9th National Convention on Statistics (NCS)
EDSA Shangri-la Hotel
1 Garden Way, Ortigas Center, Mandaluyong City
October 4-5, 2004

On the Sustainable Development Indices of Agriculture and Fisheries in Southeast Asia

by:
Erniel B. Barrios[1]
Genelyn Ma. F. Sarte

For additional information, please contact:

Author's Name:	Erniel B. Barrios and Genelyn Ma. F. Sarte
Designation:	Associate Professors
Agency:	School of Statistics, University of the Philippines
Address:	Diliman, Quezon City, Philippines
Telefax:	(632) 9280881
E-mail:	**ernielb@yahoo.com** and **gensarte@yahoo.com**

[1]Associate Professor, UP School of Statistics (ernielb@yahoo.com)
[2] Assistant Professor, UP School of Statistics

Figure 1: The CI, the Stock–Watson XCI, and Seasonally-Adjusted Quarterly Real GDP from 1979 to 1983 (1980:Q1=1). The vertical lines are the NBER business cycle reference dates.
Sources: The NBER and the home page of James Stock.

models, i.e., VAR and factor models, following the convention in the literature on index construction.

One can extend this paper in several ways. First, construction of leading and lagging indices is straightforward if we define a leading index as, say, a six-month ahead forecast of monthly real GDP and a lagging index as the final estimate of the past monthly real GDP. Second, other models are worth considering, since they may predict better with less parameters; e.g., VARMA models, dynamic factor models, Markov-switching models, etc. Third, to exploit information in various indicators, one can consider a VAR model for latent monthly real GDP and major principal components, or "diffusion indices," extracted from many indicators; see Stock and Watson (2002*a*, 2002*b*). Fourth, it seems useful to predict monthly real GDP from the production, expenditure, and distribution sides respectively, and combine the three forecasts. Fifth, our framework has other interesting applications; e.g., one can estimate monthly impulse response functions using mixed-frequency series.

The plan of the paper is as follows. Section 2 introduces what we call a mixed-frequency VAR model, i.e., we set up a VAR model for partially latent time series, and derive a state-space model for the observable mixed-frequency series. We then derive the EM algorithm for ML estimation of a Gaussian mixed-frequency VAR model. Section 3 discusses a mixed-frequency factor model. Section 4 applies the method to US data to obtain a new coincident index, and compares it with other coincident indices. Section 5 discusses remaining issues.

2 Mixed-Frequency VAR Model

2.1 VAR Model

Let $\{Y_{t,1}\}$ be an N_1-variate random sequence observable every third period (quarterly series) and $\{Y_{t,2}\}$ be an N_2-variate random sequence observable every period (monthly series). Let for all t, $Y_t := (Y'_{t,1}, Y'_{t,2})'$ and $N := N_1 + N_2$. Assume that $\{\ln Y_t\}$ is integrated of order 1.

Let $\{Y^*_{t,1}\}$ be a latent random sequence underlying $\{Y_{t,1}\}$ such that for all t,

$$\ln Y_{t,1} = \frac{1}{3}\left(\ln Y^*_{t,1} + \ln Y^*_{t-1,1} + \ln Y^*_{t-2,1}\right), \tag{1}$$

i.e., $Y_{t,1}$ is the geometric mean of $Y^*_{t,1}$, $Y^*_{t-1,1}$, and $Y^*_{t-2,1}$. Taking the three-period differences, for all t,

$$\begin{aligned}\ln Y_{t,1} - \ln Y_{t-3,1} &= \frac{1}{3}\left(\ln Y^*_{t,1} - \ln Y^*_{t-3,1}\right) + \frac{1}{3}\left(\ln Y^*_{t-1,1} - \ln Y^*_{t-4,1}\right) \\ &\quad + \frac{1}{3}\left(\ln Y^*_{t-2,1} - \ln Y^*_{t-5,1}\right),\end{aligned}$$

or

$$\begin{aligned}y_{t,1} &= \frac{1}{3}\left(y^*_{t,1} + y^*_{t-1,1} + y^*_{t-2,1}\right) + \frac{1}{3}\left(y^*_{t-1,1} + y^*_{t-2,1} + y^*_{t-3,1}\right) \\ &\quad + \frac{1}{3}\left(y^*_{t-2,1} + y^*_{t-3,1} + y^*_{t-4,1}\right) \\ &= \frac{1}{3}y^*_{t,1} + \frac{2}{3}y^*_{t-1,1} + y^*_{t-2,1} + \frac{2}{3}y^*_{t-3,1} + \frac{1}{3}y^*_{t-4,1},\end{aligned}$$

where $y_{t,1} := \Delta_3 \ln Y_{t,1}$ and $y^*_{t,1} := \Delta \ln Y^*_{t,1}$. We observe $\{y_{t,1}\}$ every third period, and never observe $\{y^*_{t,1}\}$.

Let for all t,

$$y_t := \begin{pmatrix} y_{t,1} \\ y_{t,2} \end{pmatrix}, \quad y^*_t := \begin{pmatrix} y^*_{t,1} \\ y_{t,2} \end{pmatrix},$$

where $y_{t,2} := \Delta \ln Y_{t,2}$. Let

$$\begin{aligned} H(L) &:= \begin{bmatrix} (1/3)I_{N_1} & 0 \\ 0 & I_{N_2} \end{bmatrix} + \begin{bmatrix} (2/3)I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L + \begin{bmatrix} I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L^2 \\ &\quad + \begin{bmatrix} (2/3)I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L^3 + \begin{bmatrix} (1/3)I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L^4, \end{aligned}$$

where L is the lag operator. Let $\mu := \mathrm{E}(y_t)$ and $\mu^* := \mathrm{E}(y_t^*)$. Then for all t,

$$y_t - \mu = H(L)(y_t^* - \mu^*). \tag{2}$$

Assume a Gaussian VAR(p) model for $\{y_t^*\}$ such that for all t,

$$\begin{aligned} \Phi(L)(y_t^* - \mu^*) &= w_t, && (3) \\ \{w_t\} &\sim \mathrm{NID}(0, \Sigma). && (4) \end{aligned}$$

2.2 State-Space Representation

If $p \leq 5$, then we define the state vector as for all t,

$$s_t := \begin{pmatrix} y_t^* - \mu^* \\ \vdots \\ y_{t-4}^* - \mu^* \end{pmatrix}.$$

A state-space representation of the VAR model is for all t,

$$\begin{aligned} s_{t+1} &= A s_t + B z_t, && (5) \\ y_t &= \mu + C s_t, && (6) \\ \{z_t\} &\sim \mathrm{NID}(0, I_N), && (7) \end{aligned}$$

where

$$\begin{aligned} A &:= \begin{bmatrix} \Phi_1 & \dots & \Phi_p & O_{N\times(5-p)N} \\ & I_{4N} & & O_{4N\times N} \end{bmatrix}, \\ B &:= \begin{bmatrix} \Sigma^{1/2} \\ O_{4N\times N} \end{bmatrix}, \\ C &:= \begin{bmatrix} H_0 & \dots & H_4 \end{bmatrix}. \end{aligned}$$

If $p \geq 5$, then we define the state vector as for all t,

$$s_t := \begin{pmatrix} y_t^* - \mu^* \\ \vdots \\ y_{t-p+1}^* - \mu^* \end{pmatrix}.$$

A state-space representation of the VAR model is the same except that

$$\begin{aligned} A &:= \begin{bmatrix} \Phi_1 & \dots & \Phi_{p-1} & \Phi_p \\ & I_{(p-1)N} & & O_{(p-1)N\times N} \end{bmatrix}, \\ B &:= \begin{bmatrix} \Sigma^{1/2} \\ O_{(p-1)N\times N} \end{bmatrix}, \\ C &:= \begin{bmatrix} H_0 & \dots & H_4 & O_{N\times(p-5)N} \end{bmatrix}. \end{aligned}$$

2.3 ML Estimation by a Quasi-Newton Method

Using SsfPack 2.2 by Koopman, Shephard, and Doornik (1999), which runs on Ox 3.3 by Doornik (2001), ML estimation of a linear Gaussian state-space model by a quasi-Newton method is straightforward even with missing observations. When the number of the unknown parameters is large, however, this may fail, especially with a poor initial guess.

2.4 ML Estimation by the EM Algorithm

2.4.1 Missing Observations

Following Mariano and Murasawa (2003), we fill in missing observations with random numbers independent of the model parameters, and rewrite the measurement equation accordingly. Since the Kalman filter "skips" the random numbers, we can simply put 0s for missing observations in practice.

Let for all t,

$$y_{t,1}^{+} := \begin{cases} y_{t,1} & \text{if } y_{t,1} \text{ is observable} \\ v_t & \text{otherwise} \end{cases},$$

where v_t is a random number. We assume for convenience that $\{v_t\} \sim \mathrm{NID}(0, I_{N_1})$. The measurement equation for $\{y_t\}$ is for all t,

$$\begin{pmatrix} y_{t,1} \\ y_{t,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} s_t.$$

We can write for all t,

$$\begin{pmatrix} y_{t,1}^{+} \\ y_{t,2} \end{pmatrix} = \begin{pmatrix} \mu_{t,1} \\ \mu_2 \end{pmatrix} + \begin{bmatrix} C_{t,1} \\ C_2 \end{bmatrix} s_t + \begin{pmatrix} D_{t,1} \\ 0 \end{pmatrix} v_t,$$

where

$$\begin{aligned} \mu_{t,1} &= \begin{cases} \mu_1 & \text{if } y_{t,1} \text{ is observable} \\ 0 & \text{otherwise} \end{cases}, \\ C_{t,1} &= \begin{cases} C_1 & \text{if } y_{t,1} \text{ is observable} \\ 0 & \text{otherwise} \end{cases}, \\ D_{t,1} &= \begin{cases} 0 & \text{if } y_{t,1} \text{ is observable} \\ I_{N_1} & \text{otherwise} \end{cases}. \end{aligned}$$

Let for all t,

$$y_t^{+} := \begin{pmatrix} y_{t,1}^{+} \\ y_{t,2} \end{pmatrix}, \quad \mu_t := \begin{pmatrix} \mu_{t,1} \\ \mu_2 \end{pmatrix}, \quad C_t := \begin{bmatrix} C_{t,1} \\ C_2 \end{bmatrix}, \quad D_t := \begin{bmatrix} D_{t,1} \\ 0 \end{bmatrix}.$$

Then we have a state-space model for $\left\{y_t^{+}\right\}$ such that for all t,

$$s_{t+1} = As_t + Bz_t, \tag{8}$$

$$y_t^{+} = \mu_t + C_t s_t + D_t v_t, \tag{9}$$

$$\left\{\begin{pmatrix} z_t \\ v_t \end{pmatrix}\right\} \sim \mathrm{NID}(0, I_{N+N_1}). \tag{10}$$

2.4.2 Likelihood Function

Assume for simplicity that we know μ^* to be 0. Let $\Phi := [\Phi_1, \ldots, \Phi_p]$, $\phi := \mathrm{vec}(\Phi')$, and $\theta := (\phi', \mathrm{vech}(\Sigma)')'$. Consider an approximate ML estimator of θ, taking s_0 as given. We derive the EM algorithm for solving this ML problem.

Let for $t \geq 0$, $S_t := (s_0, \ldots, s_t)$. Let $Y_0^+ := \emptyset$ and for $t \geq 1$, $Y_t^+ := \left(y_1^+, \ldots, y_t^+\right)$. Let $\Omega \subset \{1, \ldots, T\}$ be the set of periods such that $y_{t,1}$ is missing. By the prediction error decomposition, we can write the joint pdf of $\left(Y_T^+, S_T\right)$ as

$$\begin{aligned}
f\left(Y_T^+, S_T; \theta\right) &= \prod_{t=1}^{T} f\left(y_{t,1}^+, y_{t,2} | s_t, Y_{t-1}^+, S_{t-1}; \theta\right) f\left(s_t | Y_{t-1}^+, S_{t-1}; \theta\right) \\
&= \prod_{t=1}^{T} f\left(y_{t,1}^+ | s_t; \theta\right) f\left(s_t | s_{t-1}; \theta\right) \\
&= \prod_{t \in \Omega} f(v_t) \prod_{t=1}^{T} f(y_t^* | s_{t-1}; \theta).
\end{aligned}$$

Let

$$\begin{aligned}
F &:= \begin{cases} \begin{bmatrix} I_N & O_{N \times 4N} \end{bmatrix} & \text{if } p \leq 5 \\ \begin{bmatrix} I_N & O_{N \times (p-1)N} \end{bmatrix} & \text{if } p \geq 5 \end{cases}, \\
G &:= \begin{cases} \begin{bmatrix} I_{pN} & O_{pN \times (5-p)N} \end{bmatrix} & \text{if } p \leq 5 \\ I_{pN} & \text{if } p \geq 5 \end{cases}.
\end{aligned}$$

Then for all t,

$$\begin{aligned}
F s_t &= y_t^* \\
&= \Phi_1 y_{t-1}^* + \cdots + \Phi_p y_{t-p}^* + w_t \\
&= \Phi G s_{t-1} + w_t \\
&= \begin{bmatrix} \phi_1' \\ \vdots \\ \phi_N' \end{bmatrix} G s_{t-1} + w_t \\
&= \begin{bmatrix} s_{t-1}' G' \phi_1 \\ \vdots \\ s_{t-1}' G' \phi_N \end{bmatrix} + w_t \\
&= \begin{bmatrix} s_{t-1}' G' & & 0 \\ & \ddots & \\ 0 & & s_{t-1}' G' \end{bmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_N \end{pmatrix} + w_t \\
&= \left(I_N \otimes s_{t-1}' G'\right) \phi + w_t.
\end{aligned}$$

The log-likelihood function of θ given $\left(Y_T^+, S_T\right)$ is

$$\begin{aligned}
&\ln L\left(\theta; Y_T^+, S_T\right) \\
&= \sum_{t \in \Omega} \ln f(v_t) - \frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln \det(\Sigma) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} (F s_t - \Phi G s_{t-1})' \Sigma^{-1} (F s_t - \Phi G s_{t-1}) \\
&= \sum_{t \in \Omega} \ln f(v_t) - \frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln \det(\Sigma) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \left[F s_t - \left(I_N \otimes s_{t-1}' G'\right) \phi\right]' \Sigma^{-1} \left[F s_t - \left(I_N \otimes s_{t-1}' G'\right) \phi\right].
\end{aligned}$$

2.4.3 Kalman Filtering and Smoothing

Initial State Let for all t, for $s \geq 0$,

$$\begin{aligned} s_{t|s} &:= \mathrm{E}\left(s_t|Y_s^+\right), \\ P_{t|s} &:= \mathrm{V}\left(s_t|Y_s^+\right). \end{aligned}$$

To start the Kalman filter, one must specify $s_{1|0}$ and $P_{1|0}$. Given stationarity, we have

$$s_{1|0} = 0, \tag{11}$$

$$\operatorname{vec}(P_{1|0}) = (I_{M^2} - A \otimes A)^{-1} \operatorname{vec}(BB'), \tag{12}$$

where M is the dimension of the state vector; see Hamilton (1994, p. 378). The second equation involves inversion of a large matrix, which can cause a computational problem. Hence we instead assume that $s_0 := 0$, which implies that

$$s_{1|0} = 0, \tag{13}$$

$$P_{1|0} = BB'. \tag{14}$$

The resulting estimator is asymptotically equivalent to the exact ML estimator.

Updating We have for $t \geq 1$,

$$\begin{pmatrix} s_t \\ y_t^+ \end{pmatrix} |Y_{t-1}^+ \sim \mathrm{N}\left(\begin{pmatrix} s_{t|t-1} \\ C_t s_{t|t-1} \end{pmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}C_t' \\ C_t P_{t|t-1} & C_t P_{t|t-1} C_t' + D_t D_t' \end{bmatrix}\right).$$

The updating equations are for $t \geq 1$,

$$\begin{aligned} s_{t|t} &= s_{t|t-1} + P_{t|t-1}C_t'\left(C_t P_{t|t-1} C_t' + D_t D_t'\right)^{-1}\left(y_t - C_t s_{t|t-1}\right) \\ &= s_{t|t-1} + K_t e_t, \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}C_t'\left(C_t P_{t|t-1} C_t' + D_t D_t'\right)^{-1} C_t P_{t|t-1} \\ &= (I_M - K_t C_t) P_{t|t-1}, \end{aligned}$$

where

$$\begin{aligned} K_t &:= P_{t|t-1}C_t'\left(C_t P_{t|t-1} C_t' + D_t D_t'\right)^{-1}, \\ e_t &:= y_t^+ - C_t s_{t|t-1}. \end{aligned}$$

Prediction The prediction equations are for $t \geq 1$,

$$\begin{aligned} s_{t+1|t} &= A s_{t|t}, \\ P_{t+1|t} &= A P_{t|t} A' + BB'. \end{aligned}$$

Fixed-Interval Smoothing The following algorithm by de Jong (1989) avoids inversion of large matrices, and hence is more efficient than the standard smoothing equations; see also Durbin and Koopman (2001, sec. 4.3). Let $r_{T+1} := 0$, $R_{T+1} := 0$, and for $t = T, \ldots, 1$,

$$r_t = C_t'\left(C_t P_{t|t-1} C_t' + D_t D_t'\right)^{-1} e_t + L_t' r_{t+1}, \tag{15}$$

$$R_t = C_t'\left(C_t P_{t|t-1} C_t' + D_t D_t'\right)^{-1} C_t + L_t' R_{t+1} L_t, \tag{16}$$

where

$$L_t := A(I_M - K_t C_t).$$

The smoothing equations are for $t = 1, \ldots, T$,

$$\begin{aligned} s_{t|T} &= s_{t|t-1} + P_{t|t-1} r_t, & (17) \\ P_{t|T} &= P_{t|t-1} - P_{t|t-1} R_t P_{t|t-1}. & (18) \end{aligned}$$

The EM algorithm for ML estimation of a state-space model also requires the smoothed first-order autocovariance matrix of the state vector. Let for all t, s, for $r \geq 0$,

$$P_{t,s|r} := \mathrm{Cov}\left(s_t, s_s | Y_r^+\right).$$

De Jong and MacKinnon (1988) show that for $s = 1, \ldots, T-1$, for $t = 1, \ldots, T-s$,

$$P_{t+s,t|T} = \left(I_M - P_{t+s|t+s-1} R_{t+s}\right) L_{t+s-1} \cdots L_t P_{t|t-1}. \quad (19)$$

In particular, for $t = 1, \ldots, T-1$,

$$P_{t+1,t|T} = \left(I_M - P_{t+1|t} R_{t+1}\right) L_t P_{t|t-1}. \quad (20)$$

2.4.4 EM Algorithm

The score function of θ given $\left(Y_T^+, S_T\right)$ consists of

$$\begin{aligned} &\frac{\partial \ln L\left(\theta; Y_T^+, S_T\right)}{\partial \phi} \\ &= \sum_{t=1}^{T} \left(I_N \otimes s_{t-1}' G'\right)' \Sigma^{-1} \left[F s_t - \left(I_N \otimes s_{t-1}' G'\right) \phi\right] \\ &= \sum_{t=1}^{T} \left(I_N \otimes G s_{t-1}\right) \Sigma^{-1} F s_t - \sum_{t=1}^{T} \left(I_N \otimes G s_{t-1}\right) \Sigma^{-1} \left(I_N \otimes s_{t-1}' G'\right) \phi \\ &= \sum_{t=1}^{T} \Sigma^{-1} (F s_t \otimes G s_{t-1}) - \sum_{t=1}^{T} \left(\Sigma^{-1} \otimes G s_{t-1} s_{t-1}' G'\right) \phi \\ &= \sum_{t=1}^{T} \Sigma^{-1} \mathrm{vec}(G s_{t-1} s_t' F') - \sum_{t=1}^{T} \left(\Sigma^{-1} \otimes G s_{t-1} s_{t-1}' G'\right) \phi, \\ &\frac{\partial \ln L\left(\theta; Y_T^+, S_T\right)}{\partial \Sigma^{-1}} \\ &= \frac{T}{2} \Sigma - \frac{1}{2} \sum_{t=1}^{T} (F s_t - \Phi G s_{t-1})(F s_t - \Phi G s_{t-1})' \\ &= \frac{T}{2} \Sigma - \frac{1}{2} \sum_{t=1}^{T} \left(F s_t s_t' F' - F s_t s_{t-1}' G' \Phi' - \Phi G s_{t-1} s_t' F' + \Phi G s_{t-1} s_{t-1}' G' \Phi'\right). \end{aligned}$$

Let for all t, for $s \geq 0$,

$$\begin{aligned} M_{t|s} &:= \mathrm{E}\left(s_t s_t' | Y_s^+\right) \\ &= P_{t|s} + s_{t|s} s_{t|s}'. \end{aligned}$$

Let for all t, s, for $r \geq 0$,

$$\begin{aligned} M_{t,s|r} &:= \mathrm{E}\left(s_t s_s' | Y_r^+\right) \\ &= P_{t,s|r} + s_{t|r} s_{s|r}'. \end{aligned}$$

Let

$$
\begin{aligned}
\bar{M} &:= \frac{1}{T}\sum_{t=1}^{T} M_{t|T}, \\
\bar{M}_1 &:= \frac{1}{T}\sum_{t=1}^{T} M_{t,t-1|T}, \\
L\bar{M} &:= \frac{1}{T}\sum_{t=1}^{T} M_{t-1|T}.
\end{aligned}
$$

Taking the conditional expectation of the likelihood equation given Y_T^+,

$$
\begin{aligned}
\Sigma^{*-1}\mathrm{vec}\left(G\bar{M}_1'F'\right) - \left(\Sigma^{*-1}\otimes GL\bar{M}G'\right)\phi^* &= 0, \\
\Sigma^* - \left(F\bar{M}F' - F\bar{M}_1G'\Phi^{*\prime} - \Phi^*G\bar{M}_1'F' + \Phi^*GL\bar{M}G'\Phi^{*\prime}\right) &= 0,
\end{aligned}
$$

or

$$
\begin{aligned}
\phi^* &= \left(\Sigma^{*-1}\otimes GL\bar{M}G'\right)^{-1}\Sigma^{*-1}\mathrm{vec}\left(G\bar{M}_1'F'\right) \\
&= \left[I_N\otimes\left(GL\bar{M}G'\right)^{-1}\right]\mathrm{vec}\left(G\bar{M}_1'F'\right), \qquad (21) \\
\Sigma^* &= F\bar{M}F' - F\bar{M}_1G'\Phi^{*\prime} - \Phi^*G\bar{M}_1'F' + \Phi^*GL\bar{M}G'\Phi^{*\prime}. \qquad (22)
\end{aligned}
$$

The EM algorithm proceeds as follows:

1. Pick an initial guess $\theta^{(0)}$.
2. (E step) Compute $\left\{s_{t|T}\right\}$, $\left\{P_{t|T}\right\}$, and $\left\{P_{t,t-1|T}\right\}$.
3. (M step) Compute (ϕ,Σ), and use it as $\theta^{(1)}$.
4. Iterate until convergence.

The algorithm gives $\left\{\mathrm{E}\left(y_{t,1}^*|Y_T^+\right)\right\}$ as a by-product.

3 Mixed-Frequency Factor Model

3.1 Factor Model

Instead of a VAR model, assume a K-factor model for $\{y_t^*\}$ such that for all t,

$$
\begin{aligned}
y_t^* &= \mu^* + \Lambda f_t + u_t, \qquad (23) \\
\Phi_f(L)f_t &= v_t, \qquad (24) \\
\Phi_u(L)u_t &= w_t, \qquad (25) \\
\left\{\begin{pmatrix} v_t \\ w_t \end{pmatrix}\right\} &\sim \mathrm{NID}\left(0, \begin{bmatrix} \Sigma_{vv} & 0 \\ 0 & \Sigma_{ww} \end{bmatrix}\right), \qquad (26)
\end{aligned}
$$

where $\Phi_f(.)$ is the pth-order polynomial on $\Re^{K\times K}$ and $\Phi_u(.)$ is the qth-order polynomial on $\Re^{N\times N}$. For identification, assume that

$$
\Lambda := \begin{bmatrix} I_K \\ \Lambda_2 \end{bmatrix}.
$$

and that $\Phi_u(.)$ and Σ_{ww} are diagonal. Identification also requires K to be small relative to N, but this restriction depends on serial dependence of $\{y^*\}$.

3.2 State-Space Representation

Assume that $p, q \leq 5$. Let for all t,

$$s_t := \begin{pmatrix} f_t \\ \vdots \\ f_{t-4} \\ u_t \\ \vdots \\ u_{t-4} \end{pmatrix}.$$

A state-space representation of the factor model is for all t,

$$\begin{aligned} s_{t+1} &= As_t + Bz_t, & (27) \\ y_t &= \mu + Cs_t, & (28) \\ \{z_t\} &\sim \mathrm{NID}(0, I_{K+N}), & (29) \end{aligned}$$

where

$$\begin{aligned} A &:= \left[\begin{array}{cc} \begin{matrix} \Phi_{f,1} & \dots & \Phi_{f,p} & O_{K\times(5-p)K} \\ I_K & & 0 & O_{K\times K} \\ & \ddots & & \vdots \\ 0 & & I_K & O_{K\times K} \end{matrix} & O_{5K\times 5N} \\ O_{5N\times 5K} & \begin{matrix} \Phi_{u,1} & \dots & \Phi_{u,q} & O_{N\times(5-q)N} \\ I_N & & 0 & O_{N\times N} \\ & \ddots & & \vdots \\ 0 & & I_N & O_{N\times N} \end{matrix} \end{array} \right], \\ B &:= \begin{bmatrix} \begin{matrix} \Sigma_{vv}^{1/2} \\ O_{4K\times K} \end{matrix} & O_{5K\times N} \\ O_{5N\times K} & \begin{matrix} \Sigma_{ww}^{1/2} \\ O_{4N\times N} \end{matrix} \end{bmatrix}, \\ C &:= \begin{bmatrix} H_0\Lambda & \dots & H_4\Lambda & H_0 & \dots & H_4 \end{bmatrix}. \end{aligned}$$

3.3 ML Estimation

Using SsfPack 2.2 on Ox 3.3, ML estimation of a linear Gaussian state-space model by a quasi-Newton method is straightforward. The EM algorithm does not apply when the measurement equation has unknown parameters and does not have an error term, however, since the unknown parameters in the measurement equation do not appear in the complete-data likelihood function. According to Ruud (1991, p. 310), "the transformation τ from latent to observable data cannot depend on parameters to be estimated in such a way that the support of y^* conditional on y depends on θ."

To apply the EM algorithm, Watson and Engle (1983, p. 397) give an alternative state-space representation of a factor model. With mixed-frequency series, however, their representation is not applicable here. Thus we apply a quasi-Newton method from an ad hoc initial guess.

Given the parameters, we have for $t = 1, \dots, T$,

$$\begin{aligned} \mathrm{E}\left(y_{t,1}^* | Y_T^+\right) &= \mu_1^* + \lambda_1' \mathrm{E}\left(f_t | Y_T^+\right) + \mathrm{E}\left(u_{t,1} | Y_T^+\right) \\ &= \mu_1^* + \mathrm{E}\left(f_{t,1} | Y_T^+\right) + \mathrm{E}\left(u_{t,1} | Y_T^+\right). \end{aligned}$$

Table 1: US Coincident Indicators

Indicator	Description
	Quarterly
GDP	Real GDP (billions of chained 2000 dollars, SA, AR)
	Monthly
EMP	Employees on nonagricultural payrolls (thousands, SA)
INC	Personal income less transfer payments (billions of chained 1996 dollars, SA, AR)
IIP	Index of industrial production (1997 = 100, SA)
SLS	Manufacturing and trade sales (millions of chained 1996 dollars, SA)

Note: SA means "seasonally-adjusted" and AR means "annual rate."

Table 2: Summary Statistics

Indicator	Mean	S.D.	Min.	Max.
		Quarterly		
GDP	0.84	0.88	−2.04	3.86
		Monthly		
EMP	0.17	0.23	−0.88	1.23
INC	0.27	0.56	−4.95	3.70
IIP	0.26	0.83	−3.66	6.00
SLS	0.27	1.05	−3.21	3.54

Note: Statistics are for the first difference of the natural log times 100.

4 Application

4.1 Data

We apply the method described above to US coincident indicators to construct a new coincident index of business cycles. The component indicators are quarterly real GDP and the four monthly coincident indicators that currently make up the CI; see Table 1 for their descriptions. The sample period is from January 1959 to December 2002.

To stationarize the series, we take the first difference of the natural log of each series and multiply it 100, which is approximately equal to the quarterly or monthly percentage growth rate series. Table 2 gives summary statistics of these "growth rate" series.

4.2 VAR Coincident Index

We take two shortcuts in ML estimation of the state-space model, both of which are common and useful in practice. First, we "demean" the series, and delete the constant term from the model; this reduces the number of the parameters by N. Second, we use an approximate ML estimator instead of the exact one regarding the initial state for the Kalman filter, i.e., we assume that $s_0 := 0$; this avoids inversion of a large matrix, and hence saves the computational cost further. Recall that without missing observations one usually estimates a VAR model by applying OLS to the demeaned series. We take similar shortcuts here.

One must select p, the order of the VAR model for prediction of monthly real

Table 3: Model Selection (VAR Model)

p	Log-likelihood	AIC	SBIC
1	-1825.3	-3.5111	-3.6123
2	-1766.7	-3.4472	-3.6496
3	-1723.9	-3.4134	-3.7171
4	-1697.2	-3.4102	-3.8150
5	-1673.4	-3.4126	-3.9187
6	-1639.0	-3.3946	-4.0019
7	-1607.6	-3.3825	-4.0910
8	-1570.1	-3.3589	-4.1686
9	-1553.5	-3.3747	-4.2856
10	-1516.8	-3.3526	-4.3648
11	-1495.2	-3.3590	-4.4724
12	-1475.5	-3.3690	-4.5836

GDP. The two common model selection criteria are Akaike's information criterion (AIC) and Schwartz's Bayesian information criterion (SBIC). For our model,

$$\begin{aligned} \mathrm{AIC} &:= -\frac{1}{T}\left\{\ln L\left(\hat{\theta}\right) - \left[pN^2 + \frac{N(N+1)}{2}\right]\right\}, \\ \mathrm{SBIC} &:= -\frac{1}{T}\left\{\ln L\left(\hat{\theta}\right) - \frac{\ln T}{2}\left[pN^2 + \frac{N(N+1)}{2}\right]\right\}, \end{aligned}$$

where $\hat{\theta}$ is the (approximate) ML estimator of θ.

We estimate the mixed-frequency VAR model and compute the associated AIC and SBIC for $p = 1, \dots, 12$. Since a simple quasi-Newton method fails when p is large, we estimate the model in the following two steps:

1. Apply the EM algorithm to obtain a preliminary ML estimate.
2. Using this as the starting value, apply a quasi-Newton method to obtain the final ML estimate.

We use Ox 3.3 by Doornik (2001) for computation. In the second step, we also use SsfPack 2.2 by Koopman, Shephard, and Doornik (1999), which runs on Ox.

Table 3 summarizes the results. We find that AIC selects $p = 10$ while SBIC selects $p = 1$. One typically follows AIC for optimal one-step ahead prediction and SBIC for consistent model selection. Although it is unclear which is better for optimal smoothing, we follow SBIC here, preferring the simpler model.

Figure 2 plots the resulting coincident index, which we call the VAR coincident index. It seems to capture the NBER business cycle reference dates fairly well.

4.3 K-Factor Coincident Index

We construct another coincident index based on a factor model. First, we select K, p and q. AIC and SBIC are

$$\begin{aligned} \mathrm{AIC} &:= -\frac{1}{T}\left\{\ln L\left(\hat{\theta}\right) - \left[(N-K)K + pK^2 + \frac{K(K+1)}{2} + qN + N\right]\right\}, \\ \mathrm{SBIC} &:= -\frac{1}{T}\left\{\ln L\left(\hat{\theta}\right) - \frac{\ln T}{2}\left[(N-K)K + pK^2 + \frac{K(K+1)}{2} + qN + N\right]\right\}. \end{aligned}$$

Figure 2: Historical Plot of the VAR Coincident Index (1959:1=1). The vertical lines are the NBER business cycle reference dates.

Table 4 and 5 summarize the result of model selection. Since the EM algorithm is not applicable, we apply a quasi-Newton method from an ad hoc starting value. We find that numerical singularity tends to occur as K increases. This may be an identification problem as well as a numerical problem. Both AIC and SBIC select two-factor models rather than one-factor models. AIC selects VAR model, whereas SBIC selects two-factor model. Following SBIC again, we select $(K, p, q) = (2, 1, 1)$.

We can construct two different indices from a factor model. One is the smoothed estimate of latent monthly real GDP. The other is the smoothed estimate of the first common factor, or the common factor component of real GDP. Figure 3 plots these two. They seem to capture the NBER business cycle reference dates well. It appears that the 1st common factor is smoother and hence more reliable.

4.4 Comparison with Other Coincident Indices

Table 6 compares turning points determined by alternative indices with the NBER reference dates. Without an explicit criterion, it is difficult to say which index is the best. Moreover, we do not know whether the NBER reference date is the "correct" turning point. At lest, however, the turning points of the estimated monthly real GDP coincides with the NBER reference dates fairly well. Hence our proposal of considering index construction problem as interpolation problem seems approapriate.

Figure 4 plots the CI, the VAR(1) index, and the 2-factor index from 1979 to 1983, during which there are two peaks and two troughs. The VAR(1) index is clearly more volatile than the other two. In particular, it picks a small "dip" in January 1982 as the trough instead of the official trough in November 1982, and gives a "false" signal. Comparing Figure 1, we see that the 2-factor index is a

Table 4: Model Selection (One-Factor Model)

K	p	q	Log-likelihood	AIC	SBIC
1	0	0	−2026.1	−3.8635	−3.9040
1	0	1	−1941.5	−3.7125	−3.7733
1	0	2	−1884.4	−3.6136	−3.6945
1	0	3	−1866.1	−3.5885	−3.6897
1	0	4	−1861.0	−3.5883	−3.7097
1	0	5	−1854.2	−3.5848	−3.7265
1	1	0	−1953.4	−3.7275	−3.7720
1	1	1	−1884.5	−3.6062	−3.6710
1	1	2	−1827.8	−3.5082	−3.5932
1	1	3	−1808.7	−3.4814	−3.5866
1	1	4	−1804.0	−3.4819	−3.6074
1	1	5	−1797.1	−3.4784	−3.6242
1	2	0	−1950.1	−3.7231	−3.7717
1	2	1	−1882.6	−3.6045	−3.6733
1	2	2	−1826.6	−3.5079	−3.5969
1	2	3	−1807.7	−3.4814	−3.5907
1	2	4	−1802.8	−3.4817	−3.6112
1	2	5	−1795.8	−3.4778	−3.6276
1	3	0	−1949.7	−3.7242	−3.7768
1	3	1	−1882.4	−3.6061	−3.6790
1	3	2	−1826.1	−3.5088	−3.6019
1	3	3	−1807.4	−3.4828	−3.5962
1	3	4	−1804.2	−3.4862	−3.6198
1	3	5	−1795.5	−3.4792	−3.6330
1	4	0	−1949.6	−3.7260	−3.7827
1	4	1	−1882.3	−3.6078	−3.6848
1	4	2	−1825.7	−3.5098	−3.6069
1	4	3	−1807.4	−3.4846	−3.6020
1	4	4	−1802.6	−3.4850	−3.6227
1	4	5	−1795.4	−3.4808	−3.6387
1	5	0	−1949.6	−3.7279	−3.7887
1	5	1	−1882.2	−3.6095	−3.6905
1	5	2	−1825.5	−3.5114	−3.6126
1	5	3	−1806.9	−3.4856	−3.6070
1	5	4	−1802.1	−3.4859	−3.6276
1	5	5	−1794.8	−3.4817	−3.6436

Table 5: Model Selection (Two-Factor Model)

K	p	q	Log-likelihood	AIC	SBIC
2	0	0	−2018.7	−3.8571	−3.9138
2	0	1	−1896.6	−3.6350	−3.7119
2	0	2	−1874.5	−3.6026	−3.6997
2	0	3	−1856.0	−3.5769	−3.6943
2	0	4	−1853.7	−3.5820	−3.7197
2	0	5		singular	
2	1	0	−1861.3	−3.5661	−3.6390
2	1	1	−1770.3	−3.4028	−3.4959
2	1	2	−1757.9	−3.3888	−3.5021
2	1	3	−1751.2	−3.3855	−3.5191
2	1	4	−1746.3	−3.3859	−3.5397
2	1	5		singular	
2	2	0	−1852.8	−3.5574	−3.6465
2	2	1	−1767.6	−3.4052	−3.5145
2	2	2	−1755.2	−3.3913	−3.5209
2	2	3	−1747.6	−3.3864	−3.5362
2	2	4	−1742.7	−3.3865	−3.5565
2	2	5		singular	
2	3	0	−1835.8	−3.5328	−3.6381
2	3	1	−1766.8	−3.4113	−3.5368
2	3	2	−1754.0	−3.3966	−3.5424
2	3	3	−1746.0	−3.3909	−3.5568
2	3	4	−1742.2	−3.3932	−3.5794
2	3	5	−1735.9	−3.3908	−3.5973
2	4	0	−1834.2	−3.5373	−3.6588
2	4	1	−1754.7	−3.3960	−3.5377
2	4	2	−1749.1	−3.3948	−3.5568
2	4	3	−1739.8	−3.3866	−3.5688
2	4	4	−1736.2	−3.3894	−3.5918
2	4	5		singular	
2	5	0	−1828.0	−3.5332	−3.6709
2	5	1	−1761.1	−3.4157	−3.5736
2	5	2	−1743.0	−3.3909	−3.5690
2	5	3	−1738.2	−3.3912	−3.5896
2	5	4	−1733.9	−3.3926	−3.6112
2	5	5	−1729.3	−3.3934	−3.6322

Note: “singular” means that BFGS algorithm fails because of numerical singularity.

Figure 3: Historical Plot of the 2-Factor Coincident Index (1959:1=1). The vertical lines are the NBER business cycle reference dates.

Table 6: Turning Points Determined by Alternative Indices

NBER	CI	VAR(1)	2-Fac	
			GDP	1st CF
		Peaks		
1960/4	0	-1	-3	-3
1969/12	-2	-3	-3	-2
1973/11	0	0	$+1$	$+8$
1980/1	0	$+1$	0	$+1$
1981/7	$+1$	$+1$	0	$+1$
1990/7	-1	$+1$	-4	-1
2001/3	-6	-3	-3	-3
		Troughs		
1961/2	0	-3	-2	-2
1970/11	0	0	-1	0
1975/3	$+1$	0	0	0
1980/7	0	$+1$	-1	0
1982/11	$+1$	-10	0	-1
1991/3	0	0	0	0
2001/11	0	-2	-4	0

Note: The numbers are lags from the NBER business cycle reference dates.

Figure 4: Comparison of Alternative Indices from 1979 to 1983 (1980:1=1). The vertical lines are the NBER business cycle reference dates.

reasonable estiate of latent real GDP.

5 Discussion

This paper relates index construction to interpolation of real GDP, and constructs coincident indices of business cycles based on a VAR model and a two-factor model. SBIC selects a two-factor model. Still there are two possible coincident indices: one is the smoothed estimate of latent monthly real GDP, and the other is its common factor component, or the first common factor. The first common factor is smoother than monthly real GDP, but excludes the specific factor component of real GDP.

To determine turning points in monthly real GDP, we may need an an additional filter; e.g., Bry and Boschan (1971). Alternatively, one may fit a Markov-switching model; see Hamilton (1989). Harding and Pagan (2003) compare the two methods. This is a separate issue form our contribution; hence we do not discuss it in this paper.

One limitation of the paper is that we are unable to estimate some factor models because of numerical (and perhaps identification) problems. To avoid such problems, applying simulation-based Bayesian methods seems an interesting direction for future research.

Finally, our framework has other interesting applications; e.g., one can estimate monthly impulse response functions using mixed-frequency series.

References

Bernanke, B. S., Gertler, M. & Watson, M. W. (1997), 'Systematic monetary policy and the effects of oil price shocks', *Brookings Papers on Economic Activity* pp. 91–157.

Bry, G. & Boschan, C. (1971), Cyclical analysis of time series: Selected procedures and computer programs, Technical Paper 20, National Bureau of Economic Research.

Chauvet, M. (1998), 'An econometric characterization of business cycle dynamics with factor structure and regime switching', *International Economic Review* **39**, 969–996.

Chow, G. C. & Lin, A. (1971), 'Best linear unbiased interpolation, distribution, and extrapolation of time series by related series', *Review of Economics and Statistics* **53**, 372–375.

Cuche, N. A. & Hess, M. K. (2000), 'Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland', *Economic & Financial Modelling* **7**, 153–194.

Cushe, N. A. & Hess, M. K. (1999), A monthly time series of Swiss GDP: Business cycle indicator and research tool. unpublished.

de Jong, P. (1989), 'Smoothing and interpolation with the state-space model', *Journal of the American Statistical Association* **84**, 1085–1088.

de Jong, P. & MacKinnon, M. J. (1988), 'Covariances for smoothed estimates in state space models', *Biometrika* **75**, 601–602.

Diebold, F. X. & Rudebusch, G. D. (1996), 'Measuring business cycles: A modern perspective', *Review of Economics and Statistics* **78**, 67–77.

Doornik, J. A. (2001), *Ox: An Object-Oriented Matrix Language*, 4 edn, Timberlake Consultants.

Durbin, J. & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**, 357–384.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Harding, D. & Pagan, A. (2003), 'A comparison of two business cycle dating methods', *Journal of Economic Dynamics and Control* **27**, 1681–1690.

Kim, C.-J. & Nelson, C. R. (1998), 'Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching', *Review of Economics and Statistics* **80**, 188–201.

Kim, M.-J. & Yoo, J.-S. (1995), 'New index of coincident indicators: A multivariate Markov switching factor model approach', *Journal of Monetary Economics* **36**, 607–630.

Koopman, S. J., Shephard, N. & Doornik, J. A. (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal* **2**, 107–160.

Liu, H. & Hall, S. G. (2001), 'Creating high-frequency national accounts with state-space modelling: A Monte Carlo experiment', *Journal of Forecasting* **20**, 441–449.

Mariano, R. S. & Murasawa, Y. (2003), 'A new coincident index of business cycles based on monthly and quarterly series', *Journal of Applied Econometrics* **18**, 427–443.

Murasawa, Y. (2003), Statistical foundation of the composite index, Discussion Paper 2003-2, College of Economics, Osaka Prefecture University.

Ruud, P. A. (1991), 'Extensions of estimation methods using the EM algorithm', *Journal of Econometrics* **49**, 305–341.

Shumway, R. H. & Stoffer, D. S. (1982), 'An approach to time series smoothing and forecasting using the EM algorithm', *Journal of Time Series Analysis* **3**, 253–265.

Stock, J. H. & Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *NBER Macroeconomics Annual* **4**, 351–409.

Stock, J. H. & Watson, M. W. (1991), A probability model of the coincident economic indicators, *in* K. Lahiri & G. H. Moore, eds, 'Leading Economic Indicators', Cambridge University Press, chapter 4.

Stock, J. H. & Watson, M. W. (2002*a*), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. & Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics* **20**, 147–162.

Watson, M. W. & Engle, R. F. (1983), 'Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models', *Journal of Econometrics* **23**, 385–400.