1, ... , 1) and no AR roots are filtered out;

1b: W = (1, ..., 1) and |AR| AR roots > .98 are filtered out;

2a: W = (0, ..., 0, 1) and no AR roots are filtered out;

2b: W = (0, ..., 0, 1) and no |AR| roots > .98 are filtered out.