

BLS WORKING PAPERS



U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

Weighted-Covariance Factor Decomposition of VARMA Models Applied to Forecasting Quarterly U.S. Real GDP at Monthly Intervals

Peter A. Zadrozny, U.S. Bureau of Labor Statistics
Baoline Chen, Bureau of Economic Analysis

Working Paper 516
September 2019

SPECIAL ISSUE

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

PETER A. ZADROZNY^{a,b,c}  AND BAOLINE CHEN^d

^a *Division of Price Index Number Research, Bureau of Labor Statistics, Washington, DC, USA*

^b *Center for Financial Studies (CFS), Goethe University, Frankfurt, Germany*

^c *Center for Economic Studies and Ifo Institute for Economic Research (CESifo), Munich, Germany*

^d *Office of the Director, Bureau of Economic Analysis, Washington, DC, USA*

Suppose a vector autoregressive moving-average 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 article proposes a method called weighted-covariance factor decomposition (WCFD), comparable to Stock and Watson's method here called principle-components factor decomposition (PCFD), for reducing the secondary variables to fewer factors to obtain a parsimonious estimated model that is more effective in an application. The WCFD method is illustrated in the article by forecasting quarterly observed U.S. real GDP at monthly intervals using monthly observed four coincident and eight leading indicators from the Conference Board (<http://www.conference-board.org>). 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.

Received 30 June 2018; Revised 15 August 2019; Accepted 15 August 2019

Keywords: principal-components-type decomposition of multivariate time-series models.

JEL. C33; C53.

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 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 of secondary variables need 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: (i) uses eigenvectors of the sample contemporaneous covariance matrix of secondary variables to

* Correspondence to: Peter A. Zadrozny, Bureau of Labor Statistics, 2 Massachusetts Avenue, NE, Room 3105, Washington, DC 20212, USA. E-mail: peterzadrozny@gmail.com

linearly transform them to factors; (ii) based on experience and experimentation (including using information criteria), decides which factors are significant; (iii) estimates an AR equation of primary variables and significant factors; and (iv) 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 article 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 four coincident and eight 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 article 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 articles in various fields of science. For example, Delchambre (2015) develops and illustrates a modified standard PCA based on weighted covariances motivated by the same aim of down-weighting irrelevant or less-relevant observations to minimize the impact of noise.

Corresponding to PCFD, WCFD has the following steps: (i) estimates an initial VARMA model of primary and secondary variables; (ii) uses the initial model to compute factors that account for variations in primary variables; (iii) based on experience and experimentation (including using information criteria), decides which factors are significant; (iv) estimates a smaller VARMA model of primary variables and significant factors; and (v) 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 article 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 (Forni and Reichlin, 1998; Forni *et al.*, 2000; Stock and Watson, 2002a,b).

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

Table I. Normalized root mean-squared errors of GDP forecasts

#vars/model	\bar{R}^2 of GDP	BIC	Normalized root mean-squared errors (NRMSE) – 1–24 months ahead								Rank
			1	2	3	6	12	18	24	Average 1–24	
‘Early’ estimation-forecasting period: January 1959–December 1995 and January 1996–December 2007 (444 and 144 months)											
1/AR2	0.081	−6734	0.960	0.960	0.960	0.961	0.994	1.004	1.006	0.991	7
5/VAR2	0.677	1098	0.746	0.763	0.959	0.994	0.993	1.001	1.002	0.966	2
5/PCF3/VAR2	0.507	9498	0.803	0.799	0.917	0.989	1.000	1.000	1.000	0.975	3
5/WCF2/VAR4	0.596	−601	0.740	0.749	0.883	0.968	0.969	0.999	1.012	0.954	1
13/VAR1	0.546	4425	0.901	0.936	1.024	0.979	0.994	1.000	1.006	0.985	5
13/PCF3/VAR2	0.340	11,407	0.849	0.868	0.971	1.003	1.005	1.005	1.005	0.991	6
13/WCF2/VAR4	0.513	−1566	0.858	0.842	1.009	1.030	0.991	0.997	1.002	0.982	4
‘Middle’ estimation-forecasting period: January 1959–December 2007 and January 2008–June 2018 (588 and 126 months)											
1/AR2	0.098	−8927	1.029	1.029	1.029	1.120	1.209	1.223	1.227	1.179	6
5/VAR2	0.486	1506	0.724	0.864	0.793	0.991	1.115	1.132	1.136	1.053	2
5/PCF3/VAR2	0.472	12,358	0.834	0.925	0.977	1.187	1.227	1.228	1.228	1.177	5
5/WCF2/VAR3	0.451	−1440	0.799	0.872	0.864	1.037	1.150	1.174	1.182	1.099	4
13/VAR1	0.503	5792	0.699	0.752	0.801	0.995	1.119	1.135	1.137	1.045	1
13/PCF3/VAR2	0.285	14,983	0.868	0.998	1.134	1.220	1.228	1.228	1.228	1.197	7
13/WCF2/VAR2	0.425	−1555	0.699	0.765	0.851	1.008	1.141	1.168	1.179	1.075	3
‘Late’ estimation-forecasting period: January 1959–December 2009 and January 2010–June 2018 (612 and 102 months)											
1/AR2	0.117	−9269	1.192	1.192	1.192	1.193	1.193	1.193	1.193	1.193	6
5/VAR2	0.637	1330	1.096	1.097	1.099	1.137	1.155	1.166	1.160	1.154	3
5/PCF3/VAR1	0.255	2568	1.188	1.170	1.174	1.176	1.176	1.176	1.176	1.176	5
5/WCF2/VAR2	0.419	−1116	1.116	1.095	1.068	1.070	1.168	1.162	1.108	1.131	1
13/VAR1	0.545	5730	1.114	1.138	1.144	1.162	1.151	1.172	1.159	1.154	4
13/PCF3/VAR2	0.206	14,674	1.217	1.172	1.222	1.195	1.193	1.193	1.193	1.195	7
13/WCF2/VAR2	0.441	−1504	1.098	1.094	1.067	1.075	1.180	1.167	1.034	1.139	2
Overall average NRMSE and rank											
#vars/model	1/AR2	5/VAR2	5/PCF3	5/WCF2	13/VAR1	13/PCF3	13/WCF2				
Overall average NRMSE	1.121	1.058	1.109	1.061	1.061	1.128	1.065				
Overall rank	6	1	5	2	2	7	4				

and produced poor forecasts. Therefore, all reported results of the application in Table I 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 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 article conducts no asymptotic analysis, T and n are assumed to be finite and fixed.

The article illustrates WCFD by forecasting U.S. quarterly-observed real GDP at monthly intervals using monthly-observed four coincident and eight 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 Zdrozny (1990) and Mittnik and Zdrozny (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 0.01 or 0.001. 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 Zdrozny, 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 (KLI), 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 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 article 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 four coincident and eight leading indicators and evaluates the forecast accuracies. Section 6 concludes by summarizing the article and discussing estimating an initial 'diagonal' VARMA model when too many variables result in too many estimated parameters in a 'non-diagonal' VARMA model. Appendix A 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 $t = 1, \dots, T$, presumed to be generated by a stationary VARMA model:

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

where ξ_t is an $n \times 1$ vector of unobserved innovations, distributed normally, identically, independently, with zero means and positive definite covariance matrix Σ_ξ or $\xi_t \sim NIID(0_{n \times 1}, \Sigma_\xi)$, where $0_{n \times 1}$ 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 $A(L) = I_n - A_1L - \dots - A_pL^p$, $B(L) = I_n + B_1L + \dots + B_qL^q$, 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 Σ_ξ 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 do not 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 $\det[I_n \lambda^p - A_1 \lambda^{p-1} - \dots - A_{p-1} \lambda - A_p] = 0$ 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 $\det[I_n \lambda^q - B_1 \lambda^{q-1} - \dots - B_{q-1} \lambda - B_q] = 0$ implies $|\lambda| < 1$. Zadrozny (2016) stated assumptions, including those above, and proved that under them a VARMA model is locally identified with single- or MFD.

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

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

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

$$\Psi_j = \sum_{\ell=1}^{\min(j,p)} A_\ell \Psi_{j-\ell} + B_j, \tag{2.3}$$

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

In exactly or nearly non-stationary models, especially when the forecasting horizon h is large, the WCFD decomposition could ‘latch’ onto the largest exactly or nearly non-stationary AR root (the latter defined, say, by $0.99 < |\lambda_i| < 1$) 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 exactly or nearly nonstationary AR roots before computing WCFD. Let $\{\lambda_i\}_{i=1}^v$ denote the v exactly or nearly nonstationary AR roots of model (2.1). We would filter Wold representation (2.2) using $\lambda(L) = (1 - \lambda_1 L) \dots (1 - \lambda_v L)$, would obtain $\tilde{y}_t = \tilde{\Psi}(L)\xi_t$, where $\tilde{y}_t = \lambda(L)y_t$ and $\tilde{\Psi}(L) = \lambda(L)\Psi(L)$, and would compute the WCFD decomposition using $\tilde{\Psi}(L)$.

For a given forecast horizon $h \geq 1$, let $\eta_{ht} = y_{t+h} - E_t y_{t+h}$ denote the $n \times 1$ vector of errors from forecasting y_{t+h} in period t . In terms of innovations, forecast errors are

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

and have covariance matrix

$$\Gamma_h = E\eta_{ht}\eta_{ht}^T = \sum_{j=0}^{h-1} \Psi_j \Sigma_\xi \Psi_j^T. \tag{2.5}$$

We define WCFD in terms of weighted covariances of forecast errors of primary variables in y_{1t} . The primary variables are mapped from all variables as $y_{1t} = W y_t$, where W is a specified $m \times n$ weighting matrix and $1 \leq m = \text{rank}(W) \leq n$. Then, $v = E\eta_{ht}^T \Omega \eta_{ht} = \text{tr}[\Omega \Gamma_h]$ 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 $tr[\cdot]$ denotes the trace of a matrix. Using equation (2.5),

$$v = tr \left[\Omega \sum_{j=0}^{h-1} \Psi_j \Sigma_\xi \Psi_j^T \right]. \tag{2.6}$$

The following three examples illustrate different weighting matrices:

2.1. Selecting Portfolio Value as Primary

Let y_{1t} = value of a portfolio, $y_{2t} = (n - 1) \times 1$ vector of portfolio elements, and $W = (0, w_2, \dots, w_n) = 1 \times n$ 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_{1t} is the single primary variable and portfolio elements are secondary variables. Other secondary variables can be added to y_{2t} .

2.2. Selecting J of n Variables as Primary

Let $\{i_j\}_{j=1}^J \subset \{1, \dots, n\}$ index a subset of J of n variables selected as primary variables and let $e_{i_j} = (0, \dots, 0, 1, 0, \dots, 0)^T$ be the $n \times 1$ vector with 1 in position i_j and zeros elsewhere. Stack row-form $e_{i_j}^T$, for $j = 1, \dots, J$, 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_{1t} .

2.3. 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 , (2.6) can be written equivalently as $v = \sum_{i=0}^n v_i$, where

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

is the portion of the weighted h -step-ahead forecast-error variance, v , accounted for by the i th uncorrelated disturbance, ε_{it} , and $Q = \sum_{j=0}^{h-1} \Psi_j^T \Omega \Psi_j$. Q is symmetric positive semi-definite, but may be non-positive 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 δI_n to it, where δ is a small positive number.

We define the WCFD decomposition recursively. Let $\Sigma_1 = \Sigma_\xi$, $X_1 = 0_{n \times 1}$, and, for $i = 2, \dots, n - 1$, $\Sigma_i = \Sigma_\xi - X_i X_i^T$, $X_i = [r_1, \dots, r_{i-1}]$, and $Y_i = [r_{i+1}, \dots, r_n]$, a matrix of slack variables that ensures that $r_i r_i^T + Y_i Y_i^T = \Sigma_i$ holds. For $i = 1, \dots, n - 1$, we want to maximize $v_i = r_i^T Q r_i$, with respect to r_i and Y_i , so that $r_i r_i^T + Y_i Y_i^T = \Sigma_i$ 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 $r_1 r_1^T + Y_1 Y_1^T = \Sigma_1$, by eliminating Y_1 from first-order conditions (FOC; see Appendix A) 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 from the FOC and solving for r_2 and v_2 . Continuing like this, we determine $X_n = [r_1, \dots, r_{n-1}]$. Finally, given X_n and $\Sigma_n = \Sigma_\xi - X_n X_n^T$, we determine r_n such that $r_n r_n^T = \Sigma_n$ and $RR^T = \Sigma_\xi$ hold for $R = [r_1, \dots, r_n]$.

We assume that (i) Σ_ξ 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 non-null by assumption (ii). Then, it follows

from the discussion in Appendix A that the significant part of the WCFD decomposition, associated with the significantly positive eigenvalues of $\Sigma_\xi Q$, 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 Appendix A.

3. SIGNIFICANT WCFD FACTORS

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 non-stochastic test that is used in the application.

First, consider the following variant of Anderson’s (1984, ch. 11, pp. 473–5) 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 v_i of $\Sigma_\xi Q$, for $i = 1, \dots, k$, account for at least $1 - \rho$ of weighted covariances, so that H_k is true if and only if

$$\delta_k = \bar{\rho}^T \bar{v} = -\rho \sum_{i=1}^k v_i + (1 - \rho) \sum_{i=k+1}^n v_i \leq 0, \tag{3.1}$$

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

Consider the following testing sequence. Start with $\ell = 1$ and test $H_\ell = H_1$. If H_1 is accepted (strictly, not rejected), accept $k = 1$ as the number of significant largest eigenvalues of $\Sigma_\xi Q$ and associated factors; otherwise, test $H_\ell = H_2$. 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 ϕ to \bar{v} nonlinearly and differentially as $\bar{v} = \bar{v}(\phi)$, where vector ϕ collects parameters of the initial VARMA model. If ϕ is certain, for example, because it was chosen for a hypothetical model, then, H_k is accepted if and only if $\delta_k \leq 0$. However, generally, ϕ is estimated and uncertain, so that δ_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. 473–5) 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. 122–4), $\sqrt{T}(\hat{\bar{v}} - \bar{v}_0) \sim AN(0, \nabla \bar{v} S_0 \nabla \bar{v}^T)$, where $\nabla \bar{v}$ denotes the Jacobian matrix of first-partial derivatives of $\bar{v}(\phi)$ evaluated at $\hat{\phi}$. Then, because $\bar{\rho}$ is constant,

$$\sqrt{T}(\hat{\delta}_k - \delta_{k_0}) \sim AN(0, \hat{\sigma}_{\delta_k}^2), \tag{3.2}$$

where $\hat{\sigma}_{\delta_k}^2 = \bar{\rho}^T \nabla \hat{\bar{v}} \hat{S} \nabla \bar{\rho}$, \hat{S} estimates S_0 such that $\text{plim}_{T \rightarrow \infty} \hat{S} = S_0$, and $\nabla \bar{v}$ and $\hat{\sigma}_{\delta_k}^2$ can be computed using the matrix-differentiation method in Mittnik and Zadrozny (1993) and Chen and Zadrozny (2003, Appendix A).

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

$$\tau_k = \frac{\hat{\delta}_k}{\hat{\sigma}_{\delta_k}} \leq c_\alpha. \tag{3.3}$$

Test (3.3) could mislead if $\hat{\delta}_k$ is not asymptotically normally distributed. For such cases, Onatski's (2009, 2010) non-normal 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 non-stochastically 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 ρ , 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 ρ 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_{1t} . The generalization to m primary variables is straightforward but notationally more tedious. With one primary variable, weighting matrix W is a $1 \times n$ row vector and significant-factor vector f_{1t} is a $k \times 1$ column vector; with $m > 1$ primary variables, W and f_{1t} are $m \times n$ and $k \times m$ 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 $(n - 1) \times (n - 1)$ matrix M of the eigenvectors is orthogonal. Then, the $n - 1$ PCFD factors in $f_t = (\varphi_{1,t}, \dots, \varphi_{n-1,t})^T$ are defined and computed by

$$f_t = M^T y_{2t}. \tag{4.1}$$

Partition $f_t = (f_{1t}^T, f_{2t}^T)^T$, where $f_{1t} = (\varphi_{1,t}, \dots, \varphi_{k,t})^T$ and $f_{2t} = (\varphi_{k+1,t}, \dots, \varphi_{n-1,t})^T$, respectively, 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 (4.1) has the regression form

$$y_{2t} = M_1 f_{1t} + u_t, \tag{4.2}$$

where, for $M_1 = (n - 1) \times k$ matrix of regression coefficients, $f_{1t} = k \times 1$ 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 $n \times n$ non-singular decomposition matrix R whose inverse maps the model's innovations to uncorrelated disturbances, $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})^T = R^{-1} \xi_t = R^{-1} (\xi_{1t}, \dots, \xi_{nt})^T$, for $t = 1, \dots, T$. 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 $R = [r_1, \dots, r_n]$, the Wold representation can be written as $y_t = \sum_{i=1}^n \varphi_{i,t}$, where $\varphi_{i,t} = \sum_{j=0}^{\infty} \Psi_j r_i \varepsilon_{i,t-j}$, so that primary $y_{1t} = W y_t$ has the k -significant-factor representation

$$y_{1t} = \sum_{i=1}^k \varphi_{1,i,t} + \tilde{u}_t, \tag{4.3}$$

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

$$y_{1t} = \tilde{M}_1 \tilde{f}_{1t} + \tilde{u}_t \tag{4.4}$$

corresponding to (4.2), where $\tilde{M}_1 = (\sigma_1^2, \dots, \sigma_k^2) = 1 \times k$ vector of regression coefficients, $\sigma_i^2 = E(\varphi_{1,i,t})^2$, $\tilde{f}_{1t} = (\tilde{\varphi}_{1,1,t}, \dots, \tilde{\varphi}_{1,k,t})^T = k \times 1$, vector of regressors, $\tilde{\varphi}_{1,i,t} = \varphi_{1,i,t}/\sigma_i^2$, and $\tilde{u}_t =$ regression error. Because \tilde{f}_{1t} and \tilde{u}_t are uncorrelated, because ε_{is} and ε_{jt} 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 four 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 $t = 1, \dots, T$, the filter computes $y_{t|t-1} =$ forecast of y_t made in period $t - 1$, whereupon the innovation in period t is estimated as $\xi_t = \hat{y}_t - y_{t|t-1}$, 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, ξ_t , for $t = 1, \dots, T$, 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, Ψ_j , are computed by iterating on (2.3), starting with $\Psi_0 = I_n$. For the i th WCFD factor, $\varphi_{1,i,t} = \sum_{j=1}^{\infty} W\Psi_j r_i \varepsilon_{i,t-j}$, 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

$$\varphi_{1,i,t}^{(J)} = \sum_{j=1}^J W\Psi_j r_i \varepsilon_{i,t-j}. \tag{4.5}$$

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

The four computational steps appear accurate in the application. First, the Kalman-smoothed GDP estimates replicate observed GDP to within 8–9 decimal digits, that is, have ‘single precision’ accuracy. Second, because the largest absolute AR root of any estimated model is about 0.85, for the $J = 50$ used in the application, elements of the approximation-error matrices, $\varphi_{1,i,t} - \varphi_{1,i,t}^{(50)}$, for $i = 1, \dots, k$, are all about ± 0.0001 , that is, 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 I), 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 three decimal digits, whereas the other WCFD decompositions were accurate to about 13–14 decimal digits, that is, 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 March 1959 to June 2018 obtained from the Bureau of Economic Analysis (BEA) are the latest comprehensive revision produced by the Bureau, released in August 2018. Monthly observed U.S. coincident and leading indicators from January 1959 to June 2018 were obtained from the Conference Board (2018): four coincident indicators, (i) number of employees on non-agricultural payrolls, (ii) personal income less transfer payments, (iii) index of industrial production, and (iv) manufacturing and trade sales; and, eight leading indicators, (v) average weekly hours worked in manufacturing, (vi) average weekly initial claims for unemployment insurance benefits, (vii) manufacturer's new orders of consumer goods and materials, (viii) manufacturer's new orders of non-defence capital goods excluding aircraft, (ix) new permits for building private housing units, (x) Standard and Poor's index of 500 stock prices, (xi) Conference Board's index of consumer expectations, and (xii) 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 (i) naturally logged (except the interest rate spread because it has negative values), (ii) first-differenced at their sampling intervals (GDP quarterly; indicators monthly), (iii) standardized (sample means subtracted then divided by sample standard deviations), and (iv) normalized (outliers more than 3.5 standard deviations from zero treated as missing values). Figures 1 and 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 would not 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 December 2007–June 2009 by the National Bureau of Economic Research), including a financial crisis (peaking in September 2008), to the recent 'Slow Recovery' (2009–2016), to a resumption of trend-level growth (2017–2019). The data used in the application cover January 1959–June 2018. 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 January 1959–December 1995 and forecasts were made for January 1996–December 2007; a 'middle' period in which models were estimated for January 1959–December 2007 and forecasts were made for January 2008–June 2018 and, a 'late' period in which models were estimated for January 1959–December 2009 and forecasts were made for January 2010–June 2018. The three vertical lines in the graphs in Figures 1 and 2 are positioned at December 1995, December 2007, and December 2009, 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

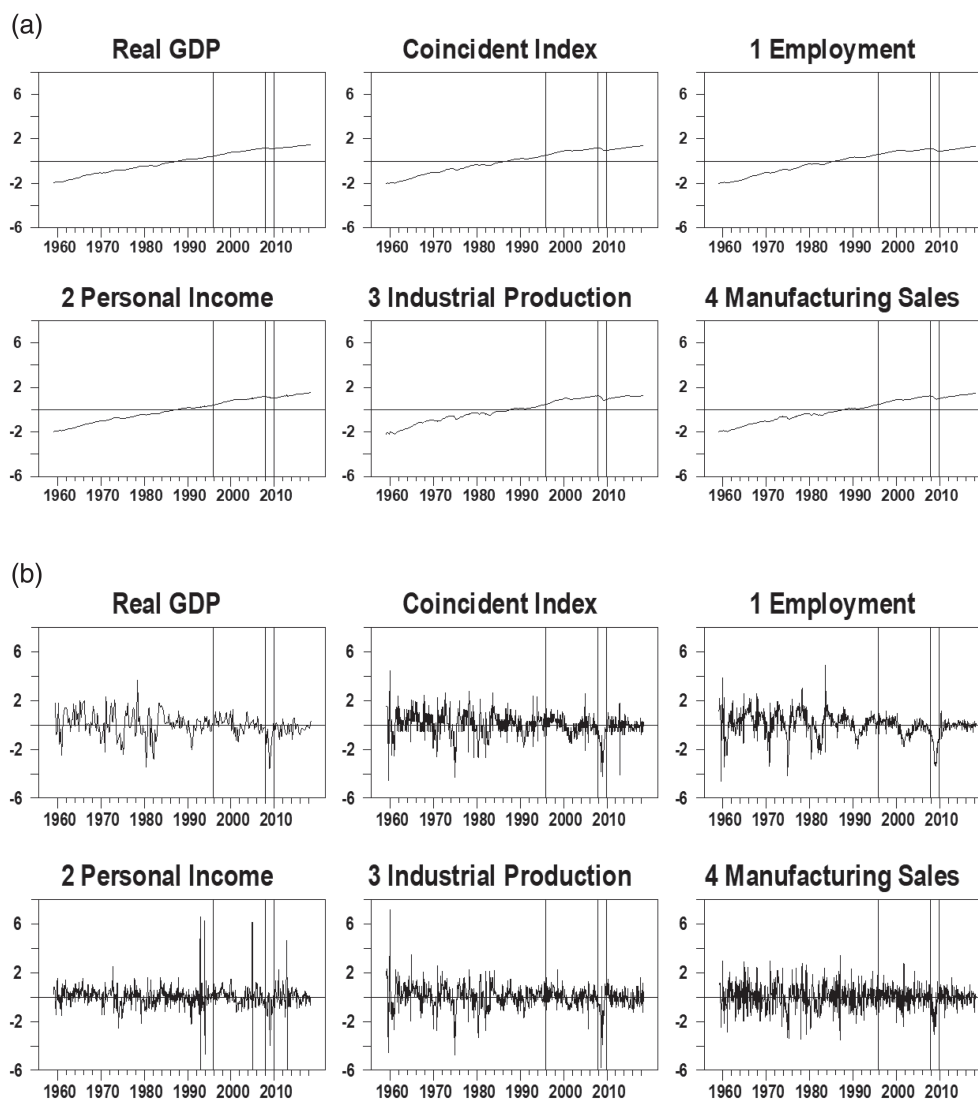


Figure 1. U.S. quarterly real GDP, monthly four coincident indicators and index of coincident indicators, January 1959–June 2018 (a) Natural logs and (b) differences of natural logs

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 (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 MLEs. The numerical method showed that all estimated monthly

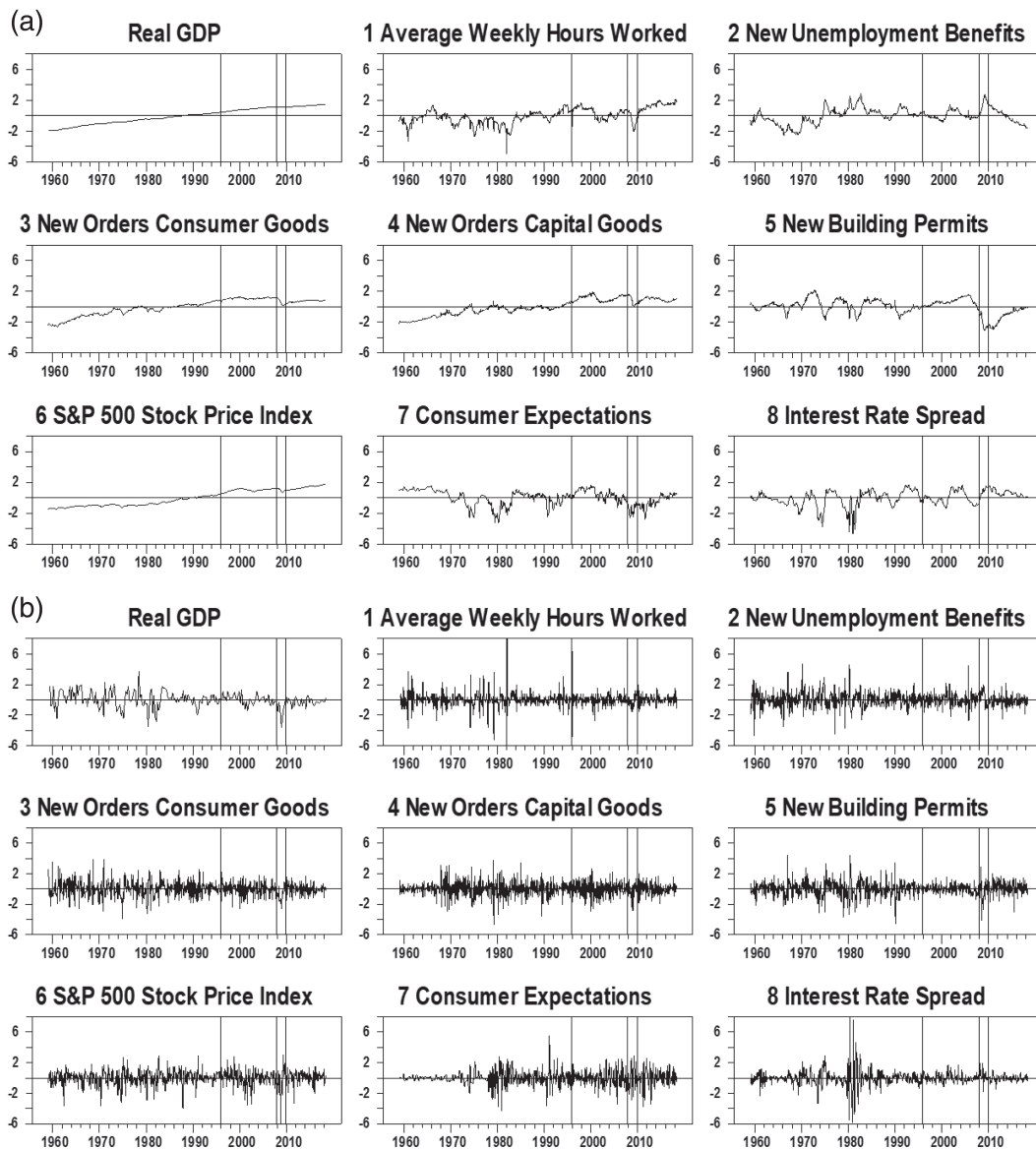


Figure 2. U.S. quarterly real GDP and monthly eight leading indicators, January 1959–June 2018. (a) Natural logs, except interest rate spread and (b) differences of natural logs, except unlogged differences of interest rate spread

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 τ 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, $g_t^{(m)} = \int_{\tau=t-1}^t g(\tau) d\tau =$ quantity of continuously generated GDP in

the month ending at moment t and $g_t^{(q)} = \int_{\tau=t-3}^t g(\tau) d\tau =$ quantity of continuously generated GDP in the quarter ending at moment t . By the linearity of integration:

$$g_t^{(q)} = g_t^{(m)} + g_{t-1}^{(m)} + g_{t-2}^{(m)}, \quad (5.1)$$

for discrete-time months $t = 1, \dots, T$.

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 (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 (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 (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 I 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 I were obtained with WCFD forecast horizon $h = 12$ months. Experimentation with neighbouring 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 $W = (1, 0, \dots, 0)$, 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 I 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(M) 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: column 2 is the \bar{R}^2 (‘ R bar squared’) of GDP; column 3 is the Schwarz’s (1978) Bayesian information criterion (BIC); columns 4–10 are the normalized root mean-squared errors (NRMSE) of GDP forecasts for 1, 2, 3, 6, 12, 18, and 24 months ahead; column 11 is the average NRMSE of GDP forecasts for 1–24 months ahead; and, column 12 is the 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 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 I 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 0.978, 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 I leads to the following five 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 I 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 and computed factors, which validates estimating multivariate monthly models with monthly-quarterly data to forecast GDP at monthly intervals.
2. However, in all three estimation-forecasting periods in Table I, 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 = 4.8% of cases in Table I. The 1–24-month-ahead-average NRMSE in column 11 of Table I of PCFD-model forecasts are 0.9–2.2%, 7.1–11.3%, and 4.0–4.9% higher 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, 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 (AIC; Akaike, 1973) to stronger (BIC; Schwarz, 1978) penalties on the number of estimated parameters. Table I 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 I, 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 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 article proposes a method called WCFD, comparable to Stock and Watson's (2002a,b) method here called PCFD, for reducing secondary variables to fewer factors to more effectively apply an estimated model. The WCFD method is illustrated in the article 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 be done by applying OLS estimation separately to each equation of the model, if $T =$ number of sample periods $>$ number of coefficients to be estimated in any equation. However, if $n =$ number of variables in the model $\geq T$, then, the estimated innovation covariance is singular, which by itself does not 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 MLE, but did not 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 non-diagonal 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 MFD. The innovation covariances could be estimated residually, like residual variances in OLS regressions. Box and Jenkins' methods are known for producing forecasts whose accuracies are often difficult to improve on. 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.

ACKNOWLEDGEMENTS

The article 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 co-editor Marcus Chambers and referees for their helpful comments.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

REFERENCES

- Aguilar O, West M. 2000. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics* **18**:338–357.
- Ahn SK, Reinsel GC. 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 extension of the maximum likelihood principle*. In *Second International Symposium on Information Theory*, Petrov BN, Csaki F (eds.) Akademia Kiado: Budapest, Hungary, pp. 267–281.
- Anderson BDO, Moore JB. 1979. *Optimal Filtering*, Prentice-Hall: Englewood Cliffs, NJ.
- Anderson TW. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley: New York, NY.
- Bai J. 2003. Inferential theory for factor models of large dimensions. *Econometrica* **71**:135–172.
- Bai J, Ng S. 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.
- Bernanke B, Boivin J, Elias P. 2005. Measuring the effects of monetary policy: a factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* **120**:387–422.
- Bowden R. 1973. The theory of parametric identification. *Econometrica* **41**:1069–1074.
- Box GEP, Jenkins GM. 1976. *Time Series Analysis: Forecasting and Control*, revised ed., Holden-Day: San Francisco, CA.
- Box GEP, Tiao G. 1977. A canonical analysis of multiple time series. *Biometrika* **6**:355–365.
- Chen B, Zadrozny PA. 1998. *An extended Yule–Walker method for estimating vector autoregressive models with mixed-frequency data*. In *Advances in Econometrics: Messy Data – Missing Observations, Outliers, and Mixed-Frequency Data*, Fomby TB, Hill RC (eds.), Vol. 13 JAI Press: Greenwich, CT, pp. 47–73.
- Chen B, Zadrozny PA. 2003. Higher moments in perturbation solution of the linear-quadratic exponential Gaussian optimal control problem. *Computational Economics* **21**:45–64.
- Conference Board. 2018. *Coincident and Leading Indicator Data*, Conference Board: New York, NY. <http://www.conference-board.org>.
- Cooley TF, LeRoy SF. 1985. Atheoretical macroeconomics: a critique. *Journal of Monetary Economics* **1**:283–308.
- Deistler M, Hamman E. 2005. Identification of factor models for forecasting returns. *Journal of Financial Econometrics* **3**:256–281.
- Delchambre L. 2015. Weighted principal component analysis: a weighted covariance eigendecomposition approach. *Journal of the Royal Astronomical Society* **446**:3545–3555.
- Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**:1–38.
- Doz C, Giannone D, Reichlin L. 2012. A quasi maximum likelihood approach for large approximate dynamic factor models. *Review of Economics and Statistics* **94**:1000–1013.
- Dufour JM, Stevanovic S. 2013. Factor-augmented VARMA models with macroeconomic applications. *Journal of Business and Economic Statistics* **31**:491–506.
- Forni M, Hallin M, Lippi M, Reichlin L. 2000. The generalized dynamic-factor model: identification and estimation. *Review of Economics and Statistics* **82**:540–554.
- Forni M, Reichlin L. 1998. Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies* **65**:453–473.
- Geweke JF, Singleton KJ. 1981. Maximum likelihood 'confirmatory' factor analysis of economic time series. *International Economic Review* **2**:37–54.
- Golub GH, Van Loan CF. 1996. *Matrix Computations*, 3rd ed., Johns Hopkins University Press: Baltimore, MD.
- Hosoya Y, Taniguchi M. 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 CJ, Nelson CR. 1999. *State space models with regime switching*, MIT Press: Cambridge, MA.
- Mariano RS, Murasawa Y. 2003. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* **18**:427–443.
- Mittnik S, Zadrozny PA. 1993. Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models. *Econometrica* **61**:857–870.
- Mittnik S, Zadrozny PA. 2004. *Forecasting quarterly German GDP at monthly intervals using monthly Ifo Business Conditions Data*. In *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, Sturm JE, Wollmershäuser T (eds.) Physica Verlag: Heidelberg, Germany, pp. 19–48.
- Nelson CR. 1972. The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review* **62**:902–917.
- Onatski A. 2009. Testing hypotheses about the number of factors in large factor models. *Econometrica* **77**:1447–1479.

- Onatski A. 2010. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* **92**:1004–1016.
- Otrok C, Whiteman CH. 1998. Bayesian leading indicators: measuring and predicting economic conditions in Iowa. *International Economic Review* **39**:997–1014.
- Sargent TJ, Sims CA. 1977. *Business cycle modeling without pretending to have too much a priori economic theory*. In *New Methods in Business Cycle Research*, Sims CA (ed.) Federal Reserve Bank of Minneapolis: Minneapolis, MN, pp. 45–109.
- Schwarz G. 1978. Estimating the dimension of a model. *Annals of Statistics* **6**:461–464.
- Serfling R. 1980. *Approximation Theorems of Mathematical Statistics*, John Wiley: New York, NY.
- Sims CA. 1980a. Macroeconomics and reality. *Econometrica* **48**:1–48.
- Sims CA. 1980b. Comparison of interwar and postwar business cycles: monetarism reconsidered. *American Economic Review* **70**:250–257.
- Sims CA. 1986. Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review* **10**(1):2–16.
- Stock JH, Watson MW. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**:1167–1179.
- Stock JH, Watson MW. 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* **20**:147–162.
- Wilkinson JH. 1965. *The Algebraic Eigenvalue Problem*, Clarendon: Oxford, UK.
- Zadrozny PA. 1990. *Estimating a multivariate ARMA model with mixed-frequency data: an application to forecasting U.S. GNP at monthly intervals*. Working Paper No. 90-6, Research Department, Federal Reserve Bank of Atlanta, Atlanta, GA.
- Zadrozny PA. 2016. Extended Yule–Walker identification of VARMA models with single or mixed-frequency data. *Journal of Econometrics* **193**:438–446.

APPENDIX A: COMPUTING THE WCFD DECOMPOSITION

The appendix first derives FOC of WCFD and then discusses computing the WCFD decomposition using two alternative closely related methods. Each method solves the FOC, for $i = 1, \dots, n - 1$. 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 $\Sigma_\xi Q$ 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 $r_i^T \Sigma_\xi^{-1} r_j = 0$ and $r_i^T Q r_j = 0$, for $i, j = 1, \dots, n$ and $i \neq j$, which hold because Σ_ξ is positive definite by assumption (i) and because Σ_ξ and Q are symmetric.

A.1. Derivation of FOC

For $i = 1, \dots, n - 1$, 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

$$\mathfrak{L}_i = r_i^T Q r_i + \text{tr}\{\Xi_i[\Sigma_i - r_i r_i^T - Y_i Y_i^T]\}, \quad (\text{A.1})$$

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

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

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

$$r_i r_i^T + Y_i Y_i^T = \Sigma_i, \quad (\text{A.4})$$

for $i = 1, \dots, n-1$, where $0_{k \times \ell}$ denotes the $k \times \ell$ zero matrix.

Postmultiply (A.4) by Ξ_i , use (A.3), and obtain

$$(\Sigma_i - r_i r_i^T) \Xi_i = 0_{n \times n}, \quad (\text{A.5})$$

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

$$(\Sigma_i Q - v_i I_n) r_i = 0_{n \times 1}, \quad (\text{A.6})$$

for $i = 1, \dots, n-1$. 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 definite square matrix, symmetric or not, has real non-negative eigenvalues and real eigenvectors. Σ_ξ 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 to it see below (2.7). The product of two real symmetric positive semi-definite matrices is generally asymmetric, but has real non-negative eigenvalues and real eigenvectors. If Σ_ξ has full rank n , then, Σ_i has rank $n - i + 1$.

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

$$r_i = \sqrt{\frac{\lambda_{i1}}{z_{i1}^T Q z_{i1}}} z_{i1}, \quad (\text{A.7})$$

where associated eigenvector z_{i1} is scaled as $z_{i1}^T z_{i1} = 1$, so that $v_i = r_i^T Q r_i$, as desired. The eigenvalues of $\Sigma_i Q$ exist and are unique in any case. The assumption that non-zero eigenvalues of $\Sigma_\xi Q$ are distinct, implies that significant $[z_{11}, \dots, z_{k1}]$ and $R_1 = [r_1, \dots, r_k]$ exist, have full column rank, and are unique (Wilkinson, 1965, pp. 4–6).

If the last $n - k$ eigenvalues of $\Sigma_\xi Q$ are close to zero, then, we may be unable to compute insignificant $R_2 = [r_{k+1}, \dots, r_n]$ accurately (Golub and Van Loan, 1996). But, because we need only significant R_1 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 (A.9) and (A.10).

A.3. Method 2 for Solving the FOC

Let $\mu_1 \geq \dots \geq \mu_n \geq 0$ denote the real and non-negative eigenvalues of $\Sigma_\xi Q$ in decreasing size and let x_1, \dots, x_n denote the associated real eigenvectors. Then, for $i = 1, \dots, n-1$, compute μ_i and x_i , set $v_i = \mu_i$, determine $k =$ number of significant v_i , and, for $i = 1, \dots, k$, set

$$r_i = \sqrt{\frac{\mu_i}{x_i^T Q x_i}} x_i, \quad (\text{A.8})$$

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

$$(\Sigma_n - \omega_i I_n) \chi_i = 0_{n \times 1}, \quad (\text{A.9})$$

for $i = 1, \dots, n$, where ω_i denotes a real non-negative eigenvalue of Σ_n and χ_i denotes its associated eigenvector. Because Σ_n has rank one, $\omega_1 > 0$ and $\omega_2 = \dots = \omega_n = 0$. Then, set

$$r_n = \sqrt{\omega_1} \chi_1, \quad (\text{A.10})$$

and scale χ_1 as $\chi_1^T \chi_1 = 1$, so that $\omega_1 = r_n^T r_n$ and $RR^T = \Sigma_\xi$, as desired.