WEIGHTED-COVARIANCE FACTOR DECOMPOSITION OF VARMA MODELS
APPLIED TO FORECASTING QUARTERLY U.S. REAL 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
4600 Silver Hill Road
Washington, DC 20233
e-mail: baoline.chen@bea.gov

October 2, 2019

Key words: principal-components-type decomposition of multivariate time-series models
JEL codes: C33, C53

ABSTRACT

Suppose a vector autoregressive moving-average (VARMA) model is estimated for m observed variables of primary interest for an application and n−m observed secondary variables to aid in the application. An application indicates the variables of primary interest but usually only broadly suggests secondary variables that may or may not be useful. Often, one has many potential secondary variables to choose from but is unsure which ones to include in or exclude from the application. The paper proposes a method called weighted-covariance factor decomposition (WCFD), comparable to Stock and Watson's (2002a,b) method here called principle-components factor decomposition (PCFD), for reducing the secondary variables to fewer factors in order to obtain a parsimonious estimated model that is more effective in an application. The WCFD method is illustrated in the paper by forecasting quarterly-observed U.S. real GDP at monthly intervals using monthly-observed 4 coincident and 8 leading indicators from the Conference Board (2018). The results show that root mean-squared errors of GDP forecasts of PCFD-factor models are 0.9%-11.3% higher than those of WCFD-factor models especially as estimation-forecasting periods pass from the pre-2007 Great Moderation through the 2007-2009 Great Recession to the 2009-2016 Slow Recovery.

∗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. Forthcoming in Journal of Time Series Analysis. We thank the co-editor, Marcus Chambers, and referees of the journal for their helpful 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.
1. Introduction.

Parsimony means minimizing the number of estimated parameters of a model while maintaining its fit of data. Parsimony is desired because, being subject to less sampling variability, parsimonious models are expected to be more effective in applications. Because variables bring estimated parameters into a model, minimizing the number variables in a model furthers its parsimony. Suppose a vector autoregressive moving-average (VARMA) model is estimated in order to forecast a few \((m)\) variables of primary interest using many \((n-m)\) secondary variables to aid in the forecasting. An application indicates the variables of primary interest but usually only broadly suggests secondary variables that may or may not be useful in the application. Often, one has many potential secondary variables to choose from but is unsure which ones to include in or exclude from the application. To further parsimony, somehow the number secondary variables needs to be minimized. Stock and Watson (2002a,b) proposed a method based on standard principal components analysis (PCA) for replacing possibly many secondary variables with fewer "factors".

We call Stock and Watson's method principal-components factor decomposition (PCFD), which has the following steps: (1) uses eigenvectors of the sample contemporaneous covariance matrix of secondary variables to linearly transform them to factors; (2) based on experience and experimentation (including using information criteria), decides which factors are significant; (3) estimates an AR equation of primary variables and significant factors; and, (4) uses the estimated AR equation to forecast the primary variables.

By eigenvalue decomposing the sample contemporaneous covariance matrix of secondary variables, PCFD uses no lagged sample correlations (no dynamic sample information) among secondary variables and no contemporaneous or lagged sample correlations between primary and secondary variables to compute factors. PCFD could include dynamic sample information by including covariances of lagged secondary variables in the covariance matrix to be decomposed, but, whether the variables are significantly serially correlated or not, the added covariances are likely to be redundant and to degrade the decomposition and lead to poorer results, as reported by Stock and Watson (2002b). PCFD accounts only for variations of secondary variables in its factorization, so that it may put the "noise" of irrelevant secondary variables into factors and resulting forecasts because it ignores correlations between primary and secondary variables.

The paper develops a corresponding method called weighted covariance factor decomposition (WCFD) that corrects these disadvantages and illustrates
the method with U.S. data on quarterly-observed real GDP as the single primary variable and monthly-observed 4 coincident and 8 leading indicators as the secondary variables. By contrast with PCFD, by eigenvalue decomposing an estimated covariance matrix of forecast errors of primary variables (at some chosen number of forecast periods ahead), WCFD can use all significant sample correlations (all significant dynamic sample information) in terms of an initial estimated VARMA model to compute factors. By eigenvalue decomposing the "weighted covariance" matrix of forecast errors of primary variables, WCFD accounts for correlations between primary and secondary variables in its factorization. These advantages are borne out in the application in the paper that reports root mean-squared errors of PCFD-based forecasts being up to 11.3% higher than those of WCFD-based forecasts. We have not seen in the statistics and econometrics literatures any similar PCA-like decompositions based on weighted covariances, although an online search of "weighted covariance PCA" returns numerous papers in various fields of science. For example, Delchambre (2015) develops and illustrates a modified standard PCA based on weighted covariances motivated by the same above aim of down-weighting irrelevant or less-relevant observations in order to minimize the impact of noise.

Corresponding to PCFD, WCFD has the following steps: (1) estimates an initial VARMA model of primary and secondary variables; (2) uses the initial model to compute factors that account for variations in primary variables; (3) based on experience and experimentation (including using information criteria), decides which factors are significant; (4) estimates a smaller VARMA model of primary variables and significant factors; and, (5) forecasts primary variables using the second estimated model.

Hotelling (1933) developed PCA to its modern form (Anderson, 1984, ch. 11). Let \( \{C_k\}_{k=0}^\infty \) 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 \( C_k \). Like classical PCA, Stock and Watson's (2002a,b) PCFD also uses only \( \hat{C}_0 \) to produce factors. By contrast, by using Fourier-transformed \( \{\hat{C}_k\}_{k=0}^K \) for a finite \( K \), Forni and Reichlin (1998) and Forni et al. (2000) use dynamic sample information to produce factors. The present paper uses the dynamic sample information \( \{\hat{C}_k\}_{k=0}^K \) in the form of an initially estimated VARMA model to produce factors.
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 parameters, because MLE is demanding computationally. By the end of the 1990s, despite large reductions in computing costs, interest shifted to applying PCA to hundreds of observed financial and macroeconomic variables to produce small dynamic factor models (Stock and Watson, 2002a,b; Forni et al., 1998, 2000).

More recently, Bernanke et al. (2005) introduced factor-augmented VAR (FAVAR) models, which have been extended to factor-augmented VARMA (FAVARMA) models (Dufour and Stevanovic, 2013). FAVARMA models are basically sums of unobserved VARMA terms with some terms being considered "dynamic factors", so that FAVARMA models are basically VARMA models with particular functional restrictions on parameters. In the present application, we initially estimated both VAR and VARMA models with zero and functional restrictions, but gave up on them and do not report their results, because they fit the mixed-frequency data (MFD) poorly and produced poor forecasts. Therefore, all reported results of the application in Table 1 are for unrestricted VAR models.

Although the literature cited above and the present application 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).

PCA and PCFD is strictly meaningful only if $\hat{C}_0$ exists, if the data generating process is stationary, so that $\hat{C}_0$ converges stochastically to $C_0$ as the number of sample periods goes to infinity. PCA/PCFD can be computed more accurately if $\hat{C}_0$ is positive definite, which occurs in practice if the data have more sample periods (T) than variables (n) and no variables exactly satisfy linear equations. Additional assumptions have often been made in terms of the approximate factor model (AFM), so that PCA/PCFD provides consistent estimates of AFM parameters, principally bounds on eigenvalues as T and n go to infinity (Bai and Ng, 2002; Stock and Watson, 2002a,b; Doz et al., 2012). Because the present paper conducts no asymptotic analysis, T and n are assumed to be finite and fixed.

The paper illustrates WCFD by forecasting U.S. quarterly-observed real GDP at monthly intervals using monthly-observed 4 coincident and 8 leading indicators as secondary variables. Although WCFD has nothing per se to do with
MFD, using MFD makes the application more realistic, challenging, and interesting. Following Zadrozny (1990) and Mittnik and Zadrozny (2004), Kalman-filtering-based MLE is used in the application to estimate VAR models with the MFD. Computational constraints still limit how many VAR parameters MLE can successfully estimate, although in the application using a FORTRAN program we were able to successfully estimate up to 260 VAR parameters in about 30 minutes, starting with setting all parameter values to .01 or .001. When this did not work, when the program did not converge in an acceptable amount of time or stopped at unreasonable parameter values, significant additional time was spent retrying the estimation from other, largely randomly selected, neighboring starting parameter values until convergence or near convergence at reasonable parameter values was achieved. These difficulties with MLE are considerably greater when using MFD than when using single-frequency data.

An alternative extended Yule-Walker method (Chen and Zadrozny, 1998) for estimating a VAR model with MFD computes as easily and quickly as any linear estimation method but was not used because it has not yet been tested thoroughly enough. Also, by optimizing over disturbance covariances as well as over VAR coefficients, MLE tends to get closer to minimal Kullback-Leibler information, hence, to the true data generating process (Bowden, 1973). For these reasons, all models in the application were estimated using MLE.

WCFD also provides a purely-data-based and economic-theory-free variance decomposition of a model, although this aspect of WCFD is not pursued here. Sims (1980a,b) advocated computing variance decompositions of estimated VAR models to judge explanatory power of one variable over another. Initially, Sims advocated variance decompositions based on Cholesky decomposition, which is a purely numerical method. 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 initial estimated 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 WCFD decomposition matrix $R$ is the counterpart of the Cholesky and structural decomposition matrices in this literature.

The paper proceeds as follows. Section 2 explains the WCFD decomposition. Section 3 discusses deciding which WCFD factors are significant. Section 4 reviews the PCFD decomposition and factorization steps and explains the WCFD factorization step. Section 5 applies PCFD and WCFD to U.S. data from 1959 to 2018 to forecast quarterly-observed real GDP at monthly intervals using monthly-observed 4 coincident and 8 leading indicators and evaluates the forecast
accuracies. Section 6 concludes by summarizing the paper and discussing estimating an initial "diagonal" VARMA model when too many variables result in too many estimated parameters in a "nondiagonal" VARMA model. The Appendix discusses two closely related methods for computing the WCFD decomposition.

2. WCFD decomposition.

In what follows all quantities are real except possibly complex AR and MA characteristic roots. Let \( y_t \) denote an \( n \times 1 \) vector of stationary sample-mean-adjusted variables observed in periods \( t = 1, \ldots, T \), presumed to be generated by a stationary VARMA model,

\[
(2.1) \quad y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \xi_t + B_1 \xi_{t-1} + \ldots + B_q \xi_{t-q},
\]

where \( \xi_t \) is an \( n \times 1 \) vector of unobserved innovations, distributed normally, identically, independently, with zero means and positive definite covariance matrix \( \Sigma_\xi \) or \( \xi_t \sim NIID(0, \Sigma_\xi) \), where \( 0 \times 1 \) denotes the \( n \times 1 \) zero vector.

Partition \( y_t = (y_{1t}^T, y_{2t}^T)^T \), where \( y_{1t} \) contains \( m \) primary variables, \( y_{2t} \) contains \( n-m \) secondary variables, and superscript \( T \) denotes transposition.

Model (2.1) is stated more concisely in terms of lag operator \( L \) as \( A(L)y_t = B(L)\xi_t \), where \( A(L) = I_n - A_1 L - \ldots - A_p L^p \), \( B(L) = I_n + B_1 L + \ldots + B_q L^q \), and \( I_n \) denotes the \( n \times n \) identity matrix. Often, some or all elements of reduced-form parameter matrices \( A_i \), \( B_i \), and \( \Sigma_\xi \) are restricted to zero or in terms of fewer structural parameters. To compute WCFD, we need to know only the values of the reduced-form parameters and don't need to know any structural restrictions on them.

We define AR and MA characteristic roots as follows. Model (2.1) is stationary if and only if the absolute characteristic AR roots are \( < 1 \), namely, if and only if \( \det[I_n \lambda^p - A_1 \lambda^{p-1} - \ldots - A_p - A_p] = 0 \) implies \( |\lambda| < 1 \), where \( \det[\cdot] \) denotes the determinant of a square matrix. Model (2.1) is invertible if and only if the absolute characteristic MA roots are \( < 1 \), namely, if and only if \( \det[I_n - B_1 \lambda^{q-1} - \ldots - B_q - B_q] = 0 \) implies \( |\lambda| < 1 \). Zadrozny (2016) stated
assumptions, including those above, and proved that under them a VARMA model is locally identified with single- or mixed-frequency data.

If model (2.1) is stationary, then, it has a unique Wold infinite moving-average representation,

\[ y_t = \Psi(L) \xi_t = \left( \sum_{j=0}^{\infty} \Psi_j L^j \right) \xi_t = \sum_{j=0}^{\infty} \Psi_j \xi_{t-j}, \]

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

\[ \Psi_j = \sum_{\ell=0}^{\min(j,p)} A_{\ell j} \Psi_{j-\ell} + B_j, \]

for \( j = 1, \ldots, J \), starting with \( \Psi_0 = I_n \), such that \( B_j = 0 \) for \( j > q \). WCDF uses Wold decomposition (2.2) to decompose \( \Sigma_\xi \) as \( RR^T = \Sigma_\xi \), such that the columns of \( R \) account for variations of primary variables in a PCA-like fashion.

In nonstationary or near-nonstationary models, especially when the forecasting horizon \( h \) is large, the WCDF decomposition could "latch" onto the largest exactly or nearly nonstationary AR root (the latter defined, say, by \( .99 < |\lambda_i| < 1 \)) and account for nearly 100\% of the weighted covariances with one factor, a result that is not useful when it masks significant subdominant cycles of a model. In such cases, we could filter out the exactly or nearly nonstationary AR roots before computing WCDF. Let \( \{\hat{\lambda}_i\}_{i=1}^v \) denote the \( v \) exactly or nearly nonstationary AR roots of model (2.1). We would filter Wold representation (2.2) using \( \hat{\lambda}(L) = (1-\hat{\lambda}_1 L) \cdots (1-\hat{\lambda}_v L) \), obtain \( \hat{y}_t = \hat{\Psi}(L) \xi_t \), where \( \hat{y}_t = \hat{\lambda}(L) y_t \) and \( \hat{\Psi}(L) = \hat{\lambda}(L) \Psi(L) \), and would compute the WCDF decomposition using \( \hat{\Psi}(L) \). Exactly or nearly nonstationary AR roots would be filtered out of \( \Psi(L) \) only to compute the WCDF decomposition but would stay in \( \Psi(L) \) to compute the WCDF factorization.

For a given forecast horizon \( h \geq 1 \), let \( \eta_{ht} = y_{t+h} - E_t y_{t+h} \) denote the \( nx1 \) vector of errors from forecasting \( y_{t+h} \) in period \( t \). In terms of innovations, forecast errors are
\[(2.4) \quad \eta_{ht} = \sum_{j=0}^{h-1} \Psi_j \xi_{h+j}, \]

and have covariance matrix

\[(2.5) \quad \Gamma_h = E\eta_{ht}\eta_{ht}^T = \sum_{j=0}^{h-1} \Psi_j \Sigma_{\xi} \Psi_j^T. \]

We define WCFD in terms of weighted covariances of forecast errors of primary variables in \(y_{ht}\). The primary variables are mapped from all variables as \(y_{ht} = Wy_t\), where \(W\) is a specified \(m \times n\) weighting matrix and \(1 \leq m = \text{rank}(W) \leq n\).

Then, \(v = E\eta_{ht}^T \Omega \eta_{ht} = tr[\Omega \Gamma_h]\) is the expected weighted h-period-ahead squared forecast error of the primary variables, where \(\Omega = W^T W\) represents the weights as an \(n \times n\) symmetric positive semi-definite matrix and \(tr[\bullet]\) denotes the trace of a matrix. Using equation (2.5),

\[(2.6) \quad v = tr[\Omega \sum_{j=0}^{h-1} \Psi_j \Sigma_{\xi} \Psi_j^T]. \]

The following three examples illustrate different weighting matrices:

**Selecting portfolio value as primary:** Let \(y_{ht}\) = value of a portfolio, \(y_{2t}\) = (n-1)x1 vector of portfolio elements, and \(W = (0, w_2, \ldots, w_n) = 1 \times n\) vector of portfolio weights, such that a portfolio weight is positive or negative depending on whether its portfolio element is an asset or a liability. Portfolio value \(y_{ht}\) is the single primary variable and portfolio elements in \(y_{2t}\) are the secondary variables. Other secondary variables can be added to \(y_{2t}\).

**Selecting J of n variables as primary:** Let \(\{i_j\}_{j=1}^J \in \{1, \ldots, n\}\) index a subset of \(J\) of \(n\) variables selected as primary variables and let \(e_{i_j} = (0, \ldots, 0, 1, 0, \ldots, 0)^T\) be the \(n \times 1\) vector with 1 in position \(i_j\) and zeros elsewhere. Stack row-form \(e_{i_j}^T\), for
\( j = 1, \ldots, J, \) on top of each other in any order and obtain the \( J \times n \) weighting matrix \( W \) that selects the \( J \) variables from \( y_t \) as the primary variables in \( y_t \).

**Setting WCFD equivalent to PCFD:** If \( W = I_n \), the initial VARMA model is stationary, and forecast horizon \( h \) is large (strictly, \( h = \infty \)), then, \( v = \) sum of equally-weighted variances of all variables in \( y_t \) and WCFD is equivalent to PCFD, except for numerical differences due to different computations.

WCFD produces the \( n \times n \) decomposition matrix \( R \) that satisfies \( RR^T = \Sigma_\xi \) and broadly corresponds to a factor-loading matrix in PCA/PCFD. For \( r_i = \) column \( i \) of \( R \), equation (2.6) can be written equivalently as \( v = \sum_{i=0}^{n} v_i \), where

\[
(2.7) \quad v_i = r_i^T Q r_i
\]

is the portion of the weighted \( h \)-step-ahead forecast-error variance, \( v \), accounted for by the \( i \)th uncorrelated disturbance, \( \varepsilon_t \), and \( Q = \sum_{j=0}^{h-1} \psi_j^T \Omega^j \psi_j^T \). \( Q \) is symmetric positive semi-definite, but may be nonpositive definite even for large \( h \). If so, a computed WCFD decomposition may be inaccurate, but can be made accurate by making \( Q \) positive definite by adding \( \delta I_n \) to it, where \( \delta \) is a small positive number.

We define the WCFD decomposition recursively. Let \( \Sigma_1 = \Sigma_\xi \), \( X_1 = 0_{n \times 1} \), and, for \( i = 2, \ldots, n-1 \), \( \Sigma_i = \Sigma_\xi - X_i X_i^T \), \( X_i = [r_{i-1}, \ldots, r_{i-1}] \), and \( Y_i = [r_{i-1}, \ldots, r_{i}] \), a matrix of slack variables that ensures that \( r_i^T + Y_i^T \Sigma_i = \Sigma_i \) holds. For \( i = 1, \ldots, n-1 \), we want to maximize \( v_i = r_i^T Q r_i \), with respect to \( r_i \) and \( Y_i \), so that \( r_i^T + Y_i^T \Sigma_i = \Sigma_i \) holds and, for \( i = n \), we want \( X_n X_n^T + r_n r_n^T = \Sigma_\xi \) to hold.

First, given \( \Sigma_1 = \Sigma_\xi \) and \( Q \), we maximize \( v_1 \) with respect to \( r_1 \) and \( Y_1 \), subject to \( r_1^T + Y_1^T = \Sigma_1 \), by eliminating \( Y_1 \) from first-order conditions (FOC; see the Appendix) 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 \( r_2^T + Y_2^T = \Sigma_2 \),
by eliminating $Y_2$ from the FOC and solving for $r_2$ and $v_2$. Continuing like this, we determine $X_n = [r_1, ..., r_{n-1}]$. Finally, given $X_n$ and $\Sigma_n = \Sigma_y - X_n X_n^T$, we determine $r_n$ such that $r_n r_n^T = \Sigma_n$ and $R R^T = \Sigma_y$ hold for $R = [r_1, ..., r_n]$.

We assume that (i) $\Sigma_y$ is positive definite; (ii) $1 \leq m = \text{rank}(W) \leq n$; and, (iii) the positive eigenvalues of $Q$ are distinct. $Q$ is symmetric positive semi-definite by construction and nonnull by assumption (ii). Then, it follows from the discussion in the Appendix that the significant part of the WCFD decomposition, associated with the significantly positive eigenvalues of $\Sigma_y Q$, exists and is unique. In practice, assumptions (i) and (iii) hold when variables are not subject to linear restrictions. The details for computing the WCFD decomposition are in the Appendix.


Because each PCFD and WCFD factor has an associated eigenvalue, deciding which PCFD and WCFD factors are significant amounts to deciding which PCFD and WCFD eigenvalues are significantly different from zero. This section first considers a stochastic test for this decision and, then, reduces it to the simpler nonstochastic test that is used in the application.

First, consider the following variant of Anderson’s (1984, ch. 11, pp. 473-475) test for significance of the largest PCA eigenvalues based on first-order normal approximation. For given tolerance $\rho \in (0,1)$, let $H_k$ denote the hypothesis that the largest eigenvalues $\lambda_i$ of $\Sigma_y Q$, for $i = 1, ..., k$, account for at least $1-\rho$ of weighted covariances, so that $H_k$ is true if and only if

$$
\delta_k = \bar{\rho}^T \bar{v} = -\rho \sum_{i=1}^k v_i + (1-\rho) \sum_{i=k+1}^n v_i \leq 0,
$$

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

Consider the following testing sequence. Start with $\ell = 1$ and test $H_1 = H_1$. If $H_1$ is accepted (strictly, not rejected), accept $k = 1$ as the number of significant largest eigenvalues of $\Sigma_y Q$ and associated factors; otherwise, test
If $H_2$ is accepted, accept $k = 2$ as the number of significant largest eigenvalues and associated factors; otherwise, continue like this until possibly reaching $\ell = n-1$. If $H_{n-1}$ is accepted, accept $k = n-1$ as the number of significant largest eigenvalues and associated factors; otherwise, accept $k = n$ as the number of significant largest eigenvalues and associated factors. Because $v_i \geq 0$ and $\sum_{i=1}^{n} v_i = 1$, the testing sequence is always conclusive.

WCFD maps $\phi$ to $\bar{V}$ non-linearly and differentiably as $\bar{V} = \bar{V}(\phi)$, where vector $\phi$ collects parameters of the initial VARMA model. If $\phi$ is certain, e.g., because it was chosen for a hypothetical model, then, $H_k$ is accepted if and only if $\delta_k \leq 0$. However, generally, $\phi$ is estimated and uncertain, so that $\delta_k$ inherits sampling variability from estimated parameters, so that $H_k$ should be tested stochastically. This can be done by expanding $\delta_k \leq 0$ to a stochastic statement similar to Anderson’s (1984, ch. 11, pp. 473-475) asymptotic test for the number of significant largest PCA eigenvalues, as follows.

If VARMA model (2.1) is stationary and invertible and additional assumptions hold (Hosoya and Taniguchi, 1982), $\sqrt{T}(\hat{\phi} - \phi_0) \sim AN(0, S_0)$, where a hat (‘) denotes an estimated value, subscript zero denotes a true value, and $\sim$ AN denotes an asymptotic normal distribution as the number of sample periods, $T$, goes to infinity. Under further assumptions (Serfling, 1980, pp. 122-124), $\sqrt{T}(\hat{\nu} - \nu_0) \sim AN(0, \nabla\hat{\nu}S_0\nabla\hat{\nu}^T)$, where $\nabla\hat{\nu}$ denotes the Jacobian matrix of first-partial derivatives of $\bar{V}(\phi)$ evaluated at $\hat{\phi}$. Then, because $\bar{\nu}$ is constant,

$$\sqrt{T}(\hat{\delta}_k - \delta_{k_0}) \sim AN(0, \hat{\sigma}^2_{\delta_k}),$$

where $\hat{\sigma}^2_{\delta_k} = \bar{\rho}' \nabla\hat{\nu} S \nabla\hat{\nu}' \bar{\rho}$, $\hat{\delta}$ estimates $S_0$ such that $\lim_{T \to \infty} \hat{S} = S_0$, and $\nabla\hat{\nu}$ and $\hat{\sigma}^2_{\delta_k}$ can be computed using the matrix-differentiation method in Mittnik and Zadrozny (1993) and Chen and Zadrozny (2003, Appendix A).

Using equation (3.1) and following standard sampling-theory testing, let $\alpha \in (0,1)$ denote a chosen significance level and let $c_\alpha$ denote a critical
value defined by \( \text{Prob}[z \leq c_\alpha] = 1 - \alpha \), where \( z \sim N(0,1) \). Then, for given \( k, \rho \), and \( \alpha \), \( H_k \) is accepted if and only if

\[
(3.3) \quad \tau_k = \frac{\hat{\delta}_k}{\hat{\sigma}_k} \leq c_\alpha.
\]

Test (3.3) could mislead if \( \hat{\delta}_k \) is not asymptotically normally distributed. For such cases, Onatski's (2009, 2010) nonnormal eigenvalue distributions or Bai and Ng (2002) and Bai's (2003) information criteria could be used. The present application ignores uncertainty about \( \hat{\delta}_k \) and decides nonstochastically on the significant largest eigenvalues and associated factors, based on experience and experimentation (including using information criteria), as did Stock and Watson (2002b). When ignoring the uncertainty about \( \hat{\delta}_k \), \( c_\alpha = 0 \) and test (3.3) reduces to \( \hat{\delta}_k \leq 0 \). The uncertainty can be ignored in the forecasting application because any choice of significant factors can and should be reviewed in terms of forecast accuracy.

In PCFD and WCFD, for a given tolerance \( \rho \), using test (3.3) stochastically or nonstochastically, one chooses as significant the smallest number of largest eigenvalues and associated factors that, respectively, account for at least \( 1 - \rho \) of the sum of variances of secondary variables and at least \( 1 - \rho \) of the sum of weighted covariances of forecast errors of primary variables. In a particular application, experimentation can determine which value of \( \rho \) leads to the best choice of significant factors with the best results. In the present application, experimentation showed that \( \rho = .10 \) led to the best choice of significant factors, 3 in PCFD and 2 in WCFD.

4. PCFD and WCFD factorizations.

For simplicity, this section proceeds with one primary variable in \( y_t \). The generalization to m primary variables is straightforward but notationally more tedious. With one primary variable, weighting matrix \( W \) is a 1x\( n \) row
vector and significant-factor vector $f_{1t}$ is a k×1 column vector; with $m > 1$
primary variables, $W$ and $f_{1t}$ are m×n and k×m matrices.

The PCFD factorization is obtained as follows. Suppose that the
eigenvalues of the sample covariance matrix of the n-1 secondary variables in
$y_{2t}$ are distinct, so that the (n-1)×(n-1) matrix $M$ of the eigenvectors is
orthogonal. Then, the n-1 PCFD factors in $f_t = (\varphi_{1t},...,\varphi_{n-1,t})^T$ are defined and
computed by

$$(4.1) \quad f_t = M^T y_{2t}.$$ 

Partition $f_t = (f_{1t}^T, f_{2t}^T)^T$, where $f_{1t} = (\varphi_{1t},...,\varphi_{k,t})^T$ and $f_{2t} = (\varphi_{k+1,t},...,\varphi_{n-1,t})^T$
contain k significant and n-k-1 insignificant factors, and, correspondingly,
partition $M = [M_1, M_2] = [(n-1)\times k, (n-1)\times(n-k-1)]$, so that equation (4.1) has
the regression form

$$(4.2) \quad y_{2t} = M_1 f_{1t} + u_t,$$

where $M_1 = (n-1)\times k$ matrix of regression coefficients, $f_{1t}$ is k×1 vector of
regressors, and $u_t = M_2 f_{2t} = (n-1)\times 1$ vector of regression errors. Because $f_{1t}$
and $u_t$ are orthogonal and, therefore, uncorrelated, regression (4.2) satisfies
this basic regression assumption.

The WCFD equation corresponding to (4.2) is obtained as follows. The WCFD
decomposition of an initial VARMA model produces the n×n nonsingular
decomposition matrix $R$ whose inverse maps the model's innovations to
uncorrelated disturbances, $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{m})^T = R^{-1} \xi_t = R^{-1}(\xi_{1t},...,\xi_{m})^T$, for $t = 1,...,T$. The initial VARMA model is presumably stationary and has the Wold
representation (2.2). Using $R \varepsilon_{t-j} = \sum_{i=1}^{n} r_{i} \varepsilon_{i,t-j}$, where $R = [r_1,...,r_n]$, the Wold
representation can be written as $y_t = \sum_{i=1}^{n} \varphi_{i,t} \varepsilon_{i,t}$, where $\varphi_{i,t} = \sum_{j=0}^{\infty} \Psi_{j} r_{i} \varepsilon_{i,t-j}$, so that
primary $y_{1t} = Wy_t$ has the k-significant-factor representation
\[ y_{it} = \sum_{j=1}^{k} \varphi_{i,j} + \tilde{u}_t, \]

where, for \( i = 1, \ldots, k \), \( \varphi_{i,j} = \sum_{j=1}^{\infty} W^j \varphi_{i,j} \) = \( i \)th significant WCFD factor and \( \tilde{u}_t \) = \( \sum_{i=k+1}^{n} \varphi_{i,j} \) = sum of insignificant WCFD factors. The \( k \)-significant-factor representation (4.3) has the regression form

\[ y_{it} = \tilde{M}_t \tilde{f}_{it} + \tilde{u}_t \]

corresponding to (4.2), where \( \tilde{M}_t = (\sigma_1^2, \ldots, \sigma_k^2) = 1 \times k \) vector of regression coefficients, \( \sigma_i^2 = E(\varphi_{i,i})^2 \), \( \tilde{f}_{it} = (\tilde{\phi}_{i,1}, \ldots, \tilde{\phi}_{i,k})^T = k \times 1 \) vector of regressors, \( \tilde{\phi}_{i,j} = \varphi_{i,j} / \sigma_i^2 \), and \( \tilde{u}_t = \sum_{i=k+1}^{n} \phi_{i,j} \) = regression error. Because \( \tilde{f}_{it} \) and \( \tilde{u}_t \) are uncorrelated, because \( \varepsilon_s \) and \( \varepsilon_t \) are uncorrelated for \( i \neq j \) and \( s \neq t \), regression (4.4) satisfies this basic regression assumption.

WCFD factors are computed for an initial estimated VARMA model and data in the following 4 steps:

**Step 1:** Innovations cannot be computed for a sample period if any data in the period are missing. In the application, missing data occur because models are estimated using monthly-quarterly MFD and because outlying observations are treated as missing. The missing data were filled in by applying Kalman smoothers (Anderson and Moore, 1979) based on initial (estimated) models. The monthly indicators also have missing data due to outliers that also need to be filled in by smoothing before either PCFD or WCFD factors can be computed. Stock and Watson (2002b) similarly suggested filling in missing data in PCFD by using the expectation-maximization (EM) algorithm (Dempster et al., 1977).

**Step 2:** If no data are missing in any period or once any missing data have been filled in, then, except for some initial periods, innovations can be computed by iterating on definition (2.1) of a VARMA model. Some initial innovations cannot be computed in this way because the data start in period \( t = 1 \). In the application, this problem is solved by applying a Kalman filter (Anderson and Moore, 1979) based on an initial VAR model. Specifically, for each sample period
t = 1, ..., T, the filter computes $y_{t-1}$ = forecast of $y_t$ made in period $t-1$, whereupon the innovation in period $t$ is estimated as $\xi_t = \hat{y}_t - y_{t-1}$, where $\hat{y}_t$ denotes either observed $y_t$ or its smoothed estimate. Both Kalman smoothers and filters were initialized in the application by setting initial state vectors to zero and their covariance matrices to unconditional covariance matrices implied by the initial VAR models.

**Step 3:** For a full set of innovations, $\xi_t$, for $t = 1, ..., T$, with no missing values, uncorrelated disturbances are computed as $\varepsilon_t = R^{-1}\xi_t$, where $R = WCFD$ decomposition matrix.

**Step 4:** Wold coefficient matrices of a model, $\Psi_j$, are computed by iterating on equation (2.3), starting with $\Psi_0 = I_n$. For the $i$th WCFD factor, $\varphi_{i,1,t} = \sum_{j=1}^{\infty} W\Psi_j \varepsilon_{i,t-j}$, to be computed approximately, its infinite sum must be truncated. Because the Wold coefficient matrices of stationary VARMA models decline exponentially to zero as $j$ increases, for every stationary model there is some $J$ at which the sum of WCFD-factor terms beyond $J$ can be considered negligible and ignored. The truncated $J$-term approximate $i$th WCFD factor is

\[(4.5) \quad \varphi_{i,1,t}^{(J)} = \sum_{j=1}^{J} W\Psi_j \varepsilon_{i,t-j}.\]

To compute $\varphi_{i,1,t}^{(J)}$ using truncation (4.5), starting from period $t = 1$, requires getting around not having presample values of $\varepsilon_{ij}$. One option is to assume that presample $\varepsilon_{ij}$ are equal to their zero means; another, used in the application, is to compute and use $\varphi_{i,1,t}^{(J)}$ only for $t \geq J+1$, so that truncation (4.5) doesn't include any presample values of $\varepsilon_{ij}$.

The 4 computational steps appear accurate in the application. First, the Kalman-smoothed GDP estimates replicate observed GDP to within 8-9 decimal digits, i.e., have "single precision" accuracy. Second, because the largest absolute AR root of any estimated model is about .85, for the $J = 50$ used in the
application, elements of the approximation-error matrices, \( \varphi_{ij} - \varphi_{ij,t}^{(50)} \), for \( i = 1, \ldots, k \), are all about \( \pm 0.0001 \), i.e., have "semi-single precision" accuracy. Finally, significant-factor threshold \( \rho = 0.10 \), used in the application for a final choice of \( k = 2 \) significant WCFD factors in every case, suggests that ordinary least squares (OLS) estimates of regression (4.4) should yield \( R^2 \) of GDP \( \cong 0.90 \). However, except in one case, OLS estimates of regression (4.4) yielded \( R^2 \) of GDP \( \cong 0.99 \), with very high \( t \) statistics of estimated coefficients up to about 50. Oddly, the best-fitting factor model, 5/WCF2/VAR4 (see Table 1), had the lowest \( R^2 \) of GDP \( \cong 0.50 \) of regression (4.4), perhaps because its WCFD decomposition was the least accurate to only about 3 decimal digits, whereas the other WCFD decompositions were accurate to about 13-14 decimal digits, i.e., close to "double precision" accuracy.

5. Application to forecasting quarterly GDP at monthly intervals.

5.1. Data sources.

Quarterly-observed U.S. real GDP data from 1959:3 (March) - 2018:6 (June) obtained from the Bureau of Economic Analysis (BEA) are the latest comprehensive revision produced by the Bureau, released in 2018:8. Monthly-observed U.S. coincident and leading indicators from 1959:1 - 2018:6 were obtained from the Conference Board (2018): 4 coincident indicators, (1) number of employees on nonagricultural payrolls, (2) personal income less transfer payments, (3) index of industrial production, and (4) manufacturing and trade sales; and, 8 leading indicators, (5) average weekly hours worked in manufacturing, (6) average weekly initial claims for unemployment insurance benefits, (7) manufacturer's new orders of consumer goods and materials, (8) manufacturer's new orders of nondefense capital goods excluding aircraft, (9) new permits for building private housing units, (10) Standard and Poor's index of 500 stock prices, (11) Conference Board's index of consumer expectations, and (12) interest rate spread (interest on 3-month Treasury bills minus interest on 1-year Treasury notes).

5.2. Data transformations.

Before being used in estimation the data were (1) naturally logged (except the interest rate spread because it has negative values), (2) first-differenced
at their sampling intervals (GDP quarterly; indicators monthly), (3) standardized (sample means subtracted then divided by sample standard deviations), and (4) normalized (outliers more than 3.5 standard deviations from zero treated as missing values). Figures 1-2 graph the data for the whole sample period in log forms (except the interest rate spread) and differenced-log forms (except the interest rate spread that is differenced without logging). (Figure 1 includes the Conference Board’s coincident index, because an even number of graphs displays better.) Each graphed series was standardized using sample means and sample standard deviations of the whole sample period.

The data were transformed by logging and differencing to make them more stationary and easier to estimate models with. Standardization simplified estimation by eliminating the need to estimate constant terms and made MLE computations easier by scaling all variables similarly. The graphs show the differenced data as more stationary than the undifferenced data, having no trends and with more uniform variations, except for some turbulent periods.

The graphed data were left unnormalized to show outliers. A hallmark of constant-parameter VAR models is their ability to account for noise-perturbed trigonometric cycles. Therefore, the data used for estimation were normalized so that outliers wouldn’t distort an estimated model’s ability to account for the trigonometric cycles, but the data used for forecasting were left unnormalized to make the forecasting more realistic.

5.3. Three estimation-forecasting periods.

In the past 20 years, the U.S. economy went from what has been called the "Great Moderation" to the "Great Recession" (dated 2007:12 - 2009:6 by the National Bureau of Economic Research), including a financial crisis (peaking in 2008:9), to the recent "Slow Recovery" (2009-2016), to a resumption of trend-level growth (2017-2019). The data used in the application cover 1959:1 - 2018:6. To capture three distinct snapshots of a GDP forecaster’s real-time experience of going from the Great Moderation to the present, the model estimation and forecasting in the application was carried out separately for each of three estimation-forecasting periods: an "early" period in which models were estimated for 1959:1-1995:12 and forecasts were made for 1996:1-2007:12; a "middle" period in which models were estimated for 1959:1-2007:12 and forecasts were made for 2008:1-2018:6; and, a "late" period in which models were estimated for 1959:1-2009:12 and forecasts were made for 2010:1-2018:6. The three vertical lines in the graphs in Figures 1-2 are positioned at 1995:12, 2007:12, and
2009:12, at the ends of the estimation periods in the three estimation-forecasting periods. It is common to evaluate a forecasting method "recursively", meaning advancing estimation and forecasting period by period. Although doing this can be expected to lead to more accurate forecasts, here this was not done because it would mask the distinct results obtained for the three estimation-forecasting periods.

5.4. Identification of univariate AR and VAR models with MFD.

Under additional assumptions on parameters, beyond those made below equation (2.1), Zadrozny (2016) proved analytically that high-frequency VARMA models are locally identified (uniquely determined) with high-low frequency data, in particular, monthly VARMA models are locally identified with monthly-quarterly data. Local identification can also be checked numerically by checking the negativity of eigenvalues of the Hessian matrix of second-partial derivatives of the likelihood function evaluated at maximum likelihood estimates. The numerical method showed that all estimated monthly models here were locally identified, including monthly univariate AR(2) models of GDP estimated using only quarterly GDP data.

5.5. GDP as stock or flow.

GDP would be considered a "continuous-time stock" if it were observed as the value of a continuous-time generating process at a moment \( t \) and a "continuous-time flow" if it were observed as an integral of the process. If a unit of continuous time measures one month, then, \( g^{(m)}_t = \int_{\tau=t-1}^{t} g(\tau) d\tau \) = quantity of continuously generated GDP in the month ending at moment \( t \) and \( g^{(q)}_t = \int_{\tau=t-3}^{t} g(\tau) d\tau \) = quantity of continuously generated GDP in the quarter ending at moment \( t \). By the linearity of integration,

\[
g^{(q)}_t = g^{(m)}_t + g^{(m)}_{t-1} + g^{(m)}_{t-2},
\]

for discrete-time months \( t = 1, \ldots, T \).

We might say that quarterly-observed discrete-time GDP is a "discrete-time flow" because it sums three consecutive monthly "discrete-time stocks" according
to equation (5.1). However, following Zadrozny (1990), we instead say that quarterly GDP is a "discrete-time stock" if it enters a discrete-time monthly VARMA model as \( g_t^{(q)} \) and is a "discrete-time flow" if enters the model as \( g_t^{(m)} \). Whereas the discrete-time-stock specification of GDP imposes no restrictions on the parameters of a discrete-time monthly VARMA model, the discrete-time-flow specification of GDP does and is so called because equation (5.1) is part of the mapping from the parameters to the likelihood function.

The imposition of restrictions on a discrete-time monthly VARMA model by equation (5.1) in the discrete-time-flow case is discussed in more detail in Zadrozny (1990) in the context of computing the likelihood function for MFD using a state-space representation of a VARMA model and the Kalman filter. Like Zadrozny (1990), the present application initially tried the discrete-time-flow specification of GDP, but, because it produced inferior results compared with the discrete-time-stock specification of GDP, it was abandoned, so that Table 1 reports only the results of the discrete-time-stock specification of GDP.

Mariano and Murasawa (2003) correspondingly estimated a discrete-time monthly FAVARMA-type single-factor model using quarterly-observed GDP and monthly-observed indicators and used it to compute monthly smoothed estimates of the single factor, considered a "new coincident index of business cycles," but used a discrete-time-flow specification of GDP based on a weighted sum that differs from equation (5.1). Whereas their monthly smoothed-estimated factor corresponds to a smoothed estimate of \( g_t^{(m)} \), like Zadrozny (1990), here we produce monthly forecasts of \( g_t^{(q)} \).

5.6. Application results.

All WCFD results in Table 1 were obtained with WCFD forecast horizon \( h = 12 \) months. Experimentation with neighboring values of \( h \) led to similar results that are not reported in the table. Also, all WCFD results in the table are based on the weighting vector \( W = (1,0,...,0) \), with 4 or 12 zeros, that picks GDP as the single primary variable, \( y_{1t} \), from the vector of variables, \( y_t \).

All results in Table 1 are grouped from top to bottom into the three "early", "middle", and "late" estimation-forecasting periods. Results for the 21 estimated models are reported in 7 rows per period. The rows are named in column 1 by model abbreviations: 1/AR2 refers to univariate AR(2) models of GDP; M/VARN refers to VAR(N) models of \( M = 5 \) variables (GDP and 4 coincident indicators) and
M = 13 variables (GDP, 4 coincident indicators, and 8 leading indicators); M/PCF3/VARN refers to VAR(N) models of GDP and 3 PCFD factors derived from initial VAR models of M = 5 and M = 13 variables; and, M/WCF2/VARN refers to VAR(N) models of GDP and 2 WCFD factors derived from initial models of M = 5 and M = 13 variables. All estimated models are monthly, including univariate AR models of GDP estimated using only quarterly GDP data.

Each row in the table reports estimation and forecasting statistics of an estimated model: col. 2 = \(\bar{R}^2\) ("R bar squared") of GDP; col. 3 = Schwarz's (1978) Bayesian Information Criterion (BIC); cols. 4-10 = normalized root mean-squared errors (NRMSE) of GDP forecasts for 1, 2, 3, 6, 12, 18, and 24 months ahead; col. 11 = average NRMSE of GDP forecasts for 1-24 months ahead; and, col. 12 = model rank in terms of 1-24-month-average NRMSE in each of the three estimation-forecasting periods. The NRMSE were normalized by dividing unnormalized RMSE by standard deviations of GDP in the forecasting periods in order to make the NRMSE comparable across the three estimation-forecasting periods with their different GDP variances.

The 1-24-month-ahead-average NRMSE in column 11 of Table 1 were further averaged for each model type over the three estimation-forecasting periods and are reported at the bottom of the table along with their overall implied ranks. The overall NRMSE at the bottom of the table show that models 5/VAR2, 5/WCF2/VAR2-4, 13/VAR1, and 13/WCF2/VAR2-4 have closely spaced overall NRMSE averaging 1.061 and models 1/AR2, 5/PCF3/VAR1-2, and 13/PCF3/VAR2 have more widely spaced overall NRMSE averaging 1.125, about 6.0% higher. The 1-24-month-ahead-average NRMSE in column 11 were also averaged for each estimation-forecasting period, yielding .978, 1.118, and 1.163 for the early, middle, and late periods, so that the GDP forecasts become increasingly less accurate going from the early to late periods.

Table 1 leads to the following 5 conclusions:

1. Nelson (1972) reported that quarterly few-parameter univariate ARMA models forecast major U.S. macroeconomic series at quarterly intervals more accurately than many-variable and many-parameter econometric models, which overturned many economist's beliefs at the time. Table 1 reports the opposite: that the univariate AR models forecast GDP much less accurately than any multivariate model in all but 1/21 = 4.8% of cases, especially 1-3 months ahead and in the middle and late estimation-forecasting periods. Thus, apparently both factor-free and factor-including VAR models successfully exploit 0-3-month-ahead correlations between quarterly-observed GDP and monthly-observed indicators,
which validates estimating multivariate monthly models with monthly-quarterly data in order to forecast GDP at monthly intervals.

2. However, in all three estimation-forecasting periods in Table 1, NRMSE begin to and continue to converge after about 12 forecast-ahead months, so that more difficult to estimate VAR models with MFD, both initial factor-free and subsequent factor-including models, begin to and then lose their forecasting advantage over simpler to estimate univariate AR models.

3. WCFD-factor models forecast GDP at monthly intervals more accurately than PCFD-factor models in all but $1/21 \approx 4.8\%$ of cases in Table 1. The 1-24-month-ahead-average NRMSE in column 11 of Table 1 of PCFD-model forecasts are 0.9% - 2.2%, 7.1% - 11.3%, and 4.0% - 4.9% higher, respectively, in the early, middle, and late estimation-forecasting periods than those of WCFD-model forecasts.

4. Stock and Watson (2002b) reduced 215 variables to a handful of PCFD factors, which indicates that most of their secondary variables are redundant for forecasting U.S. macroeconomic variables. In the early period in Table 1, models 5/VAR2, 5/PCF3/VAR2, and 5/WCF2/VAR4 forecast GDP more accurately than models 13/VAR1, 13/PCF3/VAR2, and 13/WCF2/VAR4, which indicates that the 4 coincident indicators (but not the 8 leading indicators) have little redundancy for forecasting GDP in this calmer period. By contrast, in the middle and late periods, models 13/VAR1 and 13/WCF2/VAR2 (but not 13/PCF3/VAR2) produce similarly accurate GDP forecasts as models 5/VAR2 and 5/WCF2/VAR2-3, in fact, produce NRMSE more than 1% lower in $9/14 = 64.3\%$ cases, which indicates that the 8 leading indicators contain additional information, beyond that in the 4 coincident indicators, that is useful for forecasting GDP in the latter two, more turbulent, periods.

5. The result in the middle and late periods that 260-estimated-parameter 13-variable initial models forecast GDP about as accurately or more accurately than 24-estimated-parameter 3-variable WCFD models seems to contradict the principle of parsimony. However, because a little-changing data generating process (DGP) underlies parsimony and the DGP of GDP and the 12 indicators surely changed significantly during the 2007-2009 Great Recession and financial crisis (at least, in terms of a constant-parameter VAR model's ability to fit and forecast data), the principle of parsimony might be less applicable in the middle and late periods.
The above discussion of forecast accuracy is "retrospective" in the sense of being based on both estimation-period and forecasting-period data. In practice, we want "prospective" statistics based only on estimation-period data to predict a model's ability to forecast accurately beyond the estimation period. In theory, Kullback-Leibler information (KLI; Bowden, 1973) does this by measuring a model's distance to the DGP of its variables. However, KLI cannot be evaluated, only estimated using an information criterion (IC). Many ICs have been proposed with weaker (Akaike AIC, 1973) to stronger (Schwarz BIC, 1978) penalties on the number of estimated parameters. Table 1 includes BIC in column 3 because in our experience BIC has been a good predictor of a model's beyond-estimation-sample forecast accuracy. An IC is expected to be a good predictor of a model's beyond-estimation-sample forecast accuracy if the DGP has not changed much in passing from the estimation period to the forecasting period.

BIC is strictly comparable only for models with the same variables and estimated using the same data. Therefore, BIC is strictly comparable only for models in the same estimation-forecasting period. For this reason, $\bar{R}^2$ of GDP was included in column 2 of Table 1, so that any model could be compared with any other model in the same or different estimation-forecasting period. $\bar{R}^2$ of GDP is correlated with model ranks in column 12 at 96.4%, 96.4%, and 92.9% in the three estimation-forecasting periods and BIC is correlated with the model ranks at 100%, 92.8%, and 100% in the three periods. Therefore, in all three estimation-forecasting periods of the application, both $\bar{R}^2$ of GDP and BIC are good predictors of a model's relative accuracy for forecasting quarterly GDP at monthly intervals.

6. Conclusion.

Suppose a vector autoregressive moving-average (VARMA) model is estimated for $m$ observed variables of primary interest for an application and $n-m$ observed secondary variables to aid in the application. An application indicates the variables of primary interest but usually only broadly suggests secondary variables that may or may not be useful. Often, one has many potential secondary variables to choose from but is unsure which ones to include or exclude from the application. The paper proposes a method called weighted-covariance factor decomposition (WCFD), comparable to Stock and Watson's (2002a,b) method here
called principle-components factor decomposition (PCFD), for reducing secondary variables to fewer factors in order to more effectively apply an estimated model. The WCFD method is illustrated in the paper by forecasting quarterly-observed U.S. real GDP at monthly intervals using monthly-observed 4 coincident and 8 leading indicators from the Conference Board (2018). The results show that normalized root mean-squared errors of GDP forecasts of PCFD-factor models are 0.9% - 11.3% higher than those of WCFD-factor models, especially as estimation-forecasting periods pass from the pre-2007 Great Moderation through the 2007-2009 Great Recession to the 2009-2016 Slow Recovery.

The WCFD approach requires estimating an initial VARMA model. If the initial model is a VAR model and the data are single-frequency data, then, this can usually be done by applying ordinary least-squares (OLS) estimation separately to each equation of the model, if \( T = \text{number of sample periods} > \text{number of coefficients to be estimated in any equation} \). However, if \( n = \text{number of variables in the model} \geq T \), then, the estimated innovation covariance is singular, which by itself doesn't invalidate WCFD, but the estimated covariance matrix has no statistical reliability. The present application successfully estimated 260 parameters of unrestricted 13-variable VAR(1) models using maximum likelihood (MLE), but didn't attempt to estimate higher-order 13-variable VAR models with 169 or more estimated parameters.

In cases of too many current and lagged variables and disturbances in a VARMA model resulting in too many estimated parameters, one could estimate a "diagonal" model with univariate ARMA coefficients in each equation of the model but a full nondiagonal innovation covariance matrix. Cross-variable dynamics are limited but not eliminated in such diagonal models, because, being contemporaneously correlated, innovations still propagate across different variables. Univariate ARMA coefficients of diagonal models can be estimated by applying Box and Jenkins' (1976) methods separately to each equation of the model, which greatly eases MLE, especially with mixed-frequency data. The innovation covariances could be estimated residually, like residual covariances of ordinary least-squares regressions. Box and Jenkins' methods are known for producing forecasts whose accuracies are often difficult to improve upon. This diagonal approach reduces the number of nonlinearly estimated parameters of initial VARMA models from a quadratic function of the number of variables in the model to a linear function.

**Appendix: Computing the WCFD decomposition.**
The appendix first derives first-order conditions (FOC) of WCFD and then discusses computing the WCFD decomposition using two alternative closely-related methods. Each method solves the FOC, for \( i = 1, \ldots, n-1 \). If \( r_n \) is needed for \( i = n \), then, for better expected accuracy, we recommend computing it separately at the end of each method. The second method is easier to implement, was used in the application, and requires about the same number of computations as PCA.

The difference between the two methods is that method 1 computes the first \( n-1 \) eigenvalues and eigenvectors of \( \Sigma_i Q \) sequentially, following the definition of WCFD, whereas method 2 computes them simultaneously. It can be shown that the two methods produce the same result in theory (but a slightly different result in practice because of slightly different computations), if \( r_i^T \Sigma_i^{-1} r_i = 0 \) and \( r_i^T Q r_i = 0 \), for \( i, j = 1, \ldots, n \) and \( i \neq j \), which hold because \( \Sigma_i \) is positive definite by assumption (i) and because \( \Sigma_i \) and \( Q \) are symmetric.

A.1. Derivation of first-order conditions.

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

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

(A.1)

where \( \Xi_i \) is an \( n \times n \) matrix of Lagrange multipliers. The FOC are obtained by differentiating \( \mathcal{L}_i \) with respect to \( r_i \), \( Y_i \), and \( \Xi_i \) and setting the derivatives to zero,

(A.2) \( (Q - \Xi_i) r_i = 0_{n \times 1} \),

(A.3) \( \Xi_i Y_i = 0_{n \times (n-i)} \),

(A.4) \( r_i^T + Y_i^T = \Sigma_i \),

for \( i = 1, \ldots, n-1 \), where \( 0_{k \times \ell} \) denotes the \( k \times \ell \) zero matrix.
Postmultiply equation (A.4) by $\Xi_i$, use equation (A.3), and obtain

\begin{equation}
(\Sigma_i - r_i^T)\Xi_i = 0_{n \times n},
\end{equation}

for $i = 1, \ldots, n-1$. Then, postmultiply equation (A.5) by $r_i$, use equation (A.2) and the symmetry of $Q$ to replace $r_i^T\Xi_i$ with $r_i^TQ$, replace $r_i^TQr_i$ with $v_i$, and obtain the eigenvalue problem

\begin{equation}
(\Sigma_iQ - v_iI_n)r_i = 0_{n \times 1},
\end{equation}

for $i = 1, \ldots, n-1$. Thus, FOC (A.2)-(A.4) reduce to (A.6).

**A.2. Method 1 for solving the FOC.**

Recall the following linear algebra. A real positive semi-definite square matrix, symmetric or not, has real nonnegative eigenvalues and real eigenvectors. $\Sigma_z$ is symmetric positive semi-definite by definition and positive definite by assumption (i); $Q$ is symmetric positive semi-definite by definition, is often positive definite if forecast horizon $h$ is sufficiently large, and can always be made positive definite by adding a small positive diagonal matrix added to it (see below equation (2.7)). The product of two real symmetric positive semi-definite matrices is generally asymmetric, but has real and nonnegative eigenvalues and real eigenvectors. If $\Sigma_z$ has full rank $n$, then, $\Sigma_i$ has rank $n-i+1$.

For $i = 1, \ldots, n-1$, let $\lambda_{i1} \geq \ldots \geq \lambda_{im} \geq 0$ denote the real nonnegative eigenvalues of $\Sigma_iQ$, ordered in decreasing size and let $z_{i1}, \ldots, z_{im}$ denote their associated real eigenvectors. Then, for $i = 1, \ldots, n-1$, because $v_i = r_i^TQr_i$ is being maximized, set $v_i = \lambda_{i1}$, determine $k =$ number of significant $v_i$, and set

\begin{equation}
r_i = \sqrt{\frac{\lambda_{i1}}{z_{i1}^TQz_{i1}}} z_{i1},
\end{equation}
where associated eigenvector $z_i$ is scaled as $z_i^T z_i = 1$, so that $v_i = r_i^T Q r_i$, as desired. The eigenvalues of $\Sigma Q$ exist and are unique in any case. The assumption that nonzero eigenvalues of $\Sigma z Q$ are distinct, implies that significant $[z_{i1},...,z_{ik}]$ and $R_i = [r_{i1},...,r_{ik}]$ exist, have full column rank, and are unique (Wilkinson, 1965, pp. 4-6).

If the last $n-k$ eigenvalues of $\Sigma z Q$ are close to zero, then, we may be unable to compute insignificant $R_2 = [r_{k+1},...,r_n]$ accurately (Golub and Van Loan, 1996). But, because we need only significant $R_i$ to compute the significant part of the WCFD factorization, we can forego computing insignificant $R_2$. If $r_n$ is needed, we recommend computing it using equations (A.9)-(A.10).

### A.3. Method 2 for solving the FOC.

Let $\mu_1 \geq ... \geq \mu_n \geq 0$ denote the real and nonnegative eigenvalues of $\Sigma z Q$ in decreasing size and let $x_1,\ldots,x_n$ denote the associated real eigenvectors. Then, for $i = 1,\ldots,n$, compute $\mu_i$ and $x_i$, set $V_i = \mu_i$, determine $k = \text{number of significant } v_i$, and, for $i = 1,\ldots,k$, set

$$r_i = \frac{\mu_i}{\sqrt{x_i^T Q x_i}} x_i,$$

where $x_i$ is scaled as $x_i^T x_i = 1$, so that $v_i = r_i^T Q r_i$, as desired.

### A.4. Computing $r_n$.

Postmultiply, $\Sigma_n = r_n r_n^T$ by $r_n$, which leads to the eigenvalue problem

$$\Sigma_n - \omega_i I_n x_i = 0_{n \times 1},$$
for \( i = 1, \ldots, n \), where \( \omega_i \) denotes a real and nonnegative eigenvalue of \( \Sigma_n \) and \( \chi_i \) denotes its associated eigenvector. Because \( \Sigma_n \) has rank one, \( \omega_1 > 0 \) and \( \omega_2 = \ldots = \omega_n = 0 \). Then, set

\[
(A.10) \quad r_n = \sqrt{\omega_1} \chi_1,
\]

and scale \( \chi_1 \) as \( \chi_1^T \chi_1 = 1 \), so that \( \omega_1 = r_n^T r_n \) and \( RR^T = \Sigma_\xi \), as desired.

REFERENCES


Figure 1: U.S. quarterly real GDP, monthly 4 coincident indicators and index of coincident indicators, January 1959 to June 2018

a. Natural logs.

b. Differences of natural logs.
Figure 2: U.S. quarterly real GDP and monthly 8 leading indicators, January 1959 to June 2018

a. Natural logs, except interest rate spread.

b. Differences of natural logs, except unlogged differences of interest rate spread.
Table 1: Normalized Root Mean-Squared Errors of GDP Forecasts

"Early" estimation-forecasting period

<table>
<thead>
<tr>
<th># vars model</th>
<th>RBSQ GDP</th>
<th>BIC</th>
<th>Normalized root mean-squared errors (NRMSE) 1-24 months ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1/AR2</td>
<td>.081</td>
<td>-6734</td>
<td>.960</td>
</tr>
<tr>
<td>5/VAR2</td>
<td>.677</td>
<td>1098</td>
<td>.746</td>
</tr>
<tr>
<td>5/PCF3/VAR2</td>
<td>.507</td>
<td>9498</td>
<td>.803</td>
</tr>
<tr>
<td>5/WCF2/VAR4</td>
<td>.596</td>
<td>-601</td>
<td>.740</td>
</tr>
<tr>
<td>13/VAR1</td>
<td>.546</td>
<td>4425</td>
<td>.901</td>
</tr>
<tr>
<td>13/PCF3/VAR2</td>
<td>.340</td>
<td>11407</td>
<td>.849</td>
</tr>
<tr>
<td>13/WCF2/VAR4</td>
<td>.513</td>
<td>-1566</td>
<td>.858</td>
</tr>
</tbody>
</table>

"Middle" forecasting period

<table>
<thead>
<tr>
<th># vars model</th>
<th>RBSQ GDP</th>
<th>BIC</th>
<th>Normalized root mean-squared errors (NRMSE) 1-24 months ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1/AR2</td>
<td>.098</td>
<td>-8927</td>
<td>1.029</td>
</tr>
<tr>
<td>5/VAR2</td>
<td>.486</td>
<td>1506</td>
<td>.724</td>
</tr>
<tr>
<td>5/PCF3/VAR2</td>
<td>.472</td>
<td>12358</td>
<td>.834</td>
</tr>
<tr>
<td>5/WCF2/VAR3</td>
<td>.451</td>
<td>-1440</td>
<td>.799</td>
</tr>
<tr>
<td>13/VAR1</td>
<td>.503</td>
<td>5792</td>
<td>.699</td>
</tr>
<tr>
<td>13/PCF3/VAR2</td>
<td>.285</td>
<td>14983</td>
<td>.868</td>
</tr>
<tr>
<td>13/WCF2/VAR2</td>
<td>.425</td>
<td>-1555</td>
<td>.699</td>
</tr>
</tbody>
</table>

"Late" estimation-forecasting period

<table>
<thead>
<tr>
<th># vars model</th>
<th>RBSQ GDP</th>
<th>BIC</th>
<th>Normalized root mean-squared errors (NRMSE) 1-24 months ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5/VAR2</td>
<td>.637</td>
<td>1330</td>
<td>1.096</td>
</tr>
<tr>
<td>5/WCF2/VAR2</td>
<td>.419</td>
<td>-1116</td>
<td>1.116</td>
</tr>
<tr>
<td>13/VAR1</td>
<td>.545</td>
<td>5730</td>
<td>1.114</td>
</tr>
<tr>
<td>13/WCF2/VAR2</td>
<td>.441</td>
<td>-1504</td>
<td>1.098</td>
</tr>
</tbody>
</table>

Overall average NRMSE and rank

<table>
<thead>
<tr>
<th># vars model</th>
<th>1 AR2</th>
<th>5 VAR2</th>
<th>5 PCF3</th>
<th>5 WCF2</th>
<th>13 VAR1</th>
<th>13 PCF3</th>
<th>13 WCF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall avg</td>
<td>1.121</td>
<td>1.058</td>
<td>1.109</td>
<td>1.061</td>
<td>1.061</td>
<td>1.128</td>
<td>1.065</td>
</tr>
<tr>
<td>Overall rank</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>