WEIGHTED-COVARIANCE FACTOR DECOMPOSITION OF A VARMA MODEL:

APPLIED TO FORECASTING WITH MIXED-FREQUNCY DATA

Peter A. Zadrozny

Bureau of Labor Statistics
2 Massachusetts Ave., NE
Washington, DC 20212
e-mail: zadrozny.peter@bls.gov

Baoline Chen

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

Workshop on Forecasting, National Bank of Poland,

Warsaw, Poland

November 26-27, 2018

[updated November 27, 2018]

This work represents the authors' views and does not represent any official positions of the Bureau of Economic Analysis or the Bureau of Labor Statistics.

Organization.

1. **Objective.**

2. **Define and compute WCFD.**

3. **Reduce large VARMA data model to small VARMA factor model.**

4. **Apply WCFD to U.S. quarterly GDP and 10 monthly leading indicators.**

5. **Conclusions.**

1. Objective.

Develop weighted-covariance factor decomposition (WCFD) as a method similar to principal components decomposition (PCD) for reducing a larger estimated VARMA "data" model of many variables to a smaller VARMA "factor" model of fewer "primary" variables of interest.

The factor model is based on structural disturbances of the data model that account for a desired percentage of variations of the primary variables.

WCFD is applied to U.S. time-series data on quarterly GDP and 10 monthly leading indicators from the Conference Board:

(i) A monthly VAR(1) data model of quarterly GDP and 10 monthly indicators is estimated using MLE and a missing-data Kalman filter (MDKF) to handle the mixed-frequency data (MFD) (Zadrozny, 1988, 1990a,b);

(ii) WCFD is applied to reduce the estimated VAR(1) data model to a monthly univariate ARMA(2,1) factor model of GDP as the one primary variable;

(iii) The MDKF is applied to the derived UARMA(2,1) factor model and out-of-sample data to forecast GDP at monthly intervals.

(iv) The factor-model-based GDP forecasts are only overall slightly more accurate than those of the antecedent VARMA data model.

PCD has been applied to time-series data on a large number of variables to produce a few significant factors which are, then, used as observed variables to produce forecasting models of primary variables, e.g., statically by Stock-Watson (JASA & JBES, 2002) and dynamically by Forni-Hallin-Lippi-Reichlin (REStat, 2000).

Here, we skip the intermediate step of producing factors and go directly from data to large model to small model. Previously, apparently only Box-Tiao (Biometrika, 1977) did something similar.

2. Define and compute WCFD.

(i) Start with a VARMA(p,p) data model of relatively many variables:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \xi_t + B_1 \xi_{t-1} + \ldots + B_p \xi_{t-p},$$

where y_t = n×1 vector of relatively many variables and ξ_t = n×1 vector of their innovations ~ IID$(0,\Sigma_\xi)$.

(ii) Consider the data model's transfer function or impulse responses:

$$y_t = A(L)^{-1}B(L)\xi_t = \Psi(L)\xi_t = \sum_{i=0}^{\infty} \Psi_i \xi_{t-i},$$

where $\Psi_i = \sum_{j=1}^{\min(i,p)} A_j \Psi_{i-j} + B_i$, $i \geq 1$, and $\Psi_0 = I_n$.

We need Ψ_i only for i = 1, ..., h = prespecified finite forecast horizon and don't need the full transfer function or stationarity.

(iii) Filter nonstationary roots out of $\psi(L)$:

Nonstationary AR characteristic roots, $\lambda_1, \cdots, \lambda_v$, such that $|\lambda_i| \geq 1$, can first be filtered out of $\Psi(L)$ as $\tilde{\Psi}(L) = (1 - \lambda_1 L)\cdots(1 - \lambda_v L)\psi(L)$ before computing WCFD for filtered $\tilde{\Psi}(L)$.

(iv) If n = no. variables is large:

Estimating a full VARMA or VAR data model may be infeasible. In such cases, consider "diagonal" models with A_i and B_i = diagonal, but Σ_ξ = full.

(v) Weights for selecting primary variables:

For $y_t = (y_{1t}^T, y_{2t}^T))^T = n\times 1$, $y_{1t} = r\times 1$ $(r \leq n)$ vector of primary variables of interest, $y_{2t} = (n-r)\times 1$ vector of secondary variables correlated with and informative for forecasting primary variables. Then, $y_{1t} = Wy_t$, where $W = r\times n$ selection matrix.

(vi) Examples of weighting matrices W:

(a) Principal components decomposition: If $W = I_n = n \times n$, all variables are weighted equally, all are stationary, and forecast horizon h is large, then, WCFD corresponds to PCD and all variables are primary (have nonzero weight).

(b) Forecasting a portfolio: If $W = (\omega_1, \ldots, \omega_n) = 1 \times n$, where ω_i = portfolio allocation weight of an asset ($\omega_i > 0$) or a liability ($\omega_i < 0$), where y_{it} = value of asset i, then, WCFD decomposes the forecast-error variance of the portfolio's value and all variables are primary.

(c) Forecasting one primary variable: If $W = (1, 0, \ldots, 0)^T = 1 \times n$, then, only the first variable in y_t is primary and forecast, illustrated below.

(d) Non-PCD equal weighting: $W = (1, \ldots, 1)^T = 1 \times n$ differs from $W = I_n$ by including all covariance of all variables in WCFD, illustrated below.

(vi) Some potential advantages of WCFD compared to PCD as a method for developing a small factor forecasting model.

(a) WCFD gives only primary variables nonzero weight:

- WCFD focuses the decompositon on primary variables but takes account of their correlations with secondary variables.
- PCD weights all variables equally, whether or not they are informative for primary variables.

(b) Flexible forecast horizon:

- WCFD allows any finite forecast horizon, h.
- PCD implicitly sets an infinite forecast horizon, $h = \infty$.

(c) Nonstationarity:

- WCFD works identically for stationary or nonstationary models.
- PCD strictly makes sense only for stationary variables.

(vii) Other basic components of WCFD:

(a) Decomposition matrix: WCFD computes decomposition matrix $R = [r_1, \ldots, r_n] = n \times n$, where $RR^T = \Sigma_\xi$, R is generally full, exists, is unique under mild conditions, and differs from but corresponds to a lower-triangular Cholesky factor of Σ_ξ.

(b) Structural disturbances: $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})^T = R^{-1}\xi_t = n \times 1 \sim IID(0, I_n)$.

(c) h-step-ahead forecast errors: $\eta_{ht} = \sum_{i=0}^{h-1} \Psi_i \xi_{t+h-i} = n \times 1$, with covariance matrix $\Gamma_h = E\eta_{ht}\eta_{ht}^T = \sum_{i=0}^{h-1} \Psi_i \Sigma_\xi \Psi_i^T$.

(d) Sum of weighted variances:

$$v = tr[W\Gamma_h W^T] = tr[\Sigma_\xi Q] = \sum_{i=1}^{n} v_i, \text{ where } Q = \sum_{i=0}^{h-1} \Psi_i{}^T WW^T \Psi_i, \text{ and}$$

$v_i = r_i^T Q r_i$ = part of v accounted for by ith disturbance ε_{it} through ith column r_i of R.

(viii) Define WCFD analogously to PCD:

Compute r_1 that maximizes $v_1 = r_1^T Q r_1$, such that $RR^T = \Sigma_\xi$. Given r_1, compute r_2 that maximizes $v_2 = r_2^T Q r_2$, such that $RR^T = \Sigma_\xi$. Continue like this. Given $r_1, \ldots, r_{n-2}$, compute r_{n-1} that maximizes $v_{n-1} = r_{n-1}^T Q r_{n-1}$, such that $RR^T = \Sigma_\xi$. Given v, v_1, ..., v_{n-1} and r_1, ..., r_{n-1}, compute r_n such that $RR^T = \Sigma_\xi$ and $v_n = v - \sum_{i=1}^{n-1} v_i$.

(ix) Sufficient conditions for existence of a unique WCFD:

(a) Innovation covariance matrix Σ_ξ of the data model is positive definite,

(b) Weighting matrix W is nonzero,

(c) $\Sigma_\xi Q = \sum_{i=0}^{h-1} \Psi_i^T W W^T \Psi_i$ is diagonalizable with respect to its nonzero eigenvalues.

(x) Compute WCFD analogously to PCD:

(a) First n-1 steps: Compute the $i = 1, \ldots, n-1$ largest real and nonnegative eigenvalues λ_i and associated eigenvectors z_i of $\Sigma_\xi Q$, set

$$v_i = \lambda_i \text{ and } r_i = \sqrt{\frac{\lambda_i}{z_i^T Q z_i}}\, z_i, \text{ where } z_i^T z_i = 1.$$

(b) Last nth step: Given v_i and r_i, for $i = 1, \ldots, n-1$, so that $v_n = v - \sum_{i=1}^{n-1} v_i$, compute

λ_n = real and only positive eigenvalue of $\Sigma_n = \Sigma_\xi - \sum_{j=1}^{n-1} r_j r_j^T$ and

z_n = associated eigenvector, so that $z_n^T z_n = 1$, and set $r_n = (\sqrt{\lambda_n}) z_n$.

(xi) Determine m = number of significant factors:

Let m = minimum no. of structural disturbances that account for a specified fraction of the sum of weighted covariances, $v = \sum_{i=1}^{n} v_i$.

For given ρ (e.g., .95), let m = minimal $1 \leq m \leq n$ such that $\sum_{i=1}^{m} v_i / v \geq \rho$.

This definition of m ignores sampling variability of $\sum_{i=1}^{m} v_i / v$ inherited from the data model, which can be accounted for probabilistically by, among other possibilities, T.W. Anderson's (198x) normal approximations or A. Onatski's (20xx, ..., 20yy) more exact results.

The actual factors φ_{jt} are defined below.

3. Reduce a large VARMA data model of many variables to small VARMA factor Model of primary variables.

Objective is computing a "minimal" VARMA model of y_{1t} in steps (i)-(vi):

$$y_{1t} = A_1^* y_{1,t-1} + \ldots + A_q^* y_{1,t-q} + B_0^* \zeta_t + B_1^* \zeta_{t-1} + \ldots + B_q^* \zeta_{t-q},$$

where $\zeta_t \sim IID(0, I_n)$ and "minimal" means q is minimal, $\dim(\zeta_t) = \dim(y_{1t}) = r$, and the MA part is invertible.

(i) Estimate a VARMA data model for all variables in y_t.

(ii) Compute WCFD of the estimated VARMA model.

(iii) First step for reducing a VARMA data model to a VARMA factor model.

Express the vector of primary variables y_{1t} as the sum of m significant factors and a disturbance.

The VARMA data model implies the transfer-function representation

$$y_{1t} = \sum_{i=1}^{\infty} W\Psi_i R\varepsilon_{t-i} = \sum_{i=1}^{\infty} W\Psi_i \sum_{j=1}^{n} r_j\varepsilon_{j,t-i} = \sum_{j=1}^{n} \sum_{i=1}^{\infty} W\Psi_i r_j\varepsilon_{j,t-i} = \sum_{j=1}^{n} \varphi_{jt},$$

so that $y_{1t} = \sum_{j=1}^{n} \varphi_{jt}$, where $\varphi_{jt} = \sum_{i=1}^{\infty} W\Psi_i r_j\varepsilon_{j,t-i}$ = jth "factor".

For given ρ, let m = smallest integer from 1 to n, such that $\sum_{i=1}^{m} v_i / v \geq \rho$.

Then, $y_{1t} = \sum_{j=1}^{m} \varphi_{jt} + u_t$, where $u_t = \sum_{j=m+1}^{n} \varphi_{jt}$ = "disturbance" of insignificant factors.

(iv) Second step for reducing a VARMA data model to a VARMA factor model.

(a) Consider the following state-space representation for y_{1t}:

State equation $x_{t+1} = Fx_t + G\zeta_t$, with

$$F = \begin{bmatrix} A_1^* & I_r & 0_{r\times r} & . & . & . & 0_{r\times r} \\ . & 0_{r\times r} & . & . & . & . & . \\ . & . & . & . & . & . & 0_{r\times r} \\ . & . & . & . & . & . & I_r \\ A_q^* & 0_{r\times r} & . & . & . & . & 0_{r\times r} \end{bmatrix} = rq\times rq, \quad G = \begin{bmatrix} B_1^* + A_1^* \\ B_2^* + A_2^* \\ . \\ . \\ B_q^* + A_q^* \end{bmatrix} = rq\times q,$$

and state vector $x_t = rq\times 1$, and observation equation $y_{1t} = Hx_t + B_0^*\zeta_t$, with $H = [I_r, 0_{r\times r}, \ldots, 0_{r\times r}] = r\times rq$.

The objective is to determine a minimum value of $q = 1, 2, \ldots,$ such that $\sum_{j=1}^{m} \varphi_{jt}$ acceptably accurately approximates y_{1t}.

(b) <u>First, determine minimum s = rank of Hankel matrix $\mathbf{H}$</u>:

$$\text{Hankel matrix } \mathbf{H} = \begin{bmatrix} \Psi_1^* & \cdot\ \cdot & \Psi_{rp}^* \\ \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot & \cdot \\ \Psi_{rp}^* & \cdot\ \cdot & \Psi_{2rp-1}^* \end{bmatrix} = \begin{bmatrix} HFG & \cdot\ \cdot & HFF^{rp-1}G \\ \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot & \cdot \\ HFF^{rp-1}G & \cdot\ \cdot & HFF^{2rp-2}G \end{bmatrix} = \mathbf{OFC}, \text{ where}$$

$\mathbf{O} = [H^T, F^TH^T, \ldots, (F^{rp-1})^TH^T]^T$ = observability matrix and

$\mathbf{C} = [G, FG, \ldots, F^{rp-1}]$ = controllability matrix.

We needn't consider any larger $\mathbf{H}$ with higher powers of F, because the Cayley-Hamilton theorem implies that $\mathbf{H}$, $\mathbf{O}$, and $\mathbf{C}$ achieve maximum rank when $\mathbf{H}$ is defined as above.

Apply singular value decomposition (SVD) to $\mathbf{H}$ to determine s = rank($\mathbf{H}$).

Note that the objective is to determine minimum q so that the VARMA(q,q) factor model reflects the significant dynamics of primary variables y_{1t}; we would like q = s/r, but s/r may be non-integral.

(c) Compute VARMA coefficients of y_{1t} in substeps (i) to (iii):

(i) Compute minimal q of VARMA(q,q) model of y_{1t}:

Although a minimal state-space model of y_{1t} always exits -- but not a minimal VARMA model of y_{1t} -- we focus on computing the "closest to minimal" VARMA model of y_{1t} because VARMA models are more parsimonious and are commonly used in economics and statistics.

(1) For $r = \dim(y_{1t}) = 1$, $q = \text{int}[s/r] = s$, the minimal VARMA model exists, and it can be computed uniquely by solving the Hankel system below.

(2) For $r > 1$ and $q = \text{int}[s/r] < s/r$, the Hankel system below may not be solvable, but, if it is, it won't reflect all significant y_{1t} dynamics.

(3) For $r > 1$ and $q = \text{int}[s/r] + 1 > s/r$, the Hankel system below is always solvable, will reflect all significant y_{1t} dynamics, but will be non-minimal and non-unique.

(ii) Solve linear Hankel system for y_{1t}'s VARMA coefficients:

First, solve $\begin{bmatrix} A_q^* & \cdots & A_1^* \end{bmatrix} \begin{bmatrix} \Psi_1^* & \cdot & \cdot & \Psi_q^* \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Psi_q^* & \cdot & \cdot & \Psi_{2q-1}^* \end{bmatrix} = \begin{bmatrix} \Psi_{q+1}^* & \cdots & \Psi_{2q}^* \end{bmatrix}$

for the VAR coefficients A_i^* and, then, evaluate

$$\begin{bmatrix} B_0^* \\ B_1^* \\ B_2^* \\ \cdot \\ \cdot \\ \cdot \\ B_q^* \end{bmatrix} = \begin{bmatrix} \Psi_0^* \\ \Psi_1^* \\ \Psi_2^* \\ \cdot \\ \cdot \\ \cdot \\ \Psi_q^* \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ \Psi_0^* & 0 & 0 & & & & \cdot \\ \Psi_1^* & \Psi_0^* & 0 & \cdot & & & \cdot \\ \cdot & \cdot & \Psi_0^* & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & 0 \\ \Psi_{q-1}^* & \Psi_{q-2}^* & \Psi_{q-3}^* & \cdot & \cdot & \cdot & \Psi_0^* \end{bmatrix} \begin{bmatrix} A_1^* \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ A_q^* \end{bmatrix}$$ for the VMA coefficients B_i^*.

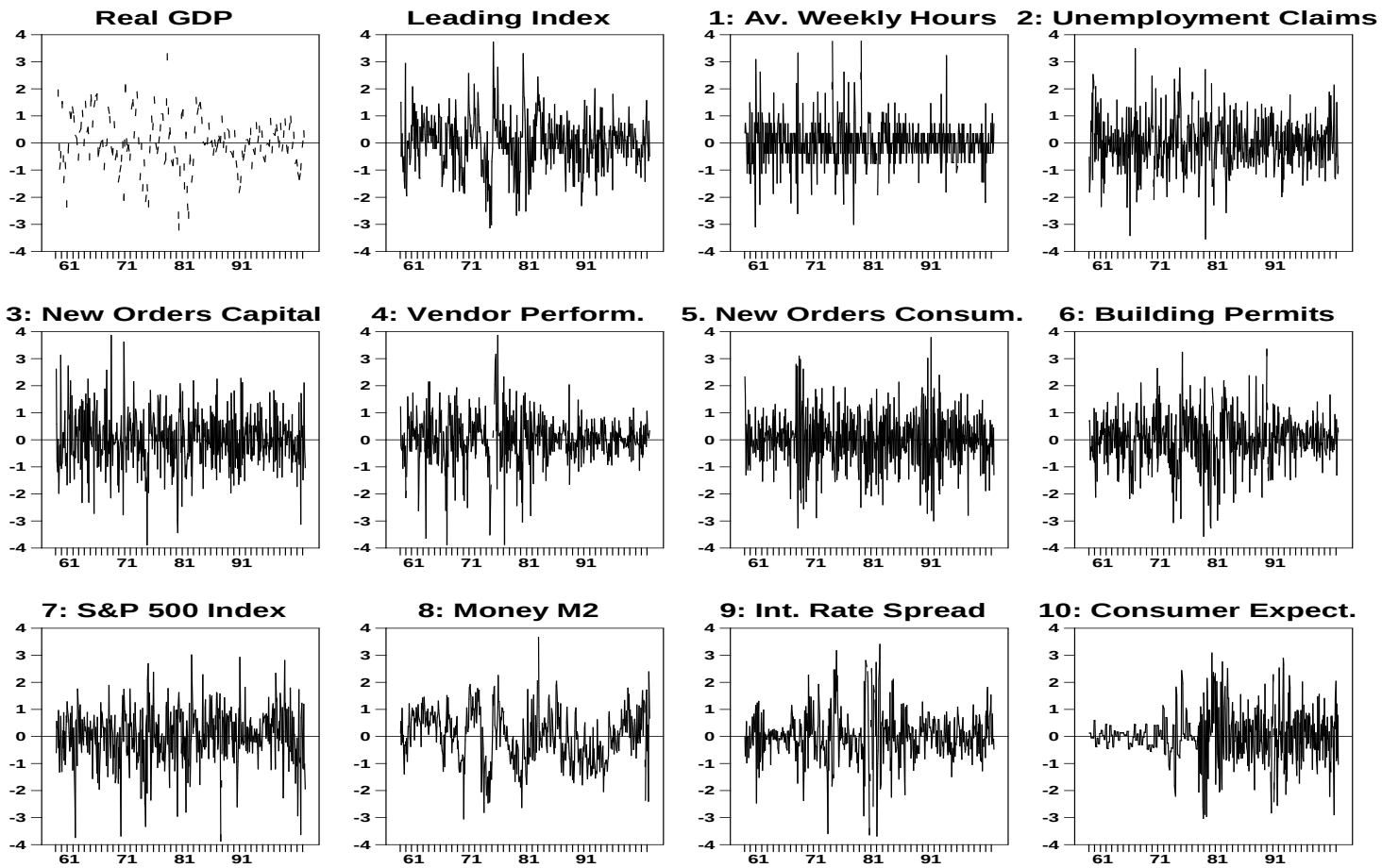
(iii) If necessary, recompute the VMA coefficients to be invertible.

(v) Choose a statistically "best" model of y_{1t}:

One could further vary m and q to find a "best" model in some sense, e.g., that minimizes some information criterion or that produces the lowest RMSEs of out-of-sample forecasts for the particular case being considered.

4. Apply WCFD to U.S. 10 monthly leading indicators and quarterly GDP.

Monthly Δln of leading index and 10 leading indicators and quarterly Δln of GDP from January 1959 to June 2002.



Estimated 11-variable monthly model and GDP-forecast RMSE.

Estimated monthly VAR(1) model of 10 monthly leading indicators and quarterly GDP in Δln form and RMSE of monthly forecasts of GDP.

Var.	LI1	LI2	LI3	LI4	LI5	LI6	LI10	GDP
R_e^2	.132	.106	.172	.135	.202	.159	.101	.443
Q	66.6	27.3	51.1	50.4	83.4	36.9	60.9	29.0
p	.000	.293	.001	.001	.000	.045	.000	.000

Months Ahead	RMSE	SRMSE	Theil U
1	.596	1.03	.744
2	.664	1.14	.829
3	.725	1.25	.905
6	.649	1.12	.834
12	.623	1.07	.720
24	.636	1.09	.680
Average 1-24	.638	1.10	.720

There are 401 in-sample model-estimation months from February 1959 to June 1992, 187 estimated VAR parameters, and 120 out-of-sample GDP-forecast-evaluation months from July 1992 to June 2002.

SRMSE denotes RMSE of the model's GDP forecasts divided by out-of-sample standard deviation of GDP and Theil U denotes RMSE of the model's GDP forecasts divided by RMSE of naive GDP forecasts taken as the last available monthly observations on quarterly GDP.

Percentage WCFD and RMSE of monthly forecasts of GDP of derived monthly UARMA(2,1) model of GDP based on accounting for ρ = 95% of GDP variation.

WCFD for h = 12 months

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

Monthly GDP model implied by WCFD = 2

$$\Delta_3 \ln GDP_t = .993\Delta_3 \ln GDP_{t-1} - .191\Delta_3 \ln GDP_{t-2} + \zeta_t - .279\zeta_{t-1}$$

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

Months Ahead	RMSE	SRMSE	Theil U
1	.632	1.09	.790
2	.632	1.09	.790
3	.632	1.09	.790
6	.619	1.07	.795
12	.636	1.10	.735
24	.636	1.10	.680
Average 1-24	.633	1.09	.714

WCFD = 1 means all variables are weighted equally and WCFD = 2 means only GDP is weighted positively.

No AR roots are filtered out.

5. Conclusions.

(i) Average RMSE of 1-24 month-ahead out-of-sample GDP forecasts of the monthly UARMA(2,1) factor model of GDP are about the same as those of the monthly VAR(1) 11-variable data model.

(ii) Nelson (AER, 1972) showed empirically that small UARMA models can forecast quarterly U.S. macroeconomic variables more accurately than large econometric models.

(iii) The application shows that the information in a large VARMA data model of many variables, informative or not for forecasting a small subset of primary variables, can be passed by WCFD to a small VARMA factor model of the primary variables.

(iv) This ability appears to depend on observability matrix $\mathbf{O}$ having full rank, hence, the VARMA factor model being observable. For single-frequency data, $\mathbf{O}$ always has full rank; for mixed-frequency data, $\mathbf{O}$ may or may not have full rank and, if it does, this depends on the model parameters. See Zadrozny (J. Econometrics, 2016).