Improving the Composite Indices of Business Cycles by Minimum Distance Factor Analysis

Yasutomo Murasawa
Institute of Economic Research, Kyoto University
Kyoto 606-8501, Japan (murasawa@kier.kyoto-u.ac.jp)

First draft: April 1999
This version: February 2000

Abstract

Minimum distance factor analysis (MD-FA) is applicable to stationary ergodic sequences. The estimated common factors, or factor scores, are weighted averages of the current observable variables, i.e., more informative variables receive larger weights. MD-FA of the U.S. coincident business cycle indicators (BCI) gives a new U.S. coincident composite index (CI) of business cycles. The weights on the BCIs for the new CI are similar to those for the Stock–Watson Experimental Coincident Index (XCI).

Keywords: Covariance structure; Time series; Generalized method of moments; Stock–Watson index.

1 Introduction

The traditional composite indices (CI) of business cycles, currently published by the Conference Board in the U.S., are descriptive statistics. No statistical model exists, at least explicitly, behind the method. So the statistical meaning of the resulting indices is unclear.

Assuming a linear one-factor structure for the coincident business cycle indicators (BCI), Stock and Watson (1989, 1991) obtain their Experimental Coincident Index (XCI). The XCI is an estimate of the realization of the common "business cycle factor." Their assumption that the factors are linear autoregressive (AR) processes with normal errors, however, is inconsistent with the observed asymmetry of business cycles between expansions and recessions.

Diebold and Rudebusch (1996) suggest to combine the factor model of Stock and Watson (1989, 1991) and the regime-switching model of Hamilton (1989). They assume that the mean of the common factor is a two-state first-order Markov chain. Kim and Yoo (1995), Chauvet (1998), and Kim and Nelson (1998) estimate this type of models in different ways and propose alternative coincident indices. Although their models are consistent with the asymmetry of business cycles, they may still be misspecified. In addition, estimation of the associated nonlinear state-space models is cumbersome.

In this paper, we assume a linear one-factor structure for the coincident BCIs, and apply minimum distance factor analysis (MD-FA). We do not assume a parametric model for the dynamics of the factors, but only require the observable variables to be stationary ergodic and square integrable. The advantages of this approach are (i) the obtained index is an estimate of the realization of the common "business cycle factor," (ii) the model is consistent with the asymmetry of business cycles, (iii) we do not have to worry about misspecification of the dynamics of the factors, and (iv) estimation is easy.

The obtained index is essentially a *weighted* average of the current standardized growth rates of the BCIs, where more informative BCIs receive larger weights, while the traditional CI is essentially the *simple* average. For the U.S. coincident BCIs, our estimation results show that "employees on nonagricultural payrolls (EMP)" and "index of industrial production (IIP)" are more informative, i.e., highly correlated with the common "business cycle factor," than "personal income less transfer payments (INC)" and "manufacturing and trade sales (SLS)," implying that it is more efficient to weight them accordingly. The weights for the new CI are similar to those for the XCI. In a sense, the new CI improves the traditional CI towards the XCI, although the new CI imposes less assumptions on the factors.

The plan of the paper is as follows. Section 2 defines a factor model and derives the implied autocovariance structure. Section 3 discusses identification of the model parameters. Section 4 discusses estimation of the model parameters and the realization of the common factors. Section 5 applies MD-FA to the U.S. coincident BCIs and proposes a new CI. Section 6 concludes.

2 Factor Model

Let $\{x_t\}_{t=-\infty}^{\infty} \subset L_2$ be an $N \times 1$ stationary sequence with mean μ_x and autocovariance matrix function $\Gamma_{xx}(.)$. We observe $\{x_t\}_{t=1}^{T}$. Assume a K-factor structure for $\{x_t\}_{t=-\infty}^{\infty}$, where $K < N$, such that $\forall t \in \mathbf{Z}$,

$$x_t = \mu_x + Bf_t + u_t, \tag{1}$$

where $B \in \Re^{N\times K}$ is an unknown factor loading matrix, $\{f_t\}_{t=-\infty}^{\infty}$ is an unobservable $K \times 1$ stationary sequence of common factor vectors with mean zero and autocovariance matrix function $\Gamma_{ff}(.)$, and $\{u_t\}_{t=-\infty}^{\infty}$ is an unobservable $N \times 1$ stationary sequence of specific factor vectors with mean zero and autocovariance matrix function $\Gamma_{uu}(.)$. Consider estimation of B and the realization of $\{f_t\}_{t=1}^{T}$.

In FA, for identification of B, we usually assume that (i) $\mathrm{rank}(B) = K$, (ii)

$\Gamma_{ff}(0)$ is positive definite (p.d.), and (ii) $u_{1,t}, \ldots, u_{N,t}$ are uncorrelated with each other and with f_t at all leads and lags. Then we have $\forall s \in \mathbf{Z}$,

$$\Gamma_{xx}(s) = B\Gamma_{ff}(s)B' + \Gamma_{uu}(s), \tag{2}$$

where $\Gamma_{uu}(s)$ is diagonal.

Assume that we use $\Gamma_{xx}(0), \ldots, \Gamma_{xx}(S)$, where $S < T$, for estimation of B. Let

$$\begin{aligned} \bar{x}_T &:= \frac{1}{T-S}\sum_{t=S+1}^{T} x_t, \\ \hat{\Gamma}_{xx,T}(s) &:= \frac{1}{T-S}\sum_{t=S+1}^{T} (x_t - \bar{x}_T)(x_{t-s} - \bar{x}_T)', \quad s = 0, \ldots, S. \end{aligned}$$

Under certain conditions which we clarify later, $\forall s \in \{0, \ldots, S\}$, $\hat{\Gamma}_{xx,T}(s)$ is consistent for $\Gamma_{xx}(s)$ as $T \to \infty$. Given this, the first issue is identification of B from $\Gamma_{xx}(0), \ldots, \Gamma_{xx}(S)$. Since (2) still has rotational indeterminacy, we need further restrictions on the model.

3 Identification

3.1 Principal Factor Model

As additional identification restrictions, we often assume that (i) $\Gamma_{ff}(0) = I_K$, (ii) $\Gamma_{uu}(0)$ is given and p.d., and (iii) $B'\Gamma_{uu}(0)^{-1}B$ is diagonal, i.e., the columns of $\Gamma_{uu}(0)^{-1/2}B$ are orthogonal. These restrictions relate FA to principal component analysis (PCA).

Consider identification of B. Since $\Gamma_{uu}(0)$ is given and p.d., rearranging (2),

$$\Gamma_{uu}(0)^{-1/2}\Gamma_{xx}(0)\Gamma_{uu}(0)^{-1/2} - I_N = \Gamma_{uu}(0)^{-1/2}BB'\Gamma_{uu}(0)^{-1/2}. \tag{3}$$

Since $\operatorname{rank}(\Gamma_{uu}(0)) = N$ and $\operatorname{rank}(B) = K < N$,

$$\begin{aligned} \operatorname{rank}\left(\Gamma_{uu}(0)^{-1/2}\Gamma_{xx}(0)\Gamma_{uu}(0)^{-1/2} - I_N\right) &= \operatorname{rank}\left(\Gamma_{uu}(0)^{-1/2}BB'\Gamma_{uu}(0)^{-1/2}\right) \\ &= K. \end{aligned}$$

Let

$$\Lambda := \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_K \end{bmatrix}, \quad W := [\, w_1 \quad \ldots \quad w_K \,],$$

where λ_k is the kth largest eigenvalue of the left-hand side of (3), and w_k is the associated normalized eigenvector. By the eigenvalue decomposition,

$$\Gamma_{uu}(0)^{-1/2}\Gamma_{xx}(0)\Gamma_{uu}(0)^{-1/2} - I_N = W\Lambda W'. \tag{4}$$

Since the columns of $\Gamma_{uu}(0)^{-1/2}B$ are orthogonal, comparing (3) and (4),

$$\Gamma_{uu}(0)^{-1/2}B = W\Lambda^{1/2},$$

or

$$B = \Gamma_{uu}(0)^{1/2}W\Lambda^{1/2}.$$

Next, consider identification of the realization of $\{f_t\}_{t=1}^T$. Notice that, given B, (1) is a generalized linear regression model for each t. Given μ_x, the GLS estimator of f_t is,

$$\begin{aligned}
\hat{f}_t &= \left(B'\Gamma_{uu}(0)^{-1}B\right)^{-1} B'\Gamma_{uu}(0)^{-1}(x_t - \mu_x) \\
&= \left(\Lambda^{1/2}W'W\Lambda^{1/2}\right)^{-1} \Lambda^{1/2}W'\Gamma_{uu}(0)^{-1/2}(x_t - \mu_x) \\
&= \Lambda^{-1/2}W'\Gamma_{uu}(0)^{-1/2}(x_t - \mu_x).
\end{aligned}$$

Notice that if we normalize x_t by $\Gamma_{uu}(0)^{-1/2}$, then FA and PCA are equivalent up to scale. Let $\tilde{x}_t := \Gamma_{uu}(0)^{-1/2}(x_t - \mu_x)$, $t = 1, \ldots, T$, and $\Gamma_{\tilde{x}\tilde{x}}(0) := \mathrm{var}(\tilde{x}_1)$. Let $\tilde{\lambda}_k$ be the kth largest eigenvalue of $\Gamma_{\tilde{x}\tilde{x}}(0) - I_N$ and $\tilde{w}_k$ be the associated normalized eigenvector, $k = 1, \ldots, K$. The kth principal component of $\tilde{x}_t$ is $\tilde{w}_k'\tilde{x}_t$. The GLS estimator, which is now equivalent to the OLS estimator, of the realization of the kth common factor of $\tilde{x}_t$ is $\tilde{w}_k'\tilde{x}_t/\sqrt{\tilde{\lambda}_k}$.

Unfortunately, this approach has some problems. First, $\Gamma_{uu}(0)$ is usually unknown in practice. Second, since the restriction on B is not explicit, it is difficult to derive the asymptotic distribution of an estimator of B. Third, it does not use $\Gamma_{xx}(1)$, …, $\Gamma_{xx}(S)$, which may contain some information.

3.2 Multivariate Errors-in-Variables Model

Alternatively, we can simply assume that $B = [I_K, B_2']'$. This restriction relates FA to multivariate errors-in-variables models (EVM).

Let $\forall t \in \mathbf{Z}$,

$$\tilde{f}_t := \begin{pmatrix} x_{1,t} - \mu_{x,1} \\ \vdots \\ x_{K,t} - \mu_{x,K} \end{pmatrix}, \quad v_t := \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{pmatrix}.$$

Then we can write (1) as

$$\begin{aligned} \tilde{f}_t &= f_t + v_t, \\ x_{K+1,t} - \mu_{x,K+1} &= \beta_{K+1}' f_t + u_{K+1,t}, \\ &\vdots \\ x_{N,t} - \mu_{x,N} &= \beta_N' f_t + u_{N,t}, \end{aligned}$$

where β_i' is the ith row of B. Eliminating f_t, we have $\forall t \in \mathbf{Z}$,

$$x_{i,t} - \mu_{x,i} = \beta_i' \tilde{f}_t - \beta_i' v_t + u_{i,t}, \quad i = K+1, \ldots, N.$$

Consider the equation for $x_{K+1,t}$. Suppose that $\forall s \in \{1, \ldots, S\}$, $\Gamma_{ff}(s) \neq 0$. Then $\forall i \in \{K+2, \ldots, N\}$, $\forall s \in \{0, \ldots, S\}$, $x_{i,t-s}$ is correlated with $\tilde{f}_t$ through f_t and f_{t-s}, but uncorrelated with $-\beta_{K+1}' v_t + u_{K+1,t}$ by our assumption. So we can use them as instrumental variables for estimation of β_{K+1}. By the order condition, we need $(N-K-1)(S+1) \geq K$, or $K \leq (N-1)(S+1)/(S+2)$, to identify β_{K+1}. The same argument applies to the other equations.

Notice that serial correlation in $\{f_t\}_{t=-\infty}^{\infty}$ helps identification of B. Without serial correlation ($S = 0$), the order condition requires that $K \leq (N-1)/2$. With sufficient serial correlation ($S \to \infty$), the order condition requires only that $K < N - 1$.

4 Estimation

4.1 Autocovariance Matrices

4.1.1 Consistency

Let

$$\gamma_0 := \begin{pmatrix} \mathrm{vech}(\Gamma_{xx}(0)) \\ \mathrm{vec}(\Gamma_{xx}(1)) \\ \vdots \\ \mathrm{vec}(\Gamma_{xx}(S)) \end{pmatrix}, \quad \hat{\gamma}_T := \begin{pmatrix} \mathrm{vech}\left(\hat{\Gamma}_{xx,T}(0)\right) \\ \mathrm{vec}\left(\hat{\Gamma}_{xx,T}(1)\right) \\ \vdots \\ \mathrm{vec}\left(\hat{\Gamma}_{xx,T}(S)\right) \end{pmatrix}.$$

We can write

$$\begin{aligned} \hat{\Gamma}_{xx,T}(s) &:= \frac{1}{T-S} \sum_{t=S+1}^{T} (x_t - \bar{x}_T)(x_{t-s} - \bar{x}_T)' \\ &= \frac{1}{T-S} \sum_{t=S+1}^{T} [(x_t - \mu_x) - (\bar{x}_T - \mu_x)][(x_{t-s} - \mu_x) - (\bar{x}_T - \mu_x)]' \\ &= \frac{1}{T-S} \sum_{t=S+1}^{T} (x_t - \mu_x)(x_{t-s} - \mu_x)' \\ &\quad -(\bar{x}_T - \mu_x)\frac{1}{T-S} \sum_{t=S+1}^{T} (x_{t-s} - \mu_x)', \quad s = 0, \ldots, S. \end{aligned}$$

Let $\forall t \in \mathbf{Z}$,

$$z_t := \begin{pmatrix} \mathrm{vech}((x_t - \mu_x)(x_t - \mu_x)') \\ \mathrm{vec}((x_t - \mu_x)(x_{t-1} - \mu_x)') \\ \vdots \\ \mathrm{vec}((x_t - \mu_x)(x_{t-S} - \mu_x)') \end{pmatrix}.$$

If $\{x_t\}_{t=-\infty}^{\infty}$ is stationary ergodic, then $\{z_t\}_{t=-\infty}^{\infty}$ is stationary ergodic; see Durrett (1996, pp. 336, 340). Note that $\mathrm{E}(z_1) = \gamma_0$.

Theorem 1 *Suppose that*

1. $\{x_t\}_{t=-\infty}^{\infty}$ *is stationary ergodic,*

2. $x_1 \in L_2$.

Then

$$\lim_{T\to\infty} \hat{\gamma}_T = \gamma_0 \quad a.s.$$

Proof. Applying the ergodic theorem to $\{x_t\}_{t=-\infty}^{\infty}$, we can write

$$\hat{\gamma}_T = \frac{1}{T-S} \sum_{t=S+1}^{T} z_t + o_{as}(1).$$

The second term is asymptotically irrelevant. The result follows by applying the ergodic theorem to $\{z_t\}_{t=-\infty}^{\infty}$. $\square$

4.1.2 Asymptotic Distribution

Let $(\Omega, \mathcal{F}, P(.))$ be the probability space under consideration. Let $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ be a filtration on $(\Omega, \mathcal{F})$. We use the notion of mixingale to simplify our discussion; see Davidson (1994, ch. 16).

Definition 1 *We say that* $\{X_t\}_{t=-\infty}^{\infty}$ *is an* L_p-mixingale *with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *if* $\exists \{c_t\}_{t=-\infty}^{\infty}, \{\psi_s\}_{s=0}^{\infty} \subset \Re_+$, *where* $\lim_{s\to\infty} \psi_s = 0$, *such that* $\forall t \in \mathbf{Z}$, $\forall s \geq 0$,

$$\|\mathrm{E}(X_t|\mathcal{F}_{t-s})\|_{L_p} \leq c_t\psi_s, \tag{5}$$

$$\|X_t - \mathrm{E}(X_t|\mathcal{F}_{t+s})\|_{L_p} \leq c_t\psi_{s+1}. \tag{6}$$

The second inequality holds trivially if $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ is adapted to $\{X_t\}_{t=-\infty}^{\infty}$. We say that a vector or matrix random sequence is an L_p-mixingale if each element is an L_p-mixingale. Note that if $\{X_t\}_{t=-\infty}^{\infty}$ is a mixingale, then $\forall t \in \mathbf{Z}$, $\mathrm{E}(X_t) = 0$.

The rate of convergence of the coefficient measures the degree of serial dependence. We say that $\{a_s\}_{s=0}^{\infty} \subset \Re_+$ is of *size* $-q$ if $\sum_{s=0}^{\infty} a_s^{1/q} < \infty$. If $a_s = O(s^{-r})$, where $r > q$, then it is of size $-q$; see Davidson (1994, p. 210). If a sequence is of size $-q$, then $\forall q' < q$, it is also of size $-q'$. We say that an L_p-mixingale is of size $-q$ if $\{\psi_s\}_{s=0}^{\infty}$ is of size $-q$.

With the notion of mixingale, we can state Gordin's central limit theorem (CLT) for stationary sequences in Durrett (1996, pp. 418–421) as follows.

Theorem 2 (Gordin's CLT) *Suppose that*

1. $\{X_t\}_{t=-\infty}^{\infty}$ *is a stationary ergodic* L_2*-mixingale with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *of size* -1,

2. $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *is adapted to* $\{X_t\}_{t=-\infty}^{\infty}$.

Then

$$s^2 := \lim_{T\to\infty} \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t\right) < \infty,$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t \xrightarrow{d} \text{N}\left(0, s^2\right).$$

If $\text{E}(X_1) \neq 0$, then the CLT applies to $\{X_t - \text{E}(X_t)\}_{t=-\infty}^{\infty}$ even if $\text{E}(X_1)$ is unknown, because $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ is adapted to $\{X_t - \text{E}(X_t)\}_{t=-\infty}^{\infty}$. The following result is now immediate.

Theorem 3 *Suppose that*

1. $\{x_t\}_{t=-\infty}^{\infty}$ *is stationary ergodic,*

2. $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *is adapted to* $\{x_t\}_{t=-\infty}^{\infty}$*,*

3. $\{x_t - \mu_x\}_{t=-\infty}^{\infty}$ *and* $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$ *are* L_2*-mixingales with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *of size* -1*.*

Then

$$\Sigma := \lim_{T\to\infty} \text{var}\left(\frac{1}{\sqrt{T-S}}\sum_{t=S+1}^{T} z_t\right) < \infty,$$

and

$$\sqrt{T-S}(\hat{\gamma}_T - \gamma_0) \xrightarrow{d} \text{N}(0, \Sigma).$$

Proof. Applying Gordin's CLT to $\{x_t - \mu_x\}_{t=-\infty}^{\infty}$, we can write

$$\sqrt{T-S}\hat{\gamma}_T = \frac{1}{\sqrt{T-S}}\sum_{t=S+1}^{T} z_t + o_p(1).$$

By asymptotic equivalence, the second term is asymptotically irrelevant. Notice that $\forall a \in \Re^{N(N+1)/2+SN^2}$, (i) $\{a' z_t\}_{t=-\infty}^{\infty}$ is stationary ergodic, (ii) $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ is adapted to $\{a'(z_t - \gamma_0)\}_{t=-\infty}^{\infty}$, and (iii) $\{a'(z_t - \gamma_0)\}_{t=-\infty}^{\infty}$ is an L_2-mixingale with respect to $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ of size -1. The result follows by Gordin's CLT and the Crámer–Wold device. □

Since a transformation of a mixingale is not a mixingale in general, we need separate mixingale conditions on $\{x_t - \mu_x\}_{t=-\infty}^{\infty}$ and $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$. Using the notion of mixing, we can give sufficient conditions on $\{x_t\}_{t=-\infty}^{\infty}$ for the mixingale conditions to hold.

Definition 2 *The* sth-order α-mixing and ϕ-mixing coefficients *of* $\{X_t\}_{t=-\infty}^{\infty}$ *are*

$$\begin{aligned} \alpha_s &:= \sup_{t \in \mathbf{Z}} \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t+s}^{\infty}} |P(A \cap B) - P(A)P(B)|, \\ \phi_s &:= \sup_{t \in \mathbf{Z}} \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t+s}^{\infty}; P(A)>0} |P(B|A) - P(B)|, \end{aligned}$$

where $\forall t \in \mathbf{Z}$, $\forall s \geq 0$, $\mathcal{F}_t^{t+s} := \sigma(X_t, \ldots, X_{t+s})$.

Definition 3 *We say that* $\{X_t\}_{t=-\infty}^{\infty}$ *is* α-mixing [ϕ-mixing] *if* $\lim_{s\to\infty} \alpha_s[\phi_s] = 0$.

Since $P(A \cap B) = P(B|A)P(A)$, ϕ-mixing implies α-mixing. As before, we say that a mixing sequence is of size $-q$ if the mixing coefficient is of size $-q$.

Mixing inequalities give the following relations between mixing sequences and mixingales; see Davidson (1994, sec. 14.2).

Theorem 4 *Suppose that*

1. $\{X_t\}_{t=-\infty}^{\infty}$ *is stationary,*
2. $X_1 \in L_p$, *where* $p > 1$,
3. $\mathrm{E}(X_1) = 0$,
4. $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *is adapted to* $\{X_t\}_{t=-\infty}^{\infty}$.

Then

1. *if* $\{X_t\}_{t=-\infty}^{\infty}$ *is* α*-mixing of size* $-q$, *where* $q > 0$, *then* $\forall p' \in [1, p)$, $\{X_t\}_{t=-\infty}^{\infty}$ *is an* $L_{p'}$*-mixingale with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *of size* $-q(1/p' - 1/p)$,
2. *if* $\{X_t\}_{t=-\infty}^{\infty}$ *is* ϕ*-mixing of size* $-q$, *where* $q > 0$, *then* $\forall p' \in [1, p]$, $\{X_t\}_{t=-\infty}^{\infty}$ *is an* $L_{p'}$*-mixingale with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *of size* $-q(1 - 1/p)$.

Proof. Since $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ is adapted to $\{X_t\}_{t=-\infty}^{\infty}$, (6) holds trivially. It remains to show that (5) also holds.

1. By Theorem 14.2 in Davidson (1994), $\forall p' \in [1, p)$, $\forall t \in \mathbf{Z}$, $\forall s \geq 0$,

$$\begin{aligned}\|\mathrm{E}(X_t|\mathcal{F}_{t-s})\|_{L_{p'}} &\leq 2\left(2^{1/p'}+1\right)\|X_{t+s}\|_{L_p}\alpha_s^{1/p'-1/p} \\ &= c\psi_s,\end{aligned}$$

   where $c := 2\left(2^{1/p'}+1\right)\|X_1\|_{L_p}$ and $\psi_s := \alpha_s^{1/p'-1/p}$. Since $\{\alpha_s\}_{s=0}^{\infty}$ is of size $-q$, $\{\psi_s\}_{s=0}^{\infty}$ is of size $-q(1/p'-1/p)$.

2. By Theorem 14.4 in Davidson (1994), $\forall p' \in [1, p]$, $\forall t \in \mathbf{Z}$, $\forall s \geq 0$,

$$\begin{aligned}\|\mathrm{E}(X_t|\mathcal{F}_{t-s})\|_{L_{p'}} &\leq 2\|X_{t+s}\|_{L_p}\phi_s^{1-1/p} \\ &= c\psi_s,\end{aligned}$$

   where $c := 2\|X_1\|_{L_p}$ and $\psi_s := \phi_s^{1-1/p}$. Since $\{\phi_s\}_{s=0}^{\infty}$ is of size $-q$, $\{\psi_s\}_{s=0}^{\infty}$ is of size $-q(1-1/p)$.

□

Next corollary gives sufficient conditions on $\{x_t\}_{t=-\infty}^{\infty}$ for $\{z_t-\gamma_0\}_{t=-\infty}^{\infty}$ to be an L_2-mixingale with respect to $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ of size -1.

Corollary 1 *Suppose that*

1. $\{x_t\}_{t=-\infty}^{\infty}$ *is stationary,*

2. $x_1 \in L_p$*, where* $p > 2$*,*

3. $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *is adapted to* $\{x_t\}_{t=-\infty}^{\infty}$.

Then

1. *if* $\{x_t\}_{t=-\infty}^{\infty}$ *is* α*-mixing of size* $-q$*, where* $q > 0$*, then* $\forall p' \in [1, p/2)$*,* $\{z_t-\gamma_0\}_{t=-\infty}^{\infty}$ *is an* $L_{p'}$*-mixingale with respect to* $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ *of size* $-q(1/p'-2/p)$*,*

2. if $\{x_t\}_{t=-\infty}^{\infty}$ is ϕ-mixing of size $-q$, where $q > 0$, then $\forall p' \in [1, p/2]$, $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$ is an $L_{p'}$-mixingale with respect to $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ of size $-q(1-2/p)$.

Proof. By Theorem 14.1 in Davidson (1994), if $\{x_t\}_{t=-\infty}^{\infty}$ is α-mixing [ϕ-mixing] of size $-q$, then $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$ is α-mixing [ϕ-mixing] of size $-q$. We have $z_1 - \gamma_0 \in L_{p/2}$, where $p > 2$, and $\mathrm{E}(z_1 - \gamma_0) = 0$. The result follows by applying the previous theorem to $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$. □

So if (i) $x_1 \in L_p$, where $p > 4$, and $\{x_t\}_{t=-\infty}^{\infty}$ is α-mixing of size $-2p/(p-4)$, or (ii) $x_1 \in L_p$, where $p \geq 4$, and $\{x_t\}_{t=-\infty}^{\infty}$ is ϕ-mixing of size $-p/(p-2)$, then $\{x_t - \mu_x\}_{t=-\infty}^{\infty}$ and $\{z_t - \gamma_0\}_{t=-\infty}^{\infty}$ are L_2-mixingales with respect to $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ of size -1.

4.1.3 Covariance Matrix Estimation

Consider estimation of Σ. If we observe $\{z_t\}_{t=S+1}^{T}$, then we can apply various heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators. Let $\Gamma_{zz}(.)$ be the autocovariance matrix function for $\{z_t\}_{t=-\infty}^{\infty}$. Then

$$\begin{aligned}
\Sigma &:= \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T-S}} \sum_{t=S+1}^{T} z_t\right) \\
&= \lim_{T\to\infty} \frac{1}{T-S} \operatorname{var}\left(\sum_{t=S+1}^{T} z_t\right) \\
&= \lim_{T\to\infty} \frac{1}{T-S} \sum_{s=-(T-S-1)}^{T-S-1} (T-S-|s|)\Gamma_{zz}(s) \\
&= \sum_{s=-\infty}^{\infty} \Gamma_{zz}(s) \\
&= \Gamma_{zz}(0) + \sum_{s=1}^{\infty} (\Gamma_{zz}(s) + \Gamma_{zz}(s)').
\end{aligned}$$

Let

$$\begin{aligned}
\bar{z}_T^* &:= \frac{1}{T-S} \sum_{t=S+1}^{T} z_t, \\
\hat{\Gamma}_{zz,T}^*(s) &:= \frac{1}{T-S} \sum_{t=s+S+1}^{T} (z_t - \bar{z}_T)(z_{t-s} - \bar{z}_T)', \quad s = 0, \ldots, T-S-1.
\end{aligned}$$

A kernel estimator of Σ is

$$\hat{\Sigma}_T^* := \hat{\Gamma}_{zz,T}^*(s) + \sum_{s=1}^{l(T)} k(s) \left(\hat{\Gamma}_{zz,T}^*(s) + \hat{\Gamma}_{zz,T}^*(s)' \right).$$

where $k(.)$ is a kernel function and $l(.)$ is a bandwidth function. Since we do not observe $\{z_t\}_{t=S+1}^T$, this estimator is infeasible.

Let

$$\begin{aligned}
z_{T,t} &:= \begin{pmatrix} \mathrm{vech}((x_t - \bar{x}_T)(x_t - \bar{x}_T)') \\ \mathrm{vec}((x_t - \bar{x}_T)(x_{t-1} - \bar{x}_T)') \\ \vdots \\ \mathrm{vec}((x_t - \bar{x}_T)(x_{t-S} - \bar{x}_T)') \end{pmatrix}, \quad t = S+1, \ldots, T, \\
\bar{z}_T &:= \frac{1}{T-S} \sum_{t=S+1}^T z_{T,t}, \\
\hat{\Gamma}_{zz,T}(s) &:= \frac{1}{T-S} \sum_{t=s+S+1}^T (z_{T,t} - \bar{z}_T)(z_{T,t-s} - \bar{z}_T)', \quad s = 0, \ldots, T-S-1.
\end{aligned}$$

Then

$$\begin{aligned}
z_{T,t} &= z_t + o_p(1), \\
\bar{z}_T &= \frac{1}{T-S} \sum_{t=S+1}^T z_t + o_p(1) \\
&= \mathrm{E}(z_t) + o_p(1), \\
\hat{\Gamma}_{zz,T}(s) &= \frac{1}{T-S} \sum_{t=s+S+1}^T (z_t - \mathrm{E}(z_t) + o_p(1))(z_t - \mathrm{E}(z_t) + o_p(1))' \\
&= \Gamma_{zz}(s) + o_p(1).
\end{aligned}$$

So we can apply a kernel estimator to $\{z_{T,t}\}_{t=S+1}^T$ to estimate Σ.

4.2 Parameters

4.2.1 Minimum Distance Estimator

Assume that $B = [I_K, B_2']'$. Then we can write (2) as $\forall s \in \mathbf{Z}$,

$$\Gamma_{xx}(s) = \begin{bmatrix} I_K \\ B_2 \end{bmatrix} \Gamma_{ff}(s) \begin{bmatrix} I_K & B_2' \end{bmatrix} + \Gamma_{uu}(s). \tag{7}$$

Let

$$\theta_0 := \begin{pmatrix} \mathrm{vec}(B_2) \\ \mathrm{vech}(\Gamma_{ff}(0)) \\ \mathrm{diag}(\Gamma_{uu}(0)) \\ \mathrm{vec}(\Gamma_{ff}(1)) \\ \mathrm{diag}(\Gamma_{uu}(1)) \\ \vdots \\ \mathrm{vec}(\Gamma_{ff}(S)) \\ \mathrm{diag}(\Gamma_{uu}(S)) \end{pmatrix}.$$

Let $\Theta \subset \Re^{(N-K)K+K(K+1)/2+N+S\left(K^2+N\right)}$ be the parameter space and $g : \Theta \to \Re^{N(N+1)/2+SN^2}$ be such that $\forall\theta \in \Theta$,

$$g(\theta) := \begin{pmatrix} \mathrm{vech}\left(\begin{bmatrix} I_K \\ B_2 \end{bmatrix} \Gamma_{ff}(0)\,[\,I_K \quad B_2'\,] + \Gamma_{uu}(0)\right) \\ \mathrm{vec}\left(\begin{bmatrix} I_K \\ B_2 \end{bmatrix} \Gamma_{ff}(1)\,[\,I_K \quad B_2'\,] + \Gamma_{uu}(1)\right) \\ \vdots \\ \mathrm{vec}\left(\begin{bmatrix} I_K \\ B_2 \end{bmatrix} \Gamma_{ff}(S)\,[\,I_K \quad B_2'\,] + \Gamma_{uu}(S)\right) \end{pmatrix}.$$

Then we have

$$\gamma_0 = g(\theta_0). \tag{8}$$

An MD estimator of θ_0 is

$$\hat{\theta}_T := \arg\min_{\theta\in\Theta} (\hat{\gamma}_T - g(\theta))' W_T (\hat{\gamma}_T - g(\theta)), \tag{9}$$

where W_T is a weighting matrix that is finite, p.d., and $\mathrm{plim}_{T\to\infty} W_T = W$, where W is fixed, finite, and p.d.

Different weighting matrices give different MD estimators of θ_0. If W_T is the identity matrix, then we have the equally-weighted MD (EMD) estimator. If W_T is such that $W = \Sigma^{-1}$, then we have an optimal MD (OMD) estimator, i.e., its asymptotic variance-covariance matrix is the smallest among the MD estimators.

4.2.2 Consistency

Let $\{Q_T : \Omega \times \Theta \to \Re\}_{T=S+1}^{\infty}$ be a sequence of random criterion functions such that $\forall T \geq S+1$, $\forall\theta \in \Theta$,

$$Q_T(\theta) := (\hat{\gamma}_T - g(\theta))' W_T (\hat{\gamma}_T - g(\theta)).$$

By definition, $\forall T \geq S+1$,

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} Q_T(\theta).$$

Let $Q : \Theta \to \Re$ be such that $\forall \theta \in \Theta$,

$$Q(\theta) := (\gamma_0 - g(\theta))' W (\gamma_0 - g(\theta)).$$

Since $\gamma_0 = g(\theta_0)$ and W is p.d., given the identification restrictions,

$$\theta_0 = \arg\min_{\theta \in \Theta} Q(\theta).$$

To prove consistency of $\hat{\theta}_T$, it suffices to check conditions (A)–(C) for Theorem 4.1.1 in Amemiya (1985). The only nontrivial part is in condition (C), where we have to show that $Q_T(.)$ converges in probability to $Q(.)$ uniformly on Θ. Next lemma gives sufficient conditions for this to hold.

Lemma 1 *Suppose that*

1. Θ *is compact,*

2. $\operatorname{plim}_{T\to\infty} \hat{\gamma}_T = \gamma_0$.

Then

$$\operatorname*{plim}_{T\to\infty} \sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| = 0.$$

Proof. By the triangle inequality, $\forall T \geq S+1$, $\forall \theta \in \Theta$,

$$\begin{aligned}
|Q_T(\theta) - Q(\theta)| &= |(\hat{\gamma}_T - g(\theta))' W_T (\hat{\gamma}_T - g(\theta)) - (\gamma_0 - g(\theta))' W (\gamma_0 - g(\theta))| \\
&= |[(\hat{\gamma}_T - \gamma_0) + (\gamma_0 - g(\theta))]'(W_T - W)[(\hat{\gamma}_T - \gamma_0) + (\gamma_0 - g(\theta))] \\
&\quad +[(\hat{\gamma}_T - \gamma_0) + (\gamma_0 - g(\theta))]' W [(\hat{\gamma}_T - \gamma_0) + (\gamma_0 - g(\theta))] \\
&\quad -(\gamma_0 - g(\theta))' W (\gamma_0 - g(\theta))| \\
&= |(\hat{\gamma}_T - \gamma_0)'(W_T - W)(\hat{\gamma}_T - \gamma_0) \\
&\quad +2(\hat{\gamma}_T - \gamma_0)'(W_T - W)(\gamma_0 - g(\theta)) \\
&\quad +(\gamma_0 - g(\theta))'(W_T - W)(\gamma_0 - g(\theta))
\end{aligned}$$

$$
\begin{aligned}
& +(\hat{\gamma}_T - \gamma_0)' W(\hat{\gamma}_T - \gamma_0) + 2(\hat{\gamma}_T - \gamma_0)' W(\gamma_0 - g(\theta))| \\
\leq \quad & |\hat{\gamma}_T - \gamma_0|' |W_T - W| |\hat{\gamma}_T - \gamma_0| \\
& +2|\hat{\gamma}_T - \gamma_0|' |W_T - W| |\gamma_0 - g(\theta)| \\
& +|\gamma_0 - g(\theta)|' |W_T - W| |\gamma_0 - g(\theta)| \\
& +|\hat{\gamma}_T - \gamma_0|' |W| |\hat{\gamma}_T - \gamma_0| + 2|\hat{\gamma}_T - \gamma_0|' |W| |\gamma_0 - g(\theta)|.
\end{aligned}
$$

Since Θ is compact and $g(.) \in C^0$, $\exists M < \infty$ such that $\sup_{\theta \in \Theta} |\gamma_0 - g(\theta)| \leq M$. Taking the supremum on both sides, $\forall T \geq S + 1$,

$$
\begin{aligned}
\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| \quad \leq \quad & |\hat{\gamma}_T - \gamma_0|' |W_T - W| |\hat{\gamma}_T - \gamma_0| \\
& +2|\hat{\gamma}_T - \gamma_0|' |W_T - W| M + M' |W_T - W| M \\
& +|\hat{\gamma}_T - \gamma_0|' |W| |\hat{\gamma}_T - \gamma_0| + 2|\hat{\gamma}_T - \gamma_0|' |W| M.
\end{aligned}
$$

The result follows by taking the probability limit on both sides and applying Slutsky's theorem. □

Consistency of $\hat{\theta}_T$ is now immediate.

Theorem 5 *Suppose that*

1. Θ *is compact,*

2. $\operatorname{plim}_{T \to \infty} \hat{\gamma}_T = \gamma_0$.

Then

$$
\operatorname*{plim}_{T \to \infty} \hat{\theta}_T = \theta_0.
$$

Proof. Verify the conditions for Theorem 4.1.1 in Amemiya (1985). □

4.2.3 Asymptotic Distribution

To prove asymptotic normality of $\hat{\theta}_T$, redefine $Q_T(.)$ and $Q(.)$ as $\forall \theta \in \Theta$,

$$
\begin{aligned}
Q_T(\theta) &:= \frac{1}{2}(\hat{\gamma}_T - g(\theta))' W_T (\hat{\gamma}_T - g(\theta)), \\
Q(\theta) &:= \frac{1}{2}(\gamma_0 - g(\theta))' W (\gamma_0 - g(\theta)).
\end{aligned}
$$

By differentiation, $\forall \theta \in \Theta$,

$$\begin{aligned}
\frac{\partial Q_T}{\partial \theta_i}(\theta) &= -\frac{\partial g}{\partial \theta_i}(\theta)' W_T(\hat{\gamma}_T - g(\theta)), \\
\frac{\partial Q}{\partial \theta_i}(\theta) &= -\frac{\partial g}{\partial \theta_i}(\theta)' W(\gamma_0 - g(\theta)), \\
& \quad i = 1, \ldots, (N-K)K + K(K+1)/2 + N + S\left(K^2 + N\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 Q_T}{\partial \theta_i \partial \theta_j}(\theta) &= -\frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\theta)' W_T(\hat{\gamma}_T - g(\theta)) + \frac{\partial g}{\partial \theta_i}(\theta)' W_T \frac{\partial g}{\partial \theta_j}(\theta), \\
\frac{\partial^2 Q}{\partial \theta_i \partial \theta_j}(\theta) &= -\frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\theta)' W(\gamma_0 - g(\theta)) + \frac{\partial g}{\partial \theta_i}(\theta)' W \frac{\partial g}{\partial \theta_j}(\theta), \\
& \quad i, j = 1, \ldots, (N-K)K + K(K+1)/2 + N + S\left(K^2 + N\right).
\end{aligned}$$

We need the following lemma first.

Lemma 2 *Suppose that* $\operatorname{plim}_{T\to\infty} \hat{\theta}_T = \theta_0$. *Then*

$$\operatorname*{plim}_{T\to\infty} \sup_{\theta \in K(\theta_0)} \left|\nabla^2 Q_T(\theta) - \nabla^2 Q(\theta)\right| = 0,$$

where $K(\theta_0)$ *is a compact neighborhood of* θ_0.

Proof. The proof is similar to that of Lemma 1. □

Given this, the proof of asymptotic normality of $\hat{\theta}_T$ is standard.

Theorem 6 *Suppose that*

1. $\operatorname{plim}_{T\to\infty} \hat{\theta}_T = \theta_0$,

2. $\sqrt{T-S}(\hat{\gamma}_T - \gamma_0) \xrightarrow{d} \mathrm{N}(0, \Sigma)$, *where* $\Sigma < \infty$.

Then

$$\sqrt{T-S}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} \mathrm{N}(0, V),$$

where

$$\begin{aligned}
V &:= (GWG')^{-1} GW\Sigma WG' (GWG')^{-1}, \\
G &:= \nabla g(\theta_0).
\end{aligned}$$

Proof. By the mean value theorem, $\forall T \geq S+1$, $\exists \bar{\theta}_T \in \left[\hat{\theta}_T, \theta_0\right]$ such that

$$\nabla Q_T\left(\hat{\theta}_T\right) = \nabla Q_T(\theta_0) + \nabla^2 Q_T\left(\bar{\theta}_T\right)\left(\hat{\theta}_T - \theta_0\right).$$

By the first-order condition, $\forall T \geq S+1$,

$$\nabla Q_T(\theta_0) + \nabla^2 Q_T\left(\bar{\theta}_T\right)\left(\hat{\theta}_T - \theta_0\right) = 0,$$

or

$$\begin{aligned}
\sqrt{T-S}\left(\hat{\theta}_T - \theta_0\right) &= -\nabla^2 Q_T\left(\bar{\theta}_T\right)^{-1}\sqrt{T-S}\nabla Q_T(\theta_0) \\
&= \nabla^2 Q_T\left(\bar{\theta}_T\right)^{-1} G W_T \sqrt{T-S}(\hat{\gamma}_T - \gamma_0) \\
&= (GWG')^{-1} GW\sqrt{T-S}(\hat{\gamma}_T - \gamma_0) \\
&\quad + \left[\nabla^2 Q_T\left(\bar{\theta}_T\right)^{-1} G W_T - (GWG')^{-1} GW\right] O_p(1).
\end{aligned}$$

By asymptotic equivalence, it suffices to show that the second term is $o_p(1)$. We can write $\forall T \geq S+1$,

$$\nabla^2 Q_T\left(\bar{\theta}_T\right) - GWG' = \left(\nabla^2 Q_T\left(\bar{\theta}_T\right) - \nabla^2 Q\left(\bar{\theta}_T\right)\right) + \left(\nabla^2 Q\left(\bar{\theta}_T\right) - GWG'\right).$$

Since $\bar{\theta}_T$ is consistent for θ_0, the first term is $o_p(1)$ by the previous lemma, and the second term is $o_p(1)$ by Slutsky's theorem. □

Note that if $W = \Sigma^{-1}$, then $V = \left(G\Sigma^{-1}G'\right)^{-1}$.

4.3 Factor Scores

Consider estimation of the realization of $\{f_t\}_{t=1}^T$. Stacking (1) for $t = 1, \ldots, T$, and taking B and μ_x as given, we have a seemingly unrelated regressions (SUR) model such that

$$\begin{aligned}
x_1 - \mu_x &= Bf_1 + u_1, \\
&\vdots \\
x_T - \mu_x &= Bf_T + u_T.
\end{aligned}$$

Let

$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix}, \quad f := \begin{pmatrix} f_1 \\ \vdots \\ f_T \end{pmatrix}, \quad u := \begin{pmatrix} u_1 \\ \vdots \\ u_T \end{pmatrix},$$

$$\Gamma_{uu} := \begin{bmatrix} \Gamma_{uu}(0) & \dots & \Gamma_{uu}(T-1) \\ \vdots & \ddots & \vdots \\ \Gamma_{uu}(T-1) & \dots & \Gamma_{uu}(0) \end{bmatrix}.$$

Let $l_T := (1, \dots, 1)'$. Then we have

$$x - l_T \otimes \mu_x = (I_T \otimes B)f + u,$$

where $\mathrm{E}(u) = 0$ and $\mathrm{var}(u) = \Gamma_{uu}$. The infeasible GLS estimator of f is

$$\hat{f}_{GLS,T} = \left[(I_T \otimes B)'\Gamma_{uu}^{-1}(I_T \otimes B)\right]^{-1} (I_T \otimes B)'\Gamma_{uu}^{-1}(x - l_T \otimes \mu_x).$$

By the Gauss–Markov theorem, $\hat{f}_{GLS,T}$ is the BLUE of f.

To make the GLS estimator feasible, we need a consistent estimator of the whole Γ_{uu}, which requires a parametric model for the dynamics of $\{u_t\}_{t=-\infty}^{\infty}$. Alternatively, we may give up system estimation and apply a feasible single-equation GLS estimator, i.e.,

$$\hat{f}_{T,t} = \left(\hat{B}_T'\hat{\Gamma}_{uu,T}(0)^{-1}\hat{B}_T\right)^{-1} \hat{B}_T'\hat{\Gamma}_{uu,T}(0)^{-1}(x_t - \bar{x}_T), \quad t = 1, \dots, T, \tag{10}$$

where $\hat{B}_T$ and $\hat{\Gamma}_{uu,T}(0)$ are consistent estimators of B and $\Gamma_{uu}(0)$ respectively. Although estimation errors in $\hat{B}_T$ and $\hat{\Gamma}_{uu,T}(0)$ cause a measurement-error bias in $\hat{f}_{T,t}$, it disappears as $T \to \infty$. Note that if $\{u_t\}_{t=-\infty}^{\infty}$ is serially uncorrelated, then Γ_{uu} is block-diagonal. So the full-equation GLS estimator and the single-equation GLS estimator are equivalent.

5 New Composite Index

5.1 Data

5.1.1 Business Cycle Indicators

We analyze the four BCIs in Table 1 that currently make up the U.S. coincident CI of business cycles. Our data are from CITIBASE. The sample period is 1959:1–1998:12 (480 observations). We take the first difference of the log of each series and

Table 1: Components of the U.S. Coincident Composite Index

BCI	Description
EMP	Employees on nonagricultural payrolls (thousands, SA)
INC	Personal income less transfer payments (billions of chained $, SA, AR)
IIP	Index of industrial production (1992 = 100, SA)
SLS	Manufacturing and trade sales (millions of chained $, SA)

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

Table 2: Descriptive Statistics of the Business Cycle Indicators

BCI	Mean	S.D.	Min.	Max.
EMP	0.19	0.24	−0.86	1.23
INC	0.26	0.42	−1.27	1.68
IIP	0.28	0.89	−4.25	6.00
SLS	0.29	1.05	−3.27	3.55

multiply it by 100, which is approximately equal to the monthly percentage growth rate series.

Table 2 and 3 summarize some descriptive statistics of the transformed series. We see that they have significantly different means and standard deviations (s.d.): EMP has lower mean than the others, and EMP and INC are much smoother than the other two (Table 2). We eliminate these differences by standardizing each series so that the sample mean is 0 and the sample s.d. is 1. We also see that all variables are positively correlated. In particular, EMP and IIP have the highest correlation, while INC and SLS have the lowest correlation (Table 3).

Figure 1 shows the sample autocorrelation functions of the BCIs. We see that they have different autocorrelation structures: EMP has persistent autocorrelation,

Table 3: Sample Correlation Coefficients of the Business Cycle Indicators

	EMP	INC	IIP	SLS
EMP	1.00			
INC	0.57	1.00		
IIP	0.64	0.50	1.00	
SLS	0.44	0.36	0.53	1.00

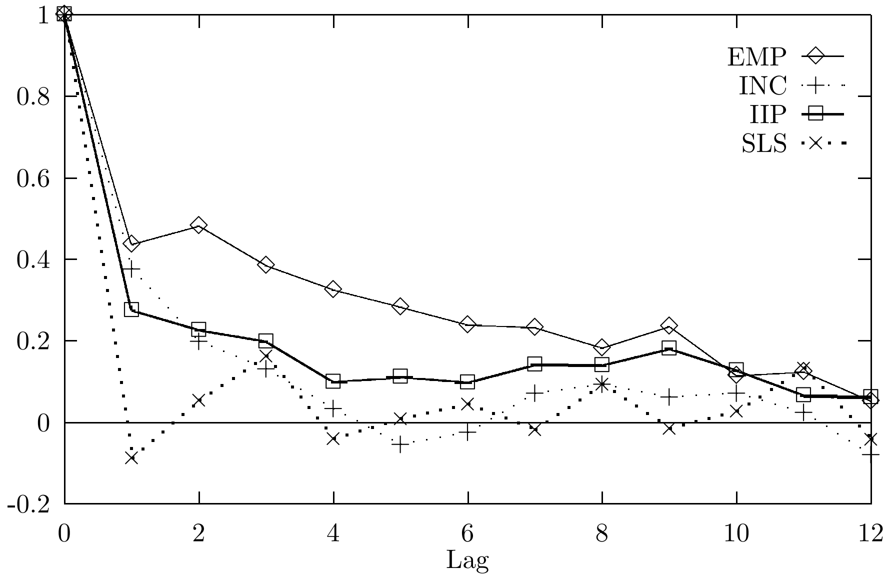

Figure 1: Sample Autocorrelation Functions of the Business Cycle Indicators

while SLS has almost no serial correlation.

5.1.2 Composite Indices

The leading, coincident, and lagging CIs of business cycles are summary statistics of the selected leading, coincident, and lagging BCIs respectively. In the U.S., the Conference Board calculates the coincident CI in the following five steps:

1. Calculate the monthly symmetric growth rate series of the BCIs.

2. Exclude outliers and normalize each symmetric growth rate series so that the sample standard deviation is 1.

3. Take the *simple* cross-section average of the normalized symmetric growth rate series. This is the monthly symmetric growth rate series of the CI.

4. Calculate the level series from the symmetric growth rate series.

5. Rebase the level series to average 100 in the base year.

See the December 1996 issue of *Business Cycle Indicators* for more details.

Table 4: Principal Component Analysis of the Business Cycle Indicators

	Principal Component			
	1st	2nd	3rd	4th
Eigenvalue	2.52	0.67	0.46	0.34
Proportion	0.63	0.17	0.12	0.09
Eigenvector				
EMP	0.53	−0.25	0.40	0.70
INC	0.48	−0.57	−0.64	−0.20
IIP	0.53	0.11	0.52	−0.66
SLS	0.45	0.78	−0.41	0.17

Note that it could be a *weighted* average in Step 3. For example, if we apply PCA, then the first principal component is a weighted average, where the weight vector is proportional to the normalized eigenvector associated with the largest eigenvalue of the sample correlation coefficient matrix.

Table 4 is the result of PCA of the U.S. coincident BCIs (in terms of the standardized first differences of their logs). The weight vector for the first principal component, which accounts for 63% of the total variation, is $(0.27, 0.24, 0.27, 0.22)'$. The weights on EMP and IIP are larger than those on INC and SLS.

PCA is still a descriptive method that reduces the dimension of a multivariate sample without assuming a statistical model. We apply FA because it is a statistical method that estimates the realization of the latent common factors underlying the observable variables.

5.2 Estimation Results

5.2.1 Parameters

We apply MD estimators to estimate a one-factor model for the standardized first differences of the logs of the U.S. coincident BCIs. Let $x_{1,t}$, …, $x_{4,t}$ be EMP, INC, IIP, and SLS respectively.

First, we have to set S, the highest order of the autocovariances included. Unfortunately, we do not have a criterion for selecting an optimal S given T. So we simply try $S = 0, 1$. Second, we have to choose W_T, a weighting matrix for the MD

Table 5: Results of Minimum Distance Estimation of the Factor Model

Parameter	EMD		OMD	
	$S=0$	$S=1$	$S=0$	$S=1$
β_1	1.00	1.00	1.00	1.00
β_2	0.81	0.86	0.89	1.08
	(0.07)	(0.08)	(0.08)	(0.12)
β_3	1.01	0.99	1.11	1.17
	(0.08)	(0.08)	(0.08)	(0.12)
β_4	0.73	0.67	0.81	0.97
	(0.06)	(0.05)	(0.07)	(0.10)
$\gamma_{ff}(0)$	0.65	0.65	0.55	0.27
	(0.14)	(0.14)	(0.13)	(0.07)
$\gamma_{uu,1}(0)$	0.35	0.35	0.35	0.25
	(0.05)	(0.06)	(0.05)	(0.04)
$\gamma_{uu,2}(0)$	0.57	0.52	0.58	0.46
	(0.08)	(0.07)	(0.08)	(0.05)
$\gamma_{uu,3}(0)$	0.35	0.35	0.32	0.27
	(0.06)	(0.05)	(0.06)	(0.05)
$\gamma_{uu,4}(0)$	0.66	0.70	0.63	0.61
	(0.06)	(0.07)	(0.06)	(0.05)

Note: Numbers in parentheses are asymptotic s.e.'s.

estimator. Although OMD estimators are asymptotically more efficient than the EMD estimator, they have small-sample bias in analysis of covariance structures; see Altonji and Segal (1996) and Clark (1996). So we try both EMD and OMD estimators. To estimate Σ, we apply a HAC covariance matrix estimator proposed by Newey and West (1994). (To implement the Newey–West estimator, we use a *GAUSS* code written by Ka-fu Wong, which is available from the *GAUSS* Sourse Code Archive at American University).

Table 5 summarizes the estimation results. Although the estimates are a little sensitive to the choice of S and W_T, EMP and IIP seem to have larger factor loadings and smaller specific-factor variances than INC and SLS. In other words, EMP and IIP are more informative about the common "business cycle factor" than INC and SLS. Hence, it is more efficient to weight them accordingly to construct a CI.

Table 6: Weights for the New Composite Index

BCI	EMD		OMD		PCA	CI
	$S=0$	$S=1$	$S=0$	$S=1$		
EMP	0.34	0.35	0.31	0.32	0.27	0.25
	(0.04)	(0.04)	(0.03)	(0.04)		
INC	0.17	0.20	0.17	0.19	0.24	0.25
	(0.03)	(0.02)	(0.03)	(0.02)		
IIP	0.35	0.34	0.38	0.36	0.27	0.25
	(0.04)	(0.04)	(0.05)	(0.04)		
SLS	0.13	0.12	0.14	0.13	0.22	0.25
	(0.02)	(0.02)	(0.02)	(0.02)		

Note: Numbers in parentheses are asymptotic s.e.'s.

5.2.2 Weights for the New Composite Index

In this application, (1) becomes

$$x_t = \beta f_t + u_t. \tag{11}$$

Given the model parameters, the single-equation GLS estimator of f_t is

$$\begin{aligned}\hat{f}_t &= \left(\beta'\Gamma_{uu}(0)^{-1}\beta\right)^{-1}\beta'\Gamma_{uu}(0)^{-1}x_t\\ &\propto \sum_{i=1}^{4}\frac{\beta_i}{\gamma_{uu,i}(0)}x_{i,t},\end{aligned}$$

i.e., $\hat{f}_t$ is essentially a weighted average of $x_{1,t}, \ldots, x_{4,t}$. Since the weight on $x_{i,t}$ is $\beta_i/\gamma_{uu,i}(0)$, more "informative" variables receive larger weights.

Table 6 summarizes estimates of the weight vector. The weights are not equal but larger on EMP and IIP than on INC and SLS. We call the CI associated with one of these weight vectors the "new CI."

5.3 Comparison with Other Indices

5.3.1 Stock–Watson Experimental Coincident Index

The Stock–Watson XCI assumes that $\forall t \in \mathbf{Z}$,

$$\begin{aligned}\phi_f(L)f_t &= v_t,\\ \Phi_u(L)u_t &= w_t,\\ \begin{pmatrix} v_t \\ w_t \end{pmatrix} &\sim \mathrm{NID}\left(0, \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \Sigma_{ww} \end{bmatrix}\right),\end{aligned}$$

where L is the lag operator, $\phi_f(.)$ is a pth-order polynomial on $\Re$, and $\Phi_u(.)$ is a qth-order polynomial on $\Re^{N\times N}$. For identification, we assume that (i) $\beta = (1, \beta_2')'$ and (ii) $\Phi_u(.)$ and Σ_{ww} are diagonal.

To obtain the maximum likelihood (ML) estimator of the model parameters, we rewrite the model into a state-space form, and apply the Kalman filter (KF) to evaluate the likelihood function; see Appendix. The XCI, obtained as a by-product of the KF, is the conditional expectation of f_t given $(y_1, \ldots, y_t)$ associated with the ML estimate of the model parameters.

To determine p and q, we use a model selection criterion such as Akaike's information criterion (AIC) or Schwartz's Bayesian information criterion (BIC). In our case, AIC and BIC are defined as

$$\begin{aligned}
AIC &:= \frac{1}{T-q}\left\{\ln L\left(\hat{\theta}_{\mathrm{ML}}\right) - \left[(N-K)K + K^2 p + Nq\right]\right\}, \\
BIC &:= \frac{1}{T-q}\left\{\ln L\left(\hat{\theta}_{\mathrm{ML}}\right) - \frac{\ln(T-q)}{2}\left[(N-K)K + K^2 p + Nq\right]\right\},
\end{aligned}$$

where $N = 4$, $K = 1$, and $T = 479$. As shown in Table 7, AIC selects $(p, q) = (1, 2)$ while BIC selects $(p, q) = (1, 0)$. Although AIC selects too large models with positive probability as $T \to \infty$, selecting too large models may be less harmful than selecting too small models. The likelihood ratio (LR) test statistic for testing $H_0 : (p, q) = (1, 0)$ against $H_1 : (p, q) = (1, 2)$ is 33.90. The asymptotic distribution of the LR test statistic under H_0 is $\chi^2(8)$. Since the test strongly rejects H_0 in favor of H_1, we select $(p, q) = (1, 2)$.

Table 8 presents the ML estimates of the model parameters for $(p, q) = (1, 2)$. Not surprisingly, the ML estimate of the factor loading vector is close to the MD estimates in Table 5.

Table 9 shows the implicit weights on the coincident BCIs for the XCI associated with the steady state KF. The weights on the current coincident BCIs for the XCI are close to those for the new CI in Table 6. The difference is that the XCI puts

Table 7: Lag-Order Selection for the Factor Model

(p, q)	$\ln L\left(\hat{\theta}_{\mathrm{ML}}\right)$	AIC	BIC
(0,1)	−592.74	−1.265	−1.317
(0,2)	−580.26	−1.250	−1.320
(0,3)	−576.16	−1.252	−1.340
(1,0)	−584.01	−1.238	−1.277
(1,1)	−577.24	−1.235	−1.292
(1,2)	−567.06	−1.224	−1.299
(1,3)	−562.90	−1.227	−1.319
(2,0)	−582.84	−1.238	−1.281
(2,1)	−576.48	−1.235	−1.296
(2,2)	−566.78	−1.226	−1.305
(2,3)	−562.71	−1.228	−1.325
(3,0)	−582.72	−1.239	−1.287
(3,1)	−576.37	−1.237	−1.303
(3,2)	−566.56	−1.228	−1.311
(3,3)	−562.50	−1.230	−1.331

Table 8: Result of Maximum Likelihood Estimation of the Factor Model

Parameter	EMP	INC	IIP	SLS
β	1.00	0.94	1.17	0.79
		(0.07)	(0.08)	(0.06)
ϕ_f		0.56		
		(0.05)		
σ_f^2		0.33		
		(0.04)		
$\phi_{u,1}$	0.10	−0.02	−0.05	−0.42
	(0.05)	(0.05)	(0.07)	(0.05)
$\phi_{u,2}$	0.45	0.04	−0.06	−0.21
	(0.05)	(0.05)	(0.06)	(0.05)
σ_w^2	0.32	0.57	0.32	0.56
	(0.03)	(0.04)	(0.03)	(0.04)

Note: Numbers in parentheses are asymptotic s.e.'s.

Table 9: Weights for the Stock–Watson Index

Lag	EMP	INC	IIP	SLS
(0)	0.27	0.15	0.32	0.13
(-1)	0.00	0.02	0.05	0.07
(-2)	-0.10	0.01	0.04	0.04
(-3)	-0.01	0.00	0.01	0.01
(-4)	-0.01	0.00	0.00	0.00

Table 10: Sample Correlation Coefficients of the Indices

	New CI	CI	XCI
New CI	1.000		
CI	0.986	1.000	
XCI	0.985	0.971	1.000

nonzero weights on the lagged coincident BCIs.

5.3.2 Comparison

Table 10 shows the sample correlation coefficient matrix of the three indices. Since the new CI has a higher correlation with the XCI than the traditional CI, the new CI improves the traditional CI towards the XCI.

A difficulty in comparing alternative indices is absence of a criterion for "good" indices. If we agree that real GDP is the most important coincident BCI, then a possible criterion is correlation with real GDP in quarterly growth rates. Table 11 shows the sample correlations of the three indices with real GDP. The XCI has the highest correlation with real GDP, while the new CI has the lowest. Moreover, the new CI hardly improves the traditional CI towards the XCI in quarterly series.

Table 11: Sample Correlation Coefficients of the Indices with Real GDP

	New CI	CI	XCI	Real GDP
New CI	1.0000			
CI	0.9945	1.0000		
XCI	0.9899	0.9898	1.0000	
Real GDP	0.7981	0.8126	0.8148	1.0000

This criterion implicitly assumes that a coincident index is an estimate of unobservable "monthly real GDP." For this purpose, however, we can obtain a better index by combining monthly coincident BCIs and quarterly real GDP using state-space models. This seems to be an interesting topic for future research.

6 Discussion

We applied MD-FA to the U.S. coincident BCIs and obtained a new CI of business cycles. The new CI, as well as the XCI, puts larger weights on more informative BCIs (EMP and IIP) and smaller weights on others (INC and SLS). So it is more efficient than the traditional CI, which simply puts equal weights.

One drawback of the new CI is that it is not unique for two reasons. First, we do not know how to choose S, the highest order of the autocovariance matrices included, in finite samples. Second, we do not know which MD estimator to use in finite samples. OMD estimators are asymptotically more efficient than the EMD estimator, but feasible OMD estimators are biased in small samples. Recently, Kitamura and Stutzer (1997) proposed an estimator that avoids the small-sample bias of feasible optimal GMM estimators. Their estimator seems to be applicable to our problem.

7 Acknowledgments

This paper is based on my Ph.D. dissertation, "Minimum Distance Factor Analysis of Time Series: Theory and Application," University of Pennsylvania, 1999. I am grateful to my advisor Roberto Mariano for his guidance and encouragement. I also thank Frank Diebold, Frank Schorfheide, Robert Stein, and Yoshiyuki Takeuchi for their valuable comments and suggestions. I am solely responsible for any remaining errors.

A Stock–Watson Experimental Coincident Index

A.1 State-Space Representation

Premultiplying both sides of (11) by $\Phi_u(L)$, $\forall t \in \mathbf{Z}$,

$$y_t = \Phi_u(L)Bf_t + w_t,$$

where $y_t := \Phi_u(L)x_t$. When $p \geq q$, a state-space model for $\{y_t\}_{t=-\infty}^{\infty}$ is $\forall t \in \mathbf{Z}$,

$$\begin{aligned}
\begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \end{pmatrix} &= \begin{bmatrix} \Phi_{f,1} & \dots & \Phi_{f,p} & O_{K\times K} \\ & I_{pK} & & O_{pK\times K} \end{bmatrix} \begin{pmatrix} f_{t-1} \\ \vdots \\ f_{t-(p+1)} \end{pmatrix} + \begin{pmatrix} I_K \\ O_{pK\times K} \end{pmatrix} v_t, \\
y_t &= \begin{bmatrix} B & -\Phi_{u,1}B & \dots & -\Phi_{u,q}B & O_{N\times(p-q)K} \end{bmatrix} \begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \end{pmatrix} + w_t,
\end{aligned}$$

where o_n is the $n \times 1$ zero vector and $O_{m\times n}$ is the $m \times n$ zero matrix. When $p \leq q$, a state-space model for $\{y_t\}_{t=-\infty}^{\infty}$ is $\forall t \in \mathbf{Z}$,

$$\begin{aligned}
\begin{pmatrix} f_t \\ \vdots \\ f_{t-q} \end{pmatrix} &= \begin{bmatrix} \Phi_{f,1} & \dots & \Phi_{f,p} & O_{K\times(q-p+1)K} \\ & & I_{qK} & O_{qK\times K} \end{bmatrix} \begin{pmatrix} f_{t-1} \\ \vdots \\ f_{t-(q+1)} \end{pmatrix} + \begin{pmatrix} I_K \\ O_{qK\times K} \end{pmatrix} v_t, \\
y_t &= \begin{bmatrix} B & -\Phi_{u,1}B & \dots & -\Phi_{u,q}B \end{bmatrix} \begin{pmatrix} f_t \\ \vdots \\ f_{t-q} \end{pmatrix} + w_t.
\end{aligned}$$

In either case, we can write $\forall t \in \mathbf{Z}$,

$$s_t = Fs_{t-1} + Gv_t, \tag{12}$$

$$y_t = Hs_t + w_t, \tag{13}$$

where s_t, F, G, and H are defined appropriately.

A.2 Likelihood Function

Let Θ be the parameter space. Let $f(.;\theta)$, $\theta \in \Theta$, be a pdf of $(y_1, \dots, y_T)$. Then

$$f(y_1, \dots, y_T; \theta) = \prod_{t=1}^{T} f_{t|t-1}(y_t|y_1, \dots, y_{t-1}; \theta),$$

where $\forall t \geq 2$, $f_{t|t-1}(.|y_1, \dots, y_{t-1}; \theta)$ is a conditional pdf of y_t given $(y_1, \dots, y_{t-1})$ (for $t = 1$, it is a marginal pdf). Let $\mathcal{F}_0 := \{\emptyset, \Omega\}$ and $\forall t \geq 1$, $\mathcal{F}_t := \sigma(y_1, \dots, y_t)$.

Let $\forall t \geq 1$,

$$\begin{aligned} \mu_{t|t-1}(\theta) &:= \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1}), \\ \Sigma_{t|t-1}(\theta) &:= \mathrm{var}_\theta(y_t|\mathcal{F}_{t-1}). \end{aligned}$$

Then $\forall t \geq 1$,

$$y_t|y_1, \dots, y_{t-1} \sim \mathrm{N}\left(\mu_{t|t-1}(\theta), \Sigma_{t|t-1}(\theta)\right).$$

So $\forall t \geq 1$,

$$\begin{aligned} & f_{t|t-1}(y_t|y_1, \dots, y_{t-1}; \theta) \\ = \; & (2\pi)^{-N/2} \det\left(\Sigma_{t|t-1}(\theta)\right)^{-1/2} \\ & \exp\left(-\frac{1}{2}\left(y_t - \mu_{t|t-1}(\theta)\right)' \Sigma_{t|t-1}(\theta)^{-1}\left(y_t - \mu_{t|t-1}(\theta)\right)\right). \end{aligned}$$

The log-likelihood function for θ given $(y_1, \dots, y_T)$ is

$$\begin{aligned} \ln L(\theta; y_1, \dots, y_T) &= -\frac{NT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\det\left(\Sigma_{t|t-1}(\theta)\right) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left(y_t - \mu_{t|t-1}(\theta)\right)' \Sigma_{t|t-1}(\theta)^{-1}\left(y_t - \mu_{t|t-1}(\theta)\right). \end{aligned}$$

To evaluate the log-likelihood function, we must evaluate $\left\{\mu_{t|t-1}(\theta), \Sigma_{t|t-1}(\theta)\right\}_{t=1}^{T}$.

Let $\forall t, s \geq 0$,

$$\begin{aligned} \hat{s}_{t|s} &:= \mathrm{E}_\theta(s_t|\mathcal{F}_s), \\ P_{t|s} &:= \mathrm{var}_\theta(s_t|\mathcal{F}_s). \end{aligned}$$

From (13), $\forall t \geq 1$,

$$\begin{aligned} \mu_{t|t-1}(\theta) &= H\hat{s}_{t|t-1}, \\ \Sigma_{t|t-1}(\theta) &= HP_{t|t-1}H' + \Sigma_{ww}. \end{aligned}$$

Given θ, we can apply the KF to evaluate $\left\{\hat{s}_{t|t-1}, P_{t|t-1}\right\}_{t=1}^{T}$.

A.3 Kalman Filter

A.3.1 Initial State

First, we must specify $\hat{s}_{1|0}$ and $P_{1|0}$. To obtain the exact ML estimator, we set

$$\begin{aligned}\hat{s}_{1|0} &= \mu_s, \\ P_{1|0} &= \Gamma_{ss}(0),\end{aligned}$$

where $\mu_s := \mathrm{E}(s_1)$ and $\Gamma_{ss}(0) := \mathrm{var}(s_1)$. Since $\{s_t\}_{t=1}^{\infty}$ is stationary, taking expectations on both sides of (12),

$$\mu_s = F\mu_s.$$

Assuming that $I_{(\max\{p,q\}+1)K} - F$ is nonsingular,

$$\mu_s = 0.$$

From (12), we also get

$$\begin{aligned}\Gamma_{ss}(0) &= F\Gamma_{ss}(1)' + G\Sigma_{uu}G', \\ \Gamma_{ss}(1) &= F\Gamma_{ss}(0).\end{aligned}$$

Eliminating $\Gamma_{ss}(1)$,

$$\Gamma_{ss}(0) = F\Gamma_{ss}(0)F' + G\Sigma_{uu}G',$$

or

$$\begin{aligned}\mathrm{vec}(\Gamma_{ss}(0)) &= \mathrm{vec}(F\Gamma_{ss}(0)T') + \mathrm{vec}(G\Sigma_{uu}G') \\ &= (F \otimes F)\mathrm{vec}(\Gamma_{ss}(0)) + \mathrm{vec}(G\Sigma_{uu}G') \\ &= \left(I_{[(\max\{p,q\}+1)K]^2} - F \otimes F\right)^{-1} \mathrm{vec}(G\Sigma_{uu}G').\end{aligned}$$

In practice, we can simply set

$$\begin{aligned}\hat{s}_{0|0} &= 0, \\ P_{0|0} &= 0,\end{aligned}$$

which implies that

$$
\begin{aligned}
\hat{s}_{1|0} &= 0, \\
P_{1|0} &= G\Sigma_{vv}G'.
\end{aligned}
$$

The resulting estimator is asymptotically equivalent to the ML estimator.

A.3.2 Updating

Since $(s_0', u_1')'$ is joint normal, s_1 is normal. Similarly, s_2, s_3, ... are also normal. So $\forall t \geq 1$,

$$
s_t | y_1, \ldots, y_{t-1} \sim \mathrm{N}\left(\hat{s}_{t|t-1}, P_{t|t-1}\right).
$$

We have $\forall t \geq 1$,

$$
y_t - \hat{y}_{t|t-1} = H\left(s_t - \hat{s}_{t|t-1}\right) + v_t,
$$

where $\hat{y}_{t|t-1} := \mathrm{E}(y_t | \mathcal{F}_{t-1})$. So $\forall t \geq 1$,

$$
\begin{pmatrix} s_t \\ y_t \end{pmatrix} | y_1, \ldots, y_{t-1} \sim \mathrm{N}\left(\begin{pmatrix} \hat{s}_{t|t-1} \\ \hat{y}_{t|t-1} \end{pmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H' \\ HP_{t|t-1} & HP_{t|t-1}H' + \Sigma_{vv} \end{bmatrix}\right).
$$

Let $\forall t \geq 1$,

$$
B_t := P_{t|t-1}H'\left(HP_{t|t-1}H' + \Sigma_{vv}\right)^{-1}.
$$

The updating equations for $\hat{s}_{t|t}$ and $P_{t|t}$ given $\hat{s}_{t|t-1}$ and $P_{t|t-1}$ are $\forall t \geq 1$,

$$
\begin{aligned}
\hat{s}_{t|t} &= \hat{s}_{t|t-1} + B_t\left(y_t - H\hat{s}_{t|t-1}\right), \\
P_{t|t} &= P_{t|t-1} - B_tHP_{t|t-1}.
\end{aligned}
$$

A.3.3 Prediction

From (12), the prediction equations for $\hat{s}_{t|t-1}$ and $P_{t|t-1}$ given $\hat{s}_{t-1|t-1}$ and $P_{t-1|t-1}$ are $\forall t \geq 1$,

$$
\begin{aligned}
\hat{s}_{t|t-1} &= F\hat{s}_{t-1|t-1}, \\
P_{t|t-1} &= FP_{t-1|t-1}F' + G\Sigma_{uu}G'.
\end{aligned}
$$

Combining the updating and prediction equations, we get $\left\{\hat{s}_{t|t-1}, P_{t|t-1}\right\}_{t=1}^{T}$.

References

Altonji, J. G. & Segal, L. M. (1996), 'Small-sample bias in GMM estimation of covariance structures', *Journal of Business and Economic Statistics* **14**, 353–366.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

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

Clark, T. E. (1996), 'Small-sample properties of estimators of nonlinear models of covariance structure', *Journal of Business and Economic Statistics* **14**, 367–373.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

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

Durrett, R. (1996), *Probability: Theory and Examples*, second edn, Duxbury Press.

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

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.

Kitamura, Y. & Stutzer, M. (1997), 'An information-theoretic alternative to generalized method of moments estimation', *Econometrica* **65**, 861–874.

Newey, W. K. & West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies* **61**, 631–653.

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