FUNDAÇÃO GETULIO VARGAS

EPGE

Escola de Pós-Graduação em Economia

Ensaios Econômicos

Escola de Pós-Graduação em Economia da Fundação Getulio Vargas

N° 429 ISSN 0104-8910

The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity

João Victor Issler, Farshid Vahid

Julho de 2001

URL: http://hdl.handle.net/10438/1025

Victor Issler, João
The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity/ João Victor Issler, Farshid Vahid - Rio de Janeiro : FGV,EPGE, 2010
(Ensaios Econômicos; 429)

Inclui bibliografia.

CDD-330

The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity*

João Victor Issler†
Graduate School of Economics – EPGE
Getulio Vargas Foundation
Praia de Botafogo 190 s. 1100
Rio de Janeiro, RJ 22253-900
Brazil
jissler@fgv.br

Farshid Vahid
Department of Econometrics and Business Statistics
Monash University
P.O. Box 11E
Victoria 3800
Australia
Farshid.Vahid@BusEco.monash.edu.au

March 2001
J.E.L. Codes: C32, E32.

*Acknowledgments: João Victor Issler acknowledges the hospitality of Monash University and Farshid Vahid acknowledges the hospitality of the Getulio Vargas Foundation. We have benefited from comments and suggestions of Heather Anderson, Antonio Fiorencio, Carlos Martins-Filho, Helio Migon, and Ajax Moreira, who are not responsible for any remaining errors in this paper. João Victor Issler acknowledges the support of CNPq-Brazil and PRONEX. Farshid Vahid is grateful to the Australian Research Council Small Grant scheme and CNPq Brazil for financial assistance.

†Corresponding author.

1. Introduction

Since Burns and Mitchell (1946) there has been a great deal of interest in making inference about the "state of the economy" from sets of monthly variables that are believed to be either concurrent or to lead the economy's business cycles (the so called "coincident" and "leading" indicators respectively). Although the business-cycle status of the economy is not directly observable, our most educated estimate of it is the binary variable announced by the NBER Business Cycle Dating Committee. These announcements are based on the consensus of a panel of experts, and they are made some time (usually six months to one year) after a turning point in the business cycle has occurred. NBER summarizes its deliberations as follows:

> "The NBER does not define a recession in terms of two consecutive quarters of decline in real GNP. Rather, a recession is a recurring period of decline in total output, income, employment, and trade, usually lasting from six months to a year, and marked by widespread contractions in many sectors of the economy."
> (Quoted from http://www.nber.org/cycles.html)

The time it takes for the NBER committee to deliberate and decide that a turning point has occurred is often too long to make these announcements practically useful. This gives importance to two constructed indices, namely the coincident index and the leading indicator index. The traditional coincident index constructed by the Department of Commerce is a combination of four representative monthly variables on total output, income, employment and trade. These variables are believed to have cycles that are concurrent with the latent "business cycle" (see Burns and Mitchell 1946). The traditional leading index is then a combination of other variables that are believed to lead the coincident index. Recently, alternative "experimental" coincident and leading indices have been proposed that are based on sophisticated statistical methods of extracting a common latent dynamic factor from the coincident variables that comprise the traditional index; see, e.g., Stock and Watson (1988a, 1988b, 1989, 1991, 1993a), and Chauvet (1998).

The basic idea behind this paper is simple: use the information content in the NBER Business Cycle Dating Committee decisions, which are generally accepted as the chronology of the U.S. business-cycles[1], to construct a coincident and a leading index of economic activity.

The NBER's Dating Committee decisions have been used extensively to validate various models of economic activity. For example, to support his econometric model, Hamilton (1989) compares the smoothed probabilities of the "recessionary regime" implied by his Markov switching model with the NBER recession indicator. Since then, this has become a routine

[1] See Stock and Watson (1993a, p. 98).

exercise for evaluating variants of Markov-switching models, see Chauvet (1998) for a recent example. Stock and Watson (1993a) use the NBER recession indicator to develop a procedure to validate the predictive performance of their experimental recession index. Estrella and Mishkin (1998) use the NBER recession indicator to compare the predictive performance of potential leading indicators of economic activity. However, as far as we know, no one has actually used the NBER recession indicator to construct coincident and leading indicators. We therefore ask "Why not?". In our opinion, this is much more appealing than imposing stringent statistical restrictions to construct a common latent dynamic factor, hoping that it represents the economy's business cycle.

The method that we employ here is based on the following ingredients. First, we use a probit regression which has the NBER recession indicator as its dependent variable. Because we are interested in constructing indices of business-cycle activity, we only use the cyclical parts of the coincident series as the regressors to explain the NBER recession indicator. This ensures that noise in the coincident series does not affect the final index[2]. Second, we allow for the possibility of measurement error and simultaneity by using instrumental-variable methods when running the probit regression. Natural candidates for the instruments are the variables that are traditionally used to construct the leading index.

The coincident index proposed here is a simple fixed-weight linear combination of the coincident series. Likewise, our leading index is also a simple fixed-weight linear combination of the leading series. This means that coincident and leading indices will be readily available to all users, who will not have to wait for them to be calculated and announced by a third party. The indices constructed by The Conference Board – TCB, formerly constructed by the Department of Commerce, are used much more widely than other proposed indices, because of their ready availability.

We like to think that our method uncovers the "Missing Link" between the pioneering research of Burns and Mitchell (1946), who proposed the coincident and leading variables to be tracked over time, and the deliberations of the NBER Business Cycle Dating Committee who define a recession as: "... a recurring period of decline in total output, income, employment, and trade, usually lasting from six months to a year, and marked by widespread contractions in many sectors of the economy".

Another feature of the present research effort is that it integrates two different strands of the modern macroeconometrics literature. The first seeks to construct indices of and to forecast business-cycle activity, and is perhaps best exemplified by the work of Stock and Watson (1988a, 1988b, 1989, 1991, 1993a), the collection of papers in Lahiri and Moore (1993) and in Stock and Watson (1993b), as well as by the recent work of Chauvet (1998).

[2] The extraction of the cyclical part of the coincident series is performed using canonical-correlation analysis due to Hotelling (1935, 1936). This method is explained in Section 2.

The second seeks to characterize and test for common-cyclical features in macroeconomic data, where a business-cycle feature is regarded as a similar pattern of serial-correlation for different macroeconomic series, showing that they display short-run co-movement; see Engle and Kozicki (1993), Vahid and Engle (1993, 1997), and Hecq, Palm and Urbain (2000) for the basic theory and Engle and Issler (1995) and Issler and Vahid (2001) for applications.

The performance of our constructed coincident index is promising. In formal econometric tests it encompasses two of the most popular indices currently in use – the TCB and Stock and Watson's coincident indices. Conversely our index is not encompassed by these others. Regarding our leading index, its behavior seems to anticipate the current state of the economy quite well.

The structure of the rest of the paper is as follows. In Section 2 we present the basic ingredients of our methodology in a non-technical way, leaving the technical details for the Appendix. Section 3 presents the coincident and leading indices, and Section 4 concludes.

2. Theoretical underpinning of the indexes

In this Section we explain the method that we use for constructing the coincident and leading indices of economic activity. Technical details are included in the Appendix.

2.1. Determining a basis for the cyclical components of coincident variables

We require that the coincident index be a linear combination of the cyclical components of coincident variables. This means that in our view, the "business cycle" is a linear combination of the cycles of the four coincident series (output, income, employment and trade), and there is no unimportant cyclical fluctuation in these variables that is excluded. This contrasts with the single latent dynamic index view of a coincident index (e.g., Stock and Watson 1989 and Chauvet 1998), which restricts the "business cycle" to be a single common cyclical factor shared by the coincident variables. In order to identify the common cycle, the single latent dynamic factor approach has to allow the coincident variables to have other idiosyncratic cyclical factors, and this provides no control over how strong these idiosyncratic cycles are relative to the common cycle; see the discussion in Appendix A.1.

We define as "cyclical" any variable which can be linearly predicted from the past information set. The past information set includes lags of both sets of coincident and leading variables. The inclusion of lags of leading variables in addition to lags of coincident variables in the information set, in effect, serves two purposes. First, it combines the estimation of coincident and leading indicator indices. Second, it allows for the possibility of asymmetric cycles in coincident series by including lags of variables such as interest rates and the spread between interest rates which are known to be nonlinear processes (Anderson 1997, Balke and

Fomby 1997) as exogenous predictors. There are infinitely many linear combinations of the coincident variables that are predictable from the past, that is, that are cyclical. We use canonical-correlation analysis to find a basis for the space of these cycles.

Canonical-correlation analysis, introduced by Hotelling (1935, 1936), has been used in multivariate statistics for a long time. It was first used in multivariate time series analysis by Akaike (1976). Akaike aptly referred to the canonical variates as "the channels of information interface between the past and the present" and he referred to canonical correlations as the "strength" of these channels. We explain the concept briefly in our context, leaving more technical details for the Appendix.

Denote the set of coincident variables (income, output, employment and trade) by the vector $x_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})'$ and the set of m ($m \geq 4$) "predictors" by the vector z_t (this includes lags of x_t as well as lags of the leading variables). Canonical correlations analysis transforms x_t into four independent linear combinations $A(x_t) = (\alpha_1' x_t, \alpha_2' x_t, \alpha_3' x_t, \alpha_4' x_t)$ with the property that $\alpha_1' x_t$ is the linear combination of x_t that is most (linearly) predictable from z_t, $\alpha_2' x_t$ is the second most predictable linear combination of x_t from z_t after controlling for $\alpha_1' x_t$, and so on[3]. These linear combinations will be uncorrelated with each other and they are restricted to have unit variances so as to identify them uniquely up-to a sign change. By-products of this analysis are four linear combination of z_t, $\Gamma(z_t) = (\gamma_1' z_t, \gamma_2' z_t, \gamma_3' z_t, \gamma_4' z_t)$, with the property that $\gamma_i' z_t$ is the linear combination of z_t that has the highest squared correlation with $\alpha_i' x_t$, for $i = 1, 2, 3, 4$. Again, the elements of $\Gamma(z_t)$ are uncorrelated with each other, and they are uniquely identified up-to a sign switch with the additional restriction that all four have unit variances. The regression R^2 between $\alpha_i' x_t$ and $\gamma_i' z_t$ for $i = 1, 2, 3, 4$, which we denote by $\left(\lambda_1^2, \lambda_2^2, \lambda_3^2, \lambda_4^2\right)$, are the squared canonical correlations between x_t and z_t.

In the present application, we call $(\alpha_1' x_t, \alpha_2' x_t, \alpha_3' x_t, \alpha_4' x_t)$ the "basis cycles" in x_t. Our view that cycles are predictable from the past information, justifies using this term. It is important to note that moving from x_t to $A(x_t)$ is just a change of coordinates. In particular, no structure is placed on these variables from outside, and no information is thrown away in this transformation. Hence, the information content in $A(x_t)$ is neither more nor less than the information content in x_t.

The advantage of this basis change is that it allows us to determine if the cyclical behavior of the coincident series can be explained by less than four basis cycles. Note that in the first basis cycle, i.e. the linear combination of x_t with maximal correlation with the past, reveals the combination of coincident series with the most pronounced cyclical feature. Analogously, the linear combination associated with the minimal canonical correlation reveals the combination

[3]The fact that canonical correlations analysis studies channels of linear dependence between x and z does not necessarily imply that it will be only useful for linear multivariate analysis. By including nonlinear basis functions (e.g. Fourier series, Tchebyschev polynomials) in z, one can use canonical correlation analysis for nonlinear multivariate modelling. See Anderson and Vahid (1998) for an example and further references.

of the x_t with the weakest cyclical feature. We can use a simple statistical-test procedure to examine whether the smallest canonical correlation (or a group of canonical correlations) is statistically equal to zero; see Appendix A.2. If this hypothesis is not rejected, then the linear combination corresponding to the statistically insignificant canonical correlation cannot be predicted from the past, i.e. it is white-noise, and therefore can be dropped from the set of basis cycles. In that case, we can conclude that all cyclical behavior in the four coincident series can be written in terms of less than four basis cycles.

Hence, the use of linear combination of x_t's that are **not** associated with a zero canonical correlation is equivalent to using only the cyclical components of the coincident series. Any linear combination of the significant basis cycles is a linear combination of the cyclical components of coincident variables, which is convenient for our purposes, because it implies that our coincident index will be a linear combination of the coincident variables themselves.

If the canonical-correlation tests suggest that only one cycle is needed to explain the dependence of the four coincident variables with the past, then that unique common cycle will be the candidate for the coincident index. In such a case, our coincident index will be close to the coincident index constructed through a single hidden dynamic factor approach. However, our analysis, which is reported in Section 3, shows that this was not the case. Jumping to our results, our proposed coincident index is a linear combination of three statistically significant basis cycles that explain the NBER recession indicator.

2.2. Using probit analysis to compute coincident and leading indices

One way to estimate the weights associated with each basis cycle is to estimate a simple probit model with the NBER indicator as the binary dependent variable and the basis cycles associated with the non-zero canonical correlations as explanatory variables. Since the basis cycles are linear combinations of the four coincident series, we will ultimately end up explaining the state of the economy by a linear combination of the coincident series. This was exactly our goal from the outset – use the information content in the NBER Business Cycle Dating Committee decisions to construct a coincident index of economic activity that is a simple linear combination of the coincident series.

The above procedure may be a good first approximation for the problem at hand. However, the basis cycles series are measured with error for two reasons. First, the coincident series are subject to constant revisions; see Stock and Watson (1988a). The data that we use in our analysis is a revised version of the data that the NBER Business Cycle Dating Committee had available to them when they decided on the state of the economy (recession vs. not recession). Second, our basis cycles are **estimates** of the population linear combinations associated with the non-zero canonical correlations. Therefore, we have a typical error-in-variables problem in estimation, which calls for instrumental-variable techniques. We use the z_t variables (i.e. lags of coincident and leading variables) as instruments for the basis cycles. Notice that canonical-correlation analysis produces estimates of $\gamma_1' z_t$, $\gamma_2' z_t$, $\gamma_3' z_t$, and $\gamma_4' z_t$, which are the best linear predictors for each of the basis cycles respectively.

Several alternative estimators have been proposed for the consistent estimation of parameters of a single equation with a limited dependent variable in a simultaneous equations model. These estimators differ in their ease of calculation versus their degree of efficiency. We use the two stage conditional maximum likelihood (2SCML) estimator proposed by Rivers and Vuong (1988). In our context, and using already the empirical results stated below, we assume that the four coincident series can be explained by three significant basis cycles $\{c_{1t}, c_{2t}, c_{3t}\}$. Denoting the NBER indicator by NBER_t, the first stage of the 2SCML estimation involves regressing $\{c_{1t}, c_{2t}, c_{3t}\}$ on the instruments z_t and saving the residuals, which we denote by $\{\hat{v}_{1t}, \hat{v}_{2t}, \hat{v}_{3t}\}$. In the second stage, both the basis cycles $\{c_{1t}, c_{2t}, c_{3t}\}$ and the residuals of the first stage $\{\hat{v}_{1t}, \hat{v}_{2t}, \hat{v}_{3t}\}$ are included in the probit model:

$$\Pr\left(\text{NBER}_t = 1\right) = \Phi\left(\beta_0 + \beta_1 c_{1t} + \beta_2 c_{2t} + \beta_3 c_{3t} + \beta_4 \hat{v}_{1t} + \beta_5 \hat{v}_{2t} + \beta_6 \hat{v}_{3t}\right),$$

where Φ is the standard normal cumulative distribution function. The estimates of β_1, β_2, and β_3 from the second stage probit will be the 2SCML estimates. Our coincident index is then given by[4]:

$$\begin{aligned}
\text{Coincident index}_t &= -\left(\widehat{\beta_1} c_{1t} + \widehat{\beta_2} c_{2t} + \widehat{\beta_3} c_{3t}\right) \\
&= -\left(\widehat{\beta_1}\alpha_1' x_t + \widehat{\beta_2}\alpha_2' x_t + \widehat{\beta_3}\alpha_3' x_t\right) \\
&= -\left(\widehat{\beta_1}\alpha_1' + \widehat{\beta_2}\alpha_2' + \widehat{\beta_3}\alpha_3'\right) x_t,
\end{aligned}$$

which shows that it can be expressed as a linear combination of the coincident series x_t. Similarly, our leading index can be expressed as a linear combination of the leading series z_t.

The simple probit standard errors cannot be used for inference and have to be modified as shown by Rivers and Vuong (1988). There is an added complication here because the value of the dependent variable is set by the Business Cycle Dating Committee several periods after time t and therefore, it embodies information that cannot be predicted at time t. This fact cannot be exploited in constructing a better index because future information is not available at the time when the index is computed. However, it implies that the model has a moving average error, and therefore its standard-errors have to be modified to make inference robust to dynamic misspecification. These technicalities are explained further in Appendix A.3.

In summary, our complete statistical model is the following:

[4] The minus sign in front of $\left(\widehat{\beta_1} c_{1t} + \widehat{\beta_2} c_{2t} + \widehat{\beta_3} c_{3t}\right)$ is to ensure that the index dips down during recessions.

$$
\begin{aligned}
\text{NBER}_t &= \begin{cases} 1 & \text{if } y_t^* > 0 \\ 0 & \text{if } y_t^* \leq 0 \end{cases} \\
y_t^* &= \psi_0 + \psi' \underset{4\times 1}{x_t} + u_t \\
x_t &= \underset{4\times m}{\Pi} \underset{m\times 1}{z_t} + \varepsilon_t,
\end{aligned}
\tag{2.1}
$$

where u_t may be correlated with ε_t, u_t and ε_t are jointly normal, and Π has rank 3.

2.3. Encompassing tests for alternative coincident indices

After obtaining our coincident index, it is desirable to evaluate its performance in explaining the state of the economy relative to alternative coincident indices. Natural candidates of alternative indices are TCB's (Dept. of Commerce) and Stock and Watson's coincident indices. The first is chosen because it is a simple linear combination of the coincident series that is widely used by practitioners. The second is chosen because it is constructed using a sophisticated dynamic factor model, which is put forward as a basis for rationalizing NBER's decisions on the state of the economy; see Appendix A.1 for more details on both indices.

In principle all three coincident indices attempt to summarize the current state of the economy. Hence, it seems natural to evaluate them with respect to their ability to predict the current state of the economy represented by the binary variable announced by the NBER Business Cycle Dating Committee. The exercise would be straightforward if the underlying models were all nested within each other. However, this is not the case, and we use non-nested tests. These can be viewed as misspecification tests for our econometric model in (2.1).

Suppose we want to compare two competing coincident indices, index_{1t} and index_{2t}. The easiest test would be to include both indices in a linear probability model, which has the NBER recession indicator as its dependent variable, i.e.,

$$
\text{NBER}_t = \theta_0 + \theta_1 \mathsf{index}_{1t} + \theta_2 \mathsf{index}_{2t} + e_t \tag{2.2}
$$

If given index_{1t}, index_{2t} is insignificant in explaining NBER_t, then index_{1t} encompasses index_{2t}. If given index_{2t}, index_{1t} is insignificant in explaining NBER_t, then index_{2t} encompasses index_{1t}. Otherwise, neither encompasses the other. If linearity is of concern, one can add higher powers of index_{1t} and index_{2t} to the right-hand side of equation (2.2), for example,

$$
\text{NBER}_t = \theta_0 + \theta_1 \mathsf{index}_{1t} + \theta_1' \mathsf{index}_{1t}^2 + \theta_1'' \mathsf{index}_{1t}^3 + \theta_2 \mathsf{index}_{2t} + \theta_2' \mathsf{index}_{2t}^2 + \theta_2'' \mathsf{index}_{2t}^3 + e_t. \tag{2.3}
$$

Then, index_{1t} encompasses index_{2t} if we cannot reject that $\theta_2 = \theta_2' = \theta_2'' = 0$, index_{2t} encompasses index_{1t} if we cannot reject that $\theta_1 = \theta_1' = \theta_1'' = 0$, and neither encompasses the other in all other cases.

Of course, linear probability models are heteroskedastic and, in our context, the above equations are likely to have serially correlated errors because of the timing issue discussed above. Therefore, heteroskedasticity and serial-correlation robust covariance matrices should be used when testing the encompassing hypotheses. The issue of measurement errors in the coincident variables that constitute the indices can be taken care of by using z_t as instruments for the indices, and basing the encompassing tests on the generalized instrumental variable (or GMM) estimates and standard errors of the parameters.

Although a linear specification is not strictly appropriate for *modelling* a binary dependent variable because it can lead to predictions for $\Pr(\text{NBER}_t = 1)$ that are outside the $[0, 1]$ interval for some t, it is convenient for *testing* hypotheses without making a restrictive assumption about the functional form. The encompassing test explained above, with all its complications arising out of heteroskedasticity and serial dependence of errors and measurement error in the independent variables, can be readily computed by standard econometrics software. Alternatively, one can take the probit specification as the correct specification under the null, and design the encompassing test along the lines of the so-called "artificial regression" approach[5] described in Davidson and McKinnon (1993, pp. 523-528); we provide some technical details of the latter in Appendix A.4.

3. Calculating coincident- and leading-indicator indices

3.1. Identification of the basis cycles

We begin our analysis by considering the coincident series, which are defined in Table 1, and are plotted in Figure 1, where shaded areas represent the NBER dating of recession periods. All four series show signs of dropping during recessions, although this behavior is more pronounced for Industrial Production ($\Delta \ln Y_t$) and Employment ($\Delta \ln N_t$). These two series also show a more visible cyclical pattern, whereas, for example, it is hard to notice the cyclical pattern in Sales ($\Delta \ln S_t$) or Income ($\Delta \ln I_t$). Before modelling the joint cyclical pattern of the coincident series in ($\Delta \ln I_t,\ \Delta \ln Y_t,\ \Delta \ln N_t,\ \Delta \ln S_t$), we performed cointegration tests to verify if the series in $(\ln I_t, \ln Y_t, \ln N_t, \ln S_t)$ share a common long-run component. As in Stock and Watson (1989), we find no cointegration among these variables.

Conditional on the evidence of no cointegration for the elements of $(\ln I_t, \ln Y_t, \ln N_t, \ln S_t)$, we model them as a Vector Autoregression (VAR) in first differences. Besides ($\Delta \ln I_t,\ \Delta \ln Y_t,\ \Delta \ln N_t,\ \Delta \ln S_t$) and their lags, the VAR also contains the lags of transformed (mostly by log first differences) leading series as a conditioning set. The latter is a sensible choice because we should expect, *a priori*, that these leading series are helpful in

[5] Indeed our test regressions in equations (2.2) and (2.3) are "artificial regressions" based on, respectively, linear and flexible nonlinear specifications.

forecasting the coincident series. A list of these leading series is also presented in Table 1. They were used by Stock and Watson(1988a) and comprise a subset of the variables initially chosen by Burns and Mitchell (1946) to be leading indicators[6].

The Akaike Information Criterion chose a VAR of order 2. Conditional on a $VAR(2)$ we calculated the canonical correlations between the coincident series ($\Delta \ln I_t$, $\Delta \ln Y_t$, $\Delta \ln N_t$, $\Delta \ln S_t$) and the respective conditioning set, comprising of two lags of ($\Delta \ln I_t$, $\Delta \ln Y_t$, $\Delta \ln N_t$, $\Delta \ln S_t$) and of two lags of the leading series. The canonical-correlation test results in Table 2 allow the conclusion that there is only one linear combination of the coincident series which is white noise. Hence, the cyclical behavior of ($\Delta \ln I_t$, $\Delta \ln Y_t$, $\Delta \ln N_t$, $\Delta \ln S_t$) can be represented by three orthogonal canonical factors. These factors, (c_{1t}, c_{2t}, c_{3t}), were labelled as the coincident basis cycles and are a linear combination of the coincident series. A plot of them is presented in Figure 2. Figure 3, on the other hand, presents the estimates of the linear combinations of the leading series in the canonical-correlation analysis, $(\gamma_1' z_t, \gamma_2' z_t, \gamma_3' z_t)$, labelled leading factors.

Below, we show the linear combinations of the four coincident indicators that yield the three basis cycles:

$$\begin{bmatrix} c_{1t} \\ c_{2t} \\ c_{3t} \end{bmatrix} = \begin{bmatrix} 1.03 & 0.31 & 19.44 & -0.68 \\ -1.68 & 1.12 & 1.12 & 4.64 \\ -0.27 & 7.78 & -13.46 & -2.33 \end{bmatrix} \times \begin{bmatrix} \Delta \ln I_t \\ \Delta \ln Y_t \\ \Delta \ln N_t \\ \Delta \ln S_t \end{bmatrix} \tag{3.1}$$

A correlation matrix for all six (coincident and leading) factors is presented in Table 3. To investigate their ability in explaining NBER recessions we include in this correlation matrix the NBER recession indicator dummy (which is equal to one during periods identified by NBER as recessions and zero otherwise). As could be expected **a priori**, the first factor (either coincident or leading) is the one with the highest correlation with the NBER dummy variable, followed by the second, and finally by the third.

3.2. "The Missing Link": computing simple indices with a probit regression

As a basic procedure to calculate a coincident index, which is a linear combination of ($\Delta \ln I_t$, $\Delta \ln Y_t$, $\Delta \ln N_t$, $\Delta \ln S_t$), we estimate a probit regression of the NBER recession indicator on the three coincident basis cycles. The result of this probit estimation are presented in Table 4. The overall fit of this simple probit regression is 0.55. Using a cutoff probability of 0.5, this model has a 70% success rate in predicting recessions (94% success rate in estimating the correct state overall). Because the NBER recession indicator is a highly dependent series, there is some evidence of significant serial correlation in the pseudo-residuals of this regression.

[6] Stock and Watson smooth some of these leading indicators. Here, we make no use of such transformations.

However, this serial correlation cannot be used to improve the quality of the leading indicator index in real time, because NBER recession indicator is often announced with a considerable lag. For correct inference though, all reported standard errors are calculated to be robust to serial correlation and heteroskedasticity.

Since the three coincident factors are a linear combination of the four coincident series, from the estimated coefficients of this simple probit regression we can construct a "simple" coincident index which is a linear combination of the four coincident series. Using the weights in (3.1) and the estimated coefficients in the probit regression (Table 4), we arrive at the following coincident-indicator linear combination (ΔCI_t):

$$\Delta CI_t = 15.52 \times \Delta \ln I_t + 50.67 \times \Delta \ln Y_t + 522.83 \times \Delta \ln N_t + 17.78 \times \Delta \ln S_t. \qquad (3.2)$$

Normalizing the weights in (3.2) to add up to unity, we arrive at the following standardized coincident-indicator linear combination ($\Delta CI_t'$):

$$\Delta CI_t' = 0.03 \times \Delta \ln I_t + 0.08 \times \Delta \ln Y_t + 0.86 \times \Delta \ln N_t + 0.03 \times \Delta \ln S_t. \qquad (3.3)$$

The formula in (3.3) shows that most of the weight is given to employment. This is not surprising given our previous analysis of Figure 1, since this series has a more pronounced coherence with the NBER recession indicator.

It is interesting to compare our measure of the "simple" coincident indicator with other measures currently in the literature. The corresponding weights that are used by the Conference Board to calculate the coincident index[7] are $(0.28, 0.13, 0.48, 0.11)$. Although both indices place the largest weight on employment, the weights are quite different.

The results of the encompassing test explained in equation (2.2) are reported in Table 6. They show that our simple index encompasses the TCB index (p-value of 0.46) but it is not encompassed by it (p-value <0.01)[8]. We also compare our simple index with the coincident index[9] (XCI) proposed by Stock and Watson (1989). The encompassing tests suggest that, at usual significance levels, our index encompasses XCI but is not encompassed by it (with respective p-values of 0.40 and <0.01)[10].

3.3. "The Missing Link": computing a more sophisticated index

For a good reason, one may suspect that the simple probit regression is not an appropriate framework for revealing the implicit weights on the four coincident series used by the NBER Dating Committee. Coincident series are subject to constant revisions, i.e., they are measured with error. Moreover, our basis cycles used in probit regressions are estimates, not the

[7] The Conference Board Index is also known as the Department of Commerce Index (or the DOC Index).

[8] Corresponding p-values for the Davidson-McKinnon encompassing tests are 0.07 and <0.01 respectively.

[9] We have downloaded this series from http://ksghome.harvard.edu/~.JStock.Academic.Ksg/xri/0012/xindex.asc.

[10] Corresponding p-values for the Davidson-McKinnon encompassing tests are 0.34 and 0.06 respectively.

actual underlying cycles. Therefore, we have to use instrumental variables to consistently estimate weights in probit regressions. As discussed above, natural candidates for instruments are the leading series. Notice that canonical-correlation analysis has already produced $(\hat{\gamma}_1' z_t, \hat{\gamma}_2' z_t, \hat{\gamma}_3' z_t)$ which are the best linear predictors for each of the three basis cycles. We use the two stage conditional maximum likelihood (2SCML) estimator proposed by Rivers and Vuong (1988) to obtain instrumental variable estimates for the coefficients of each basis cycle.

The 2SCML estimates are presented in Table 5. The "implicit weights" for the instrumental-variable coincident series ($\Delta IVCI_t$) are:

$$\Delta IVCI_t = 0.02 \times \Delta \ln I_t + 0.13 \times \Delta \ln Y_t + 0.80 \times \Delta \ln N_t + 0.05 \times \Delta \ln S_t. \tag{3.4}$$

Equation (3.4) shows that, again, most of the weight is given to employment, and that employment and industrial production get 93% of the weight altogether. A plot of this index is presented in Figure 4. Once again, the striking difference between our weights and the TCB's is that Income is weighed much more heavily in the TCB index than in ours, and employment is weighed more heavily in our index than in theirs.

The encompassing test that compares our coincident index with the TCB index (reported in Table 7) suggests that our index encompasses the TCB index (p-value of 0.83) but it is not encompassed by it (p-value of 0.04)[11]. Regarding the XCI series of Stock and Watson, the encompassing tests suggest that, at usual significance levels, our index encompasses XCI, but is not encompassed by it (with respective p-values of 0.51 and <0.01)[12].

Finally, as a by-product of this analysis, we can construct a leading index, which uses the same weights estimated by instrumental-variable probit and the leading factors $(\gamma_1' z_t, \gamma_2' z_t, \gamma_3' z_t)$ weighed by their respective canonical correlations. This index is labelled $\Delta IVLI_t$ and is presented in Figure 5. Its behavior is very similar to that of our coincident index and it tracks reasonably well NBER recessions.

4. Conclusion

The basic idea behind this paper is simple: use the information content in the NBER Business Cycle Dating Committee decisions to construct a coincident index of economic activity. Although several authors have devised sophisticated coincident indices with the ultimate goal of matching NBER recessions, no one has used the information in the NBER decisions to construct a coincident index. A second ingredient of our method is that we use canonical correlation analysis to filter out the noisy information contained in the coincident series. As a result, our final index is only influenced by the cyclical components of the coincident series.

[11] Corresponding p-values for the Davidson-McKinnon encompassing tests are 0.43 and 0.01 respectively.

[12] Corresponding p-values for the Davidson-McKinnon encompassing tests are 0.34 and 0.03 respectively.

Finally, in our preferred coincident index of the U.S. business cycle, we take account of measurement error in the coincident series by using instrumental-variable methods. The resulting index is a simple linear combination of the four coincident series originally proposed by Burns and Mitchell (1946).

As explained in the Introduction, we like to think that our method uncovers the "Missing Link" between the pioneering research of Burns and Mitchell (1946) and the deliberations of the NBER Business-Cycle Dating Committee. This is a consequence of the way we have constructed our coincident and leading indices: the coincident index is a linear combination of the four coincident series proposed by Burns and Mitchell chosen to match, using an appropriate probit regression technique, the deliberations of the NBER Business Cycle Dating Committee.

Our methodology also conveniently produces a leading index of economic activity which is a linear combination of lags of coincident and leading variables. Moreover, the probit model that produces our coincident index is in fact a model of probability of recessions. Therefore, this model can easily produce estimates of the probability of a recession.

The performance of our constructed coincident index is promising. In encompassing tests, it encompassed two currently popular constructed indices – the TCB and Stock and Watson's coincident indices. However, it was not encompassed by any of them in formal testing. Some may object to our encompassing tests as being unfair on the grounds that our indices use the NBER recession indicator in their construction, while TCB and XCI indices don't. Our reply to such objections would be, "Exactly. Why do TCB and XCI indices ignore this vital piece of information in their construction?"

References

Akaike, H. (1976), "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion", in R.K. Mehra and D.G. Lainiotis (Eds) System Identification: Advances and Case Studies, New York: Academic Press, pp. 27-96.

Anderson, T. W. (1984), An Introduction to Multivariate Statistical Analysis (2nd ed.). NewYork: John Wiley and Sons.

Anderson, H. M. and F. Vahid (1998), "Testing Multiple Equation Systems for Common Nonlinear Factors", Journal of Econometrics, 84, 1 - 37.

Balke, N.S. and T.B. Fomby (1997), "Threshold Cointegration", International Economic Review, 38, 627-647.

Burns, A. F. and Mitchell, W. C.(1946), "Measuring Business Cycles," New York: 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.

Davidson, R. and MacKinnon, J.G.(1993), "Estimation and Inference in Econometrics," Oxford: Oxford University Press.

Engle, R. F. and Kozicki, S. (1993), "Testing for Common Features", Journal of Business and Economic Statistics, 11, 369-395, with discussions.

Engle, R. F. and Issler, J. V. (1995), "Estimating Sectoral Cycles Using Cointegration and Common Features", Journal of Monetary Economics, 35, 83-113.

Estrella A. and F. S. Mishkin (1998), "Predicting U.S. Recessions: Financial Variables as Leading Indicators", Review of Economics and Statistics, 80, 45 - 61.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," Chapter 19 in D.J. Aigner and A.S. Goldberger (eds) Latent Variables in Socio-Economic Models, Amsterdam: North Holland.

Hamilton, J.D. (1994), Time Series Analysis, Princeton: Princeton 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.

Hecq, A., F. Palm and J.P. Urbain (2000), "Permanent-Transitory Decomposition in VAR Models with Cointegration and Common Cycles", Oxford Bulletin of Economics and Statistics, 62, 511-532.

Hotelling, H. (1935), "The Most Predictable Criterion", Journal of Educational Psychology, 26, 139-142.

Hotelling, H. (1936), "Relations Between Two Sets of Variates", Biometrika, 28, 321-377.

Issler, J.V. and Vahid, F. (2001), "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," Journal of Monetary Economics, 47, 449-475.

Lahiri, K. and G. Moore (1993), Editors, "Leading Economic Indicators: New Approaches and Forecasting Records." New York: Cambridge University Press.

Newey, W. and K. West (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, 703-708.

Rivers, D. and Q.H. Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", Journal of Econometrics, 39, 347-366.

Stock, J. and Watson, M.(1988a), "A New Approach to Leading Economic Indicators," mimeo, Harvard University, Kennedy School of Government.

Stock, J. and Watson, M.(1988b), "A Probability Model of the Coincident Economic Indicator", NBER Working Paper # 2772.

Stock, J. and Watson, M.(1989) "New Indexes of Leading and Coincident Economic Indicators", NBER Macroeconomics Annual, pp. 351-95.

Stock, J. and Watson, M.(1991), "A Probability Model of the Coincident Economic Indicators", in "Leading Economic Indicators: New Approaches and Forecasting Records," K. Lahiri and G. Moore, Eds. New York: Cambridge University Press.

Stock, J. and Watson, M.(1993a), "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experiences," in "New Research on Business Cycles, Indicators and Forecasting," J.H. Stock and M.W. Watson, Eds., Chicago: University of Chicago Press, for NBER.

Stock, J. and Watson, M.(1993b), Editors, "New Research on Business Cycles, Indicators and Forecasting" Chicago: University of Chicago Press, for NBER.

The Conference Board (1997), "Business Cycle Indicators," Mimeo, The Conference Board, downloadable from http://www.tcb-indicators.org/bcioverview/bci4.pdf.

Vahid, F. and Engle, R. F. (1993), "Common Trends and Common Cycles", Journal of Applied Econometrics, 8, 341-360.

Vahid, F. and Engle, R.F.(1997), "Codependent Cycles," Journal Econometrics, vol. 80, pp. 199-121.

Wooldridge, J.M.(1994), "Estimation and Inference for Dependent Processes," in "Handbook of Econometrics IV," R.F. Engle and D. McFadden, Editors. Amsterdam: Elsevier Press.

A. Econometric and statistical techniques

A.1. Statistical foundation of TCB and XCI indices

A coincident index, which is widely used by practitioners, is the index constructed by The Conference Board – TCB. This coincident index is a weighted average of the coincident variables – employment, output, sales and income, where weights are the reciprocal of the standard deviation of each component's growth rate and add up to unity; see The Conference Board (1997).

Stock and Watson's experimental coincident index (XCI) is based on an "unobserved single index" or "dynamic factor" model; see Geweke(1977), for example. There, the growth rate of the four coincident series (output, sales, income and employment) share a common cycle, ΔXCI_t, which is a latent dynamic factor that represents (the change of) "the state of the economy." Denoting the growth rates of the coincident series in a vector $x_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})'$, their proposed statistical model is as follows:

$$\begin{aligned} x_t &= \beta + \gamma(L)\Delta XCI_t + u_t, \\ \phi(L)\Delta XCI_t &= \delta + \eta_t, \\ D(L)u_t &= \epsilon_t, \end{aligned} \qquad \text{(A.1)}$$

where $\phi(L)$ and $\gamma(L)$ are scalar polynomials on the lag operator L, and $D(L)$ is a matrix polynomial on L. The error structure is restricted so as to have $E\left[\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}'\right] = \text{diag}(\sigma^2_\eta, \sigma^2_{\epsilon_1}, ..., \sigma^2_{\epsilon_4})$, and $D(L) = \text{diag}[d_{ii}(L)]$, which makes innovations mutually uncorrelated.

The model (A.1) assumes that there is a single source of comovement among the growth rates of the coincident series – ΔXCI_t. Still, these series are allowed to have their own idiosyncratic cycle, since the vector of error terms u_t is composed of serially correlated components that are mutually orthogonal. Hence, each of the four coincident series in x_t has two cyclical components: a common and an idiosyncratic one. In this view, the "business cycle" is the intersection of the cycles in output, income, employment, and trade. Moreover, there is no guarantee that idiosyncratic cycles do not dominate the common cycle in explaining the variation of the four series in x_t.

In contrast, in our view, the "business cycle" is the union of the cycles in output, income, employment, and trade. There are no idiosyncratic cycles that can be put aside, the only part of x_t that we leave out is the non-cyclical combination resulting from the canonical-correlation analysis. Comparing our method with Stock and Watson's clearly shows that neither model is a special case of the other. Hence, neither model is nested within the other one, and comparisons between them have to be made using non-nested tests. Chauvet (1998) has generalized the framework in Stock and Watson by allowing a two-state mean for the latent factor ΔXCI_t in (A.1), representing recession and non-recession regimes.

A.2. Canonical correlations

Consider two (stationary) random vectors $x_t' = (x_{1t}, x_{2t}, ..., x_{nt})$ and $z_t' = (z_{1t}, z_{2t}, ..., z_{mt})$, $m \geq n$ such that:

$$\begin{pmatrix} x_t \\ z_t \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XZ} \\ \mathbf{\Sigma}_{ZX} & \mathbf{\Sigma}_{ZZ} \end{pmatrix} \right).$$

The zero mean assumption is to simplify notation and does not involve any loss of generality. Canonical-correlation analysis seeks to rotate x_t and z_t so as to maximize the correlation between their transformed images. Formally, it seeks to find matrices

$$\underset{(n\times n)}{A'} = \begin{pmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \vdots \\ \alpha_n' \end{pmatrix} \quad \text{and} \quad \underset{(n\times m)}{\Gamma'} = \begin{pmatrix} \gamma_1' \\ \gamma_2' \\ \vdots \\ \vdots \\ \gamma_n' \end{pmatrix}$$

such that:

1. The elements of $A'x_t$ have unit variance and are uncorrelated with each other:

$$E(A'x_t x_t' A) = A'\Sigma_{XX}A = \mathbf{I}_n$$

2. The elements of $\Gamma' z_t$ have unit variance and are uncorrelated with each other:

$$E(\Gamma' z_t z_t' \Gamma) = \Gamma'\Sigma_{ZZ}\Gamma = \mathbf{I}_n, \text{ and,}$$

3. The i-th element of $A'x_t$ is uncorrelated with the j-th element of $\Gamma' z_t$, $i \neq j$. For $i = j$, this correlation is a called canonical correlation, denoted by λ_i, such that:

$$E(A'x_t z_t' \Gamma) = A'\Sigma_{XZ}\Gamma = \Lambda,$$

   where,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_n \end{pmatrix}, \text{ and,}$$

$$1 \geq |\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n| \geq 0.$$

The following basic results in Anderson (1984) and Hamilton (1994) are worth reporting here.

Proposition A.1. The k-th canonical correlation between $\mathbf{x}_t$ and $\mathbf{z}_t$ is given by k-th highest root of $\begin{vmatrix} -\lambda\Sigma_{XX} & \Sigma_{XZ} \\ \Sigma_{ZX} & -\lambda\Sigma_{ZZ} \end{vmatrix} = 0$, denoted by λ_k. The linear combinations α_k and γ_k associated with λ_k can be found by making $\lambda = \lambda_k$ in $\begin{pmatrix} -\lambda\Sigma_{XX} & \Sigma_{XZ} \\ \Sigma_{ZX} & -\lambda\Sigma_{ZZ} \end{pmatrix}\begin{pmatrix} \alpha_k \\ \gamma_k \end{pmatrix} = 0$, considering also the unit-variance restrictions in 1 and 2 above.

Proposition A.2. Let $X = (x_1,x_2,...,x_T)'$ and $Z= (z_1, z_2,...,z_T)$ be samples of T observations of x_t and z_t. The n first eigenvalues of the matrix $H= (X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$ are consistent estimates of the squared populational canonical correlations $(\lambda_1^2, \lambda_2^2, ..., \lambda_n^2)$. The corresponding eigenvectors are consistent estimates of the parameters in A. Moreover, the first n eigenvalues of H are identical to the first n eigenvalues of the matrix $G = (Z'Z)^{-1}Z'X(X'X)^{-1}X'Z$, whose corresponding eigenvectors are consistent estimates of the elements of Γ.

Proposition A.3. The likelihood ratio test statistic for the null hypothesis that the smallest $n-k$ canonical correlations are jointly zero, $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_n = 0$, can be computed using the squared sample canonical correlations $\widehat{\lambda}_i^2$, $i = k+1, \cdots, n$, in the following fashion:

$$LR = -T \sum_{i=k+1}^{n} \ln(1 - \widehat{\lambda}_i^2).$$

Moreover, the asymptotic distribution of this likelihood-ratio test statistic is chi-squared, as follows:

$$LR \xrightarrow{d} \chi^2_{(n-k)(m-k)}.$$

Canonical-correlation analysis can be applied in the present context for analyzing a large multivariate data set, summarizing the correlations between a group of stationary series x and a group of stationary series z. For example, we suppose that the coincident series in x_t can be modelled using a Vector Autoregression (VAR), using its own the lags, $x_{t-1}, \cdots, x_{t-p}$, and also the lags of some other (leading) series, $w_{t-1}, \cdots, w_{t-p}$, as follows:

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + B_1 w_{t-1} + \cdots + B_p w_{t-p} + \varepsilon_t, \qquad \text{(A.2)}$$

where ε_t is a white-noise process.

Here, we are interested in summarizing the correlations between the variables in x_t and the variables in $z_t = \left(x'_{t-1}, \cdots, x'_{t-p}, w'_{t-1}, \cdots, w'_{t-p}\right)'$. In this framework, the cyclical feature in x_t has to arise from the elements in z_t, since ε_t is a white-noise process, devoid of any cyclical features; see Engle and Kozicki(1993).

A.3. Instrumental-variable probit regressions

Denoting by $c_{1t}, \cdots, c_{kt}$, $(c_{it} = \alpha_i' x_t,\ i = 1, \cdots, k)$, the k basis cycles associated with the first k non-zero canonical correlations, the NBER business-cycle indicator is linked to them through the latent variable y_t^*:

$$\begin{aligned} y_t^* &= \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + u_t \\ \text{NBER}_t &= \begin{cases} 1 & \text{if } y_t^* > 0 \\ 0 & \text{if } y_t^* \leq 0 \end{cases} . \end{aligned} \tag{A.3}$$

As argued above, because the series $c_{1t}, \cdots, c_{kt}$, are measured with error, we cannot use a simple probit-regression procedure to estimate β_i, $i = 0, 1, \cdots, k$. The possible correlation between $c_{1t}, \cdots, c_{kt}$ and the errors u_t is modelled as follows,

$$\begin{aligned} c_{it} &= \lambda_i \left(\gamma_i' z_t\right) + v_{it}, \qquad i = 1, \cdots, k \\ \begin{pmatrix} u_t \\ v_t \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{vu}' \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}\right) \end{aligned} \tag{A.4}$$

where the v_{it}, $i = 1, \cdots, k$, are collected into a k-vector v_t, λ_i and $\gamma_i' z_t$ for $i = 1, \cdots, k$ come from the canonical-correlation analysis, Σ_{vv} is a $k \times k$ diagonal variance-covariance matrix of v_t, σ_{vu} is a $k \times 1$ vector of covariances between u_t and v_t. Because of measurement error, the basis cycles $c_{1t}, \cdots, c_{kt}$ are correlated with u_t. Joint normality of u_t and v_t implies:

$$u_t = v_t' \delta + \eta_t$$

where $\delta = \Sigma_{vv}^{-1} \sigma_{vu}$, $\eta_t \sim N\left(0, \sigma_u^2 - \sigma_{vu}' \Sigma_{vv}^{-1} \sigma_{vu}\right)$ and η_t is independent of v_t. Substituting for u_t in equation (A.3), we obtain,

$$\begin{aligned} y_t^* &= \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t' \delta + \eta_t \\ \text{NBER}_t &= \begin{cases} 1 & \text{if } \eta_t < \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t' \delta \\ 0 & \text{if } \eta_t \geq \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t' \delta \end{cases} . \end{aligned} \tag{A.5}$$

Notice that, by construction, all the regressors in (A.5) are uncorrelated with the error term η_t. As usual for probit models the mean parameters $\theta = \left(\beta_0, \beta_1, \beta_2, \beta_3, \delta'\right)'$ and the variance parameter $\left(\sigma_\eta^2 = \sigma_u^2 - \sigma_{vu}' \Sigma_{vv}^{-1} \sigma_{vu}\right)$ are not separately identifiable. The convenient normalization $\sigma_\eta^2 = 1$ will identify the mean parameters. Obtaining the two stage conditional maximum likelihood (2SCML) estimator proposed by Rivers and Vuong (1988) entails the following steps:

1. Regress c_{it}, $i = 1, \cdots, k$, on z_t to get $\widehat{v}_{it}$ and $\widehat{\Sigma}_{vv}$, a consistent estimate of Σ_{vv}.

2. From $\widehat{v}_{it}$, $i = 1, \cdots, k$, form $\widehat{v}_t$ and then run a probit regression (A.5) to get consistent estimates of $\theta = \left(\beta_0, \beta_1, \beta_2, \beta_3, \delta'\right)'$, denoted by $\widehat{\theta}$.

For inference on θ, if η_t is i.i.d., the following central-limit theorem holds:

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \xrightarrow{d} N\left(0, V\left(\widehat{\theta}\right)\right),$$

where the appropriate formula for the asymptotic covariance matrix $V\left(\widehat{\theta}\right)$ is given in Rivers and Vuong(1988, p. 354).

The error term η_t in (A.5) is likely to be a moving average process since the NBER dating committee uses future information in deciding on the state of the economy. Since this future information is unpredictable at time t, it is still valid to use z_t as instruments for estimation. However, autocorrelation robust standard errors have to be used for correct inference. See Newey and West (1987) or Wooldridge (1994).

A.4. Encompassing tests for probit regressions

Davidson and McKinnon(1993) discuss testing non-nested models in a limited-dependent variable framework. They start with two competing models to explain y_t^*:

$$\begin{aligned} H_1 &: E\left(y_t^* \left|\Omega_{t-1}\right.\right) = F_1\left(X_{1t-1}\gamma_1\right) \\ H_2 &: E\left(y_t^* \left|\Omega_{t-1}\right.\right) = F_2\left(X_{2t-1}\gamma_2\right), \end{aligned} \tag{A.6}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are cumulative density functions, X_{1t-1}, X_{2t-1}, γ_1 and γ_2 are respectively the explanatory variables and associated coefficients present in models one and two. These models may differ either because $F_1(\cdot)$ and $F_2(\cdot)$ are different, because X_{1t-1} and X_{2t-1} are different, or because of both. To nest these two models in a compound model that can serve as a basis for estimation they propose:

$$H_C : E\left(y_t^* \left|\Omega_{t-1}\right.\right) = (1-\alpha) F_1\left(X_{1t-1}\gamma_1\right) + \alpha F_2\left(X_{2t-1}\widehat{\gamma}_2\right), \tag{A.7}$$

where $\widehat{\gamma}_2$, the maximum likelihood estimate of γ_2 under H_2, is necessarily used in (A.7) for its parameters to be identified.

Testing H_1 against H_C is simply a test of the null that $\alpha = 0$. i.e., a test of irrelevance of model two. Because H_1 and H_2 are non-linear models, a convenient transformation of H_C, by means of a first-order Taylor expansion is usually preferable to work with for hypothesis testing. In the present context, we consider expanding the following function:

$$g\left(\alpha, \beta_1\right) = (1-\alpha) F_1\left(X_{1t-1}\gamma_1\right) + \alpha F_2\left(X_{2t-1}\widehat{\gamma}_2\right),$$

around $\alpha = 0$, and $\beta_1 = \widehat{\beta}_1$, where $\widehat{\beta}_1$ is the probit estimate of β_1 under H_1. Then, it can be shown that testing $\alpha = 0$ in (A.7) using a t-test is asymptotically equivalent to testing $a = 0$ using a t-test in the following normalized artificial regression:

$$\widehat{V}_t^{-1/2}\left(y_t^* - \widehat{F}_{1t}\right) = \widehat{V}_t^{-1/2}\widehat{f}_{1t}X_{1t-1}b + a\widehat{V}_t^{-1/2}\left(\widehat{F}_{2t} - \widehat{F}_{1t}\right) + \text{residual}, \tag{A.8}$$

where $\widehat{F}_{1t}$ and $\widehat{f}_{1t}$ are respectively consistent estimates of the cumulative density and density functions of model one, where the maximum likelihood estimate of γ_2 is used in constructing $\widehat{F}_{2t}$, and $\widehat{V}_t^{-1/2}$ is an estimate of the conditional variance of y_t^* used to normalize the variables in H_C taking into account the fact that the regression model is heteroskedastic. An analogous test can be constructed for testing H_2 against H_C.

B. Tables and figures

Table 1: Coincident and Leading Series: Definitions and Transformations

Series Definition	Transformation
Coincident Series	
INDUSTRIAL PRODUCTION: TOTAL INDEX (1992=100,SA) – Y_t	$\Delta \ln(\cdot)$
EMPLOYEES ON NONAG. PAYROLLS: TOTAL (THOUS.,SA) – N_t	$\Delta \ln(\cdot)$
MANUFACTURING & TRADE SALES (MIL$, 92 CHAINED $) – S_t	$\Delta \ln(\cdot)$
PERS. INCOME LESS TRANSF. PMTS. (CHAINED, BIL 92$,SAAR) – I_t	$\Delta \ln(\cdot)$
Leading Series	
MFG UNFIL.ORD.: DUR.GOODS IND., TOT.(82$,SA) = MDU/PWDMD	$\Delta \ln(\cdot)$
MANUFACT. & TRADE INVENT.:TOTAL (MIL OF CHAINED 1992, SA)	$\Delta \ln(\cdot)$
NEW PRIV. OWNED HOUSING: UNITS AUTH. BUILD. PERMITS SAAR	$\Delta \ln(\cdot)$
IND. PRODUCTION: DURABLE CONSUMER GOODS (1992=100,SA)	$\Delta \ln(\cdot)$
INT. RATE: U.S.TRS. CONST MATUR.,10-YR.(% PER ANN,NSA)	$\Delta(\cdot)$
INT. RATE SPREAD = 3 MONTHS - 10 YEARS (FYGM3-FYGT10)	NONE
NOMINAL WEIGHTED EXCHANGE RATE OF G7 (EXCL. CANADA)	$\Delta \ln(\cdot)$
EMPLOYEES ON NONAG. PAYROLLS: SERVICE-PROD.(THOUS.,SA)	$\Delta \ln(\cdot)$
UNEMPL. BY DURATION: PERSONS UN. < 5 WEEKS (THOUS.,SA)	$\Delta \ln(\cdot)$

Table 2: Squared Canonical Correlations and Canonical-Correlation Test

Sq. Canonical Correlations	Degrees of Freedom	λ_j^2 *and all smaller* $\lambda_j^2 = 0$
λ_j^2		P-Values
0.4397	104	0.0000
0.2791	75	0.0000
0.1976	48	0.0000
0.0654	23	0.1332

Table 3: Correlation Matrix for Factors and NBER Recession-Indicator

	NBER	Basis Cycle 1	Basis Cycle 2	Basis Cycle3	Leading Factor 1	Leading Factor 2	Leading Factor 3
NBER	1						
Basis Cycle 1	0.6127	1					
Basis Cycle 2	0.1658	0	1				
Basis Cycle 3	0.0937	0	0	1			
Leading Factor 1	0.6099	0.6630	0	0	1		
Leading Factor 2	0.1458	0	0.5283	0	0	1	
Leading Factor 3	0.0866	0	0	0.4445	0	0	1

Table 4: Simple Probit Regression Results

Regressor	Est. Coeff.	Robust S.E.
c_{1t}	33.24	3.80
c_{2t}	10.53	2.24
c_{3t}	3.65	2.06
Constant	-0.67	0.22

Table 5: Instrumental-Variable Probit Regression Results

Regressor	Est. Coeff.	Robust S.E.
c_{1t}	75.13	10.71
c_{2t}	32.56	7.45
c_{3t}	14.01	8.27
Constant	-0.15	0.23

Table 6: Encompassing Test Results - Simple Probit Index

Null hypothesis Competing Coincident Indicator Models	index_1 encompasses index_2 (p-value)	index_2 encompasses index_1 (p-value)
Issler-Vahid (index_1) vs. TCB (index_2)	0.46	$<$0.01
Issler-Vahid (index_1) vs. Stock-Watson (index_2)	0.40	$<$0.01

Table 7: Encompassing Test Results - Instrumental-Variable Probit Index

Null hypothesis Competing Coincident Indicator Models	index_1 encompasses index_2 (p-value)	index_2 encompasses index_1 (p-value)
Issler-Vahid (index_1) vs. TCB (index_2)	0.83	$<$0.01
Issler-Vahid (index_1) vs. Stock-Watson (index_2)	0.51	0.04

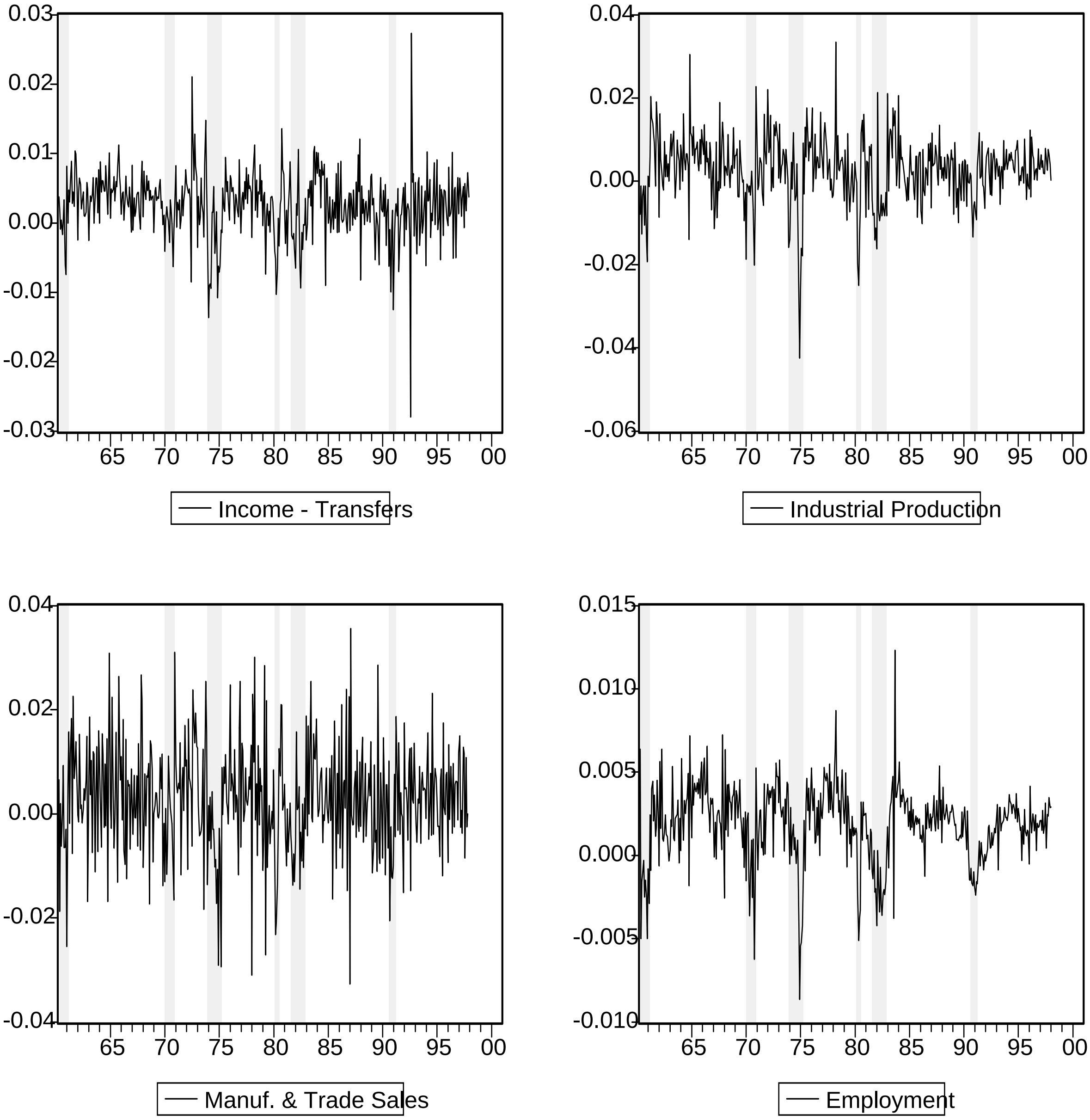

Figure 1: The Coincident Series

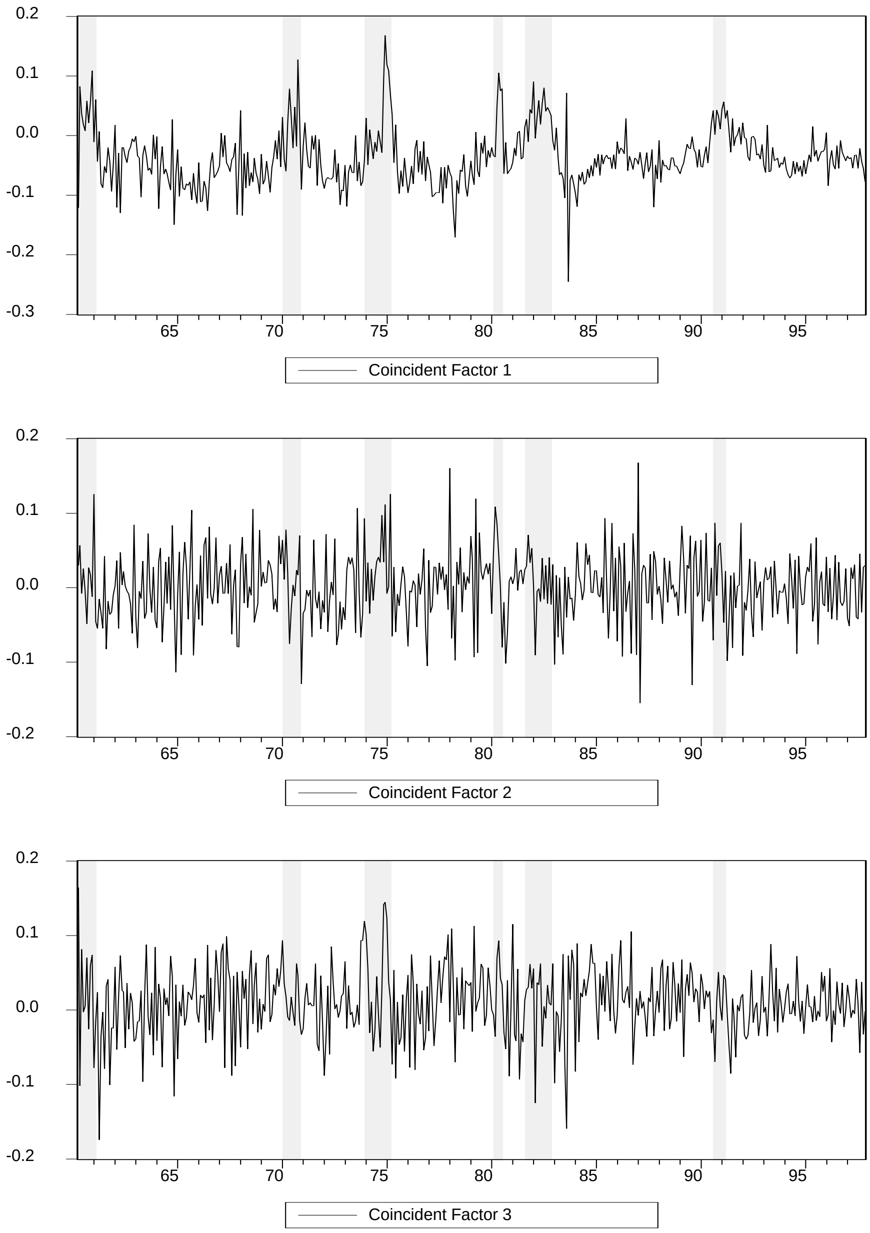

Figure 2: Basis Cycles

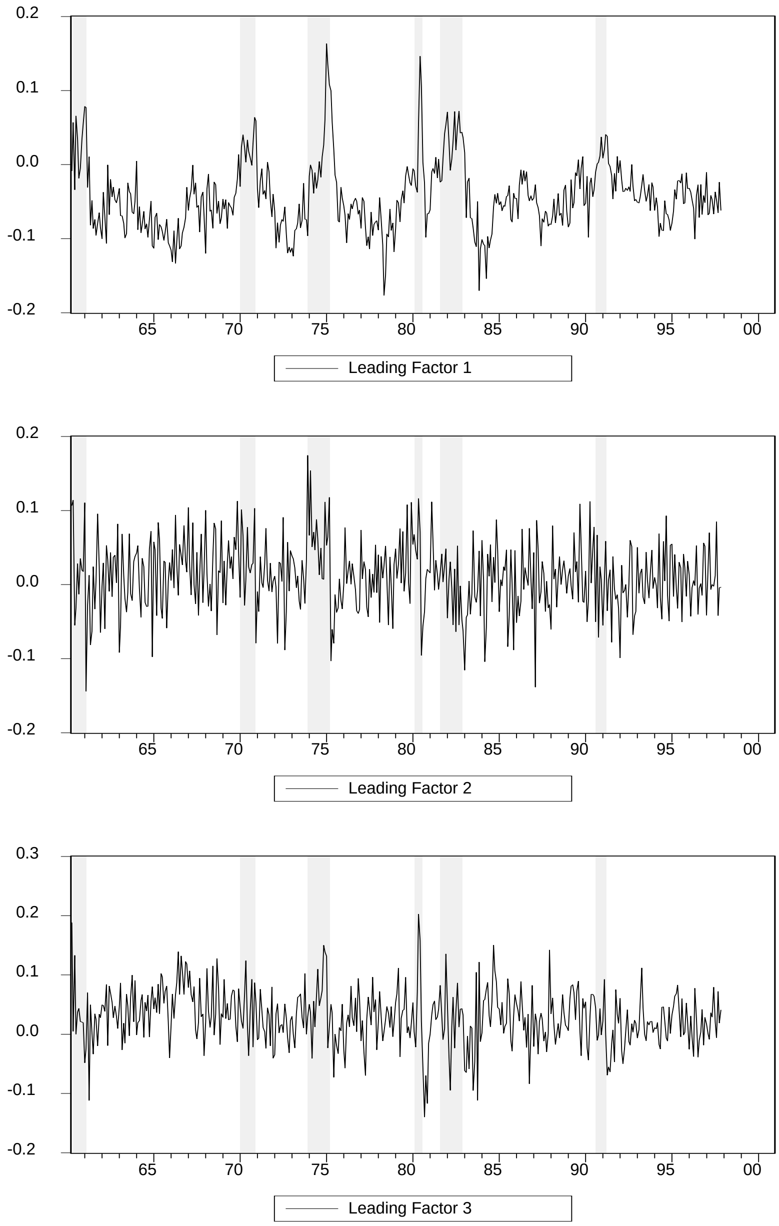

Figure 3: Leading Factors

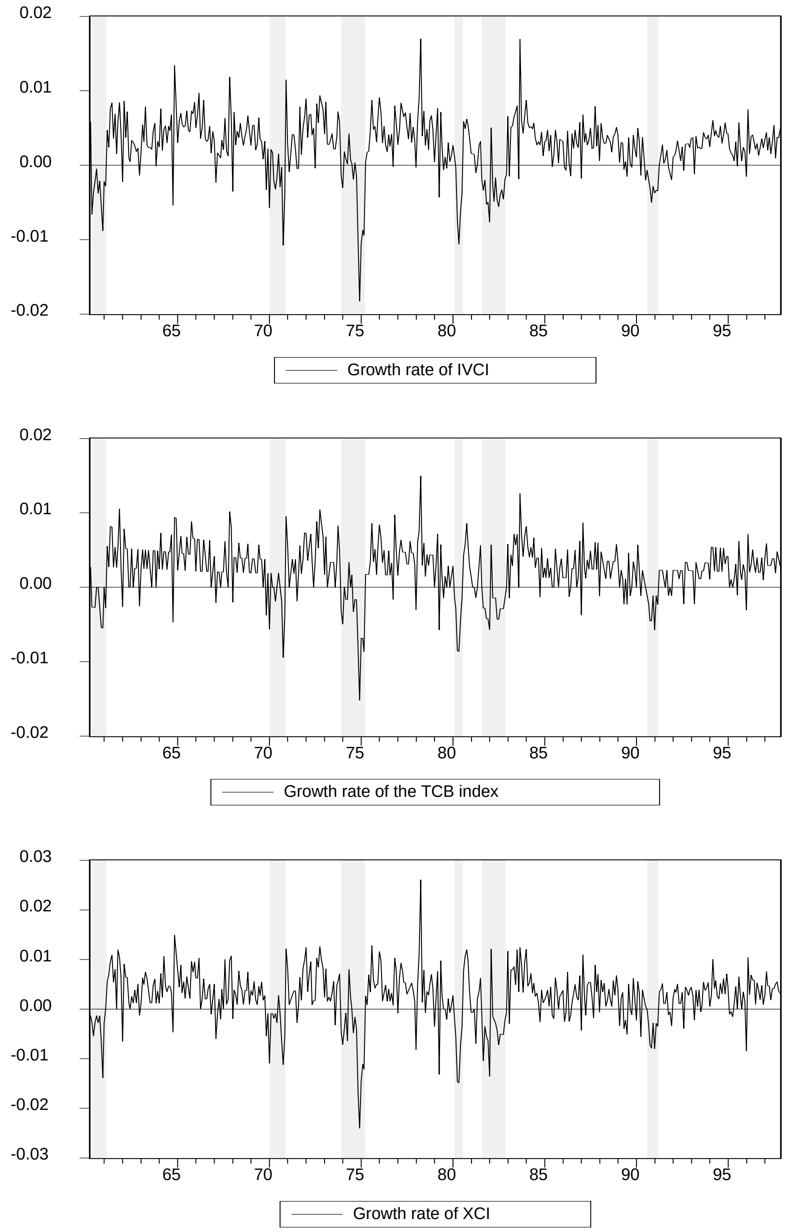

Figure 4: Coincident Indices

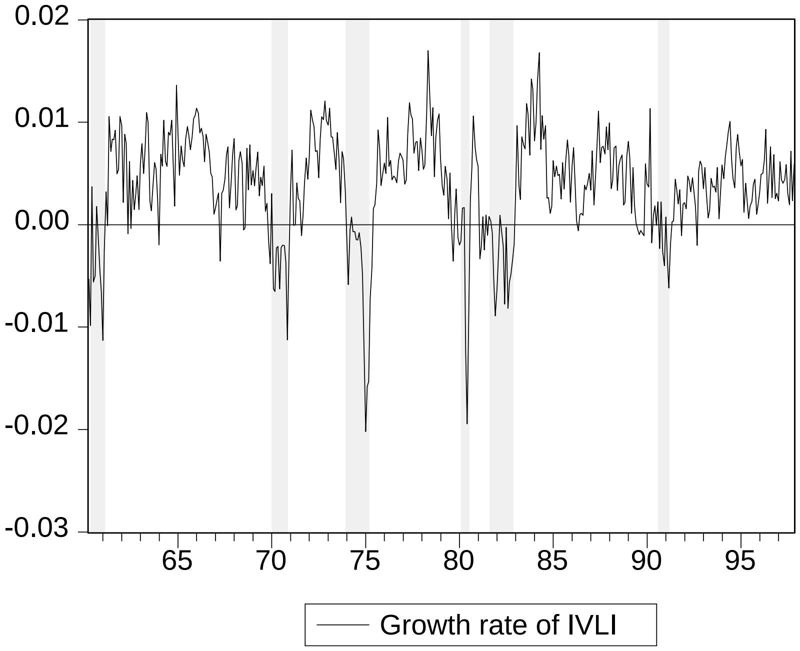

Figure 5: The Leading Index