FUNDAÇÃO
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaios Econômicos

N° 445   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

Maio de 2002

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity[1]

João Victor Issler[2]

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

This Version: May, 2002

[2]Corresponding author.

Abstract

We use the information content in the decisions of the NBER Business Cycle Dating Committee to construct coincident and leading indices of economic activity for the United States. We identify the coincident index by assuming that the coincident variables have a common cycle with the unobserved state of the economy, and that the NBER business cycle dates signify the turning points in the unobserved state. This model allows us to estimate our coincident index as a linear combination of the coincident series. We establish that our index performs better than other currently popular coincident indices of economic activity.

Keywords: Coincident and Leading Indicators, Business Cycle, Canonical Correlation, Instrumental Variable Probit, Encompassing.

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

1   Introduction

Suppose that we are asked to construct an index of health status of a patient. Also, suppose that we know that the best indicator of the health of the patient is the results of a blood test. However, blood samples cannot be taken too frequently, and test results are only available with a lag, sometimes too long to be useful. Our index therefore must be a function of variables such as blood pressure, pulse rate and body temperature that are readily available at regular frequencies. In order to estimate the best way to combine these variables into an index, would we (i) use the historical data on these variables only, or, (ii) use the historical blood test results as well? The answer is, obviously, the latter. This analogy, we hope, illustrates what is missing in the recent attempts to construct new coincident indices of economic activity for the United States. In this literature, researchers have used historical data on coincident series only, and ignored the vital information in the NBER recession indicator.

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 its turning points is embodied in 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 the time of a turning point in the business cycle. 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 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 a structural equation that links the NBER recession indicator to the coincident series. Because we are interested in constructing indices of business-cycle activity, we only use the cyclical parts of the coincident series in this structural equation. This ensures that noise in the coincident series does not affect the final index[2]. We estimate this equation using limited information quasi-maximum likelihood method. Natural candidates for the instrumental variables used in this method are the variables that are traditionally used to construct the leading index. With formal specification tests we establish that data does not reject the assumptions of our model.

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,

---

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

[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.

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 in terms of these same coincident variables 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). 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 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

3

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[3]. 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[4]. 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

---

[3]Although this definition may sound different from the engineering definition of "cyclical", which is a process that is explained by dominant regular periodic functions (such as cosine waves), it is similar to it. Cramér Representation Theorem states that any stationary process can be written as integrals of cosine and sine functions of different frequencies with independent stochastic amplitudes, and as long as the process is not white noise, some of these periodic functions will dominate the rest in explaining the total variation in the process. This justifies using "not white noise" or "predictable from the past" as a definition for "cyclical".

[4]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.

4

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 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 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 has a common cycle with the unobserved business cycle state of the economy.

2.2 Estimating a structural equation for the unobserved business cycle state

One might think that to estimate the weights associated with each basis cycle it suffices 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 NBER indicator by a linear combination of the coincident series. However, it is important to note that the coincident index that we are after is a linear combination of the coincident series that has *cyclical* features similar to the unobserved state of the economy[5]. The NBER recession indicator is important because it embodies some information about the unobserved business cycle state of the economy. As it will be clear below, the linear combination of the coincident series that has a serial correlation pattern similar to that of the unobserved state of the economy, is neither the conditional expectation of the NBER recession indicator given the past information set, nor the conditional expectation of the NBER indicator given the coincident series.

We state the key assumption that enables us to estimate the coincident index here:

Assumption 1: There exists a linear index of (the cyclical parts of) the coincident series that has the exact same correlation pattern with the past information as the unobserved state of the economy.

Note that we have enclosed "the cyclical parts of" in parentheses because it is redundant. Although the index that has the same correlation pattern with the past will only involve the significant basis cycles (i.e. will not involve white noise combinations of the coincident series), these basis cycles are themselves linear combinations of coincident series. Hence, the index will ultimately be a linear combination of coincident series.

Let y_t^* denote the unobserved state of the economy and $\{c_{1t}, c_{2t}, c_{3t}\}$ denote the significant basis cycles of the coincident series at time t. Assumption 1 clearly implies that there must be a linear combination of y_t^* and $\{c_{1t}, c_{2t}, c_{3t}\}$ that is unpredictable from the information before time t. That is,

$$E(y_t^* - \beta_0 - \beta_1 c_{1t} - \beta_2 c_{2t} - \beta_3 c_{3t} \mid I_{t-1}) = 0. \tag{1}$$

where I_{t-1} is the information available at time $t-1$. If y_t^* was observed, we could estimate β_1, β_2 and β_3 directly by GMM or limited information maximum likelihood.

However, y_t^* is not observed. Instead, we have the NBER indicator that is equal to 1 when, to the best knowledge of the NBER Dating Committee at time $t+h$, the economy

---

[5]Using the technical terms introduced in Engle and Kozicki (1993), we are assuming that the unobserved business cycle state of the economy and the coincident variables have a *serial correlation common feature*, and we want to estimate the *cofeature vector* associated with this common feature.

was in a recession at time t. That is, when the "smoothed" estimate of the unobserved state of the economy based on information at time $t+h$ is below a critical value[6].

$$\text{NBER}_t = \begin{cases} 1 & \text{if } E(y_t^* \mid I_{t+h}) < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Using equation (1), we obtain:

$$\begin{aligned} E\left(y_t^* \mid I_{t-1}\right) &= \beta_0 + \beta_1 E\left(c_{1t} \mid I_{t-1}\right) + \beta_2 E\left(c_{2t} \mid I_{t-1}\right) + \beta_3 E\left(c_{3t} \mid I_{t-1}\right) \\ &= \beta_0 + \beta_1 c_{1t} + \beta_2 c_{2t} + \beta_3 c_{3t} + \omega_t, \text{ where } E\left(\omega_t \mid I_{t-1}\right) = 0, \end{aligned}$$

and obviously ω_t is correlated with c_{it}, $i = 1, 2, 3$. Because we can always write

$$E\left(y_t^* \mid I_{t+h}\right) = E\left(y_t^* \mid I_{t-1}\right) + \xi_t + \xi_{t+1} \cdots + \xi_{t+h},$$

where ξ_{t+i} is the "surprise" associated with new information arriving in period $t+i$. It is straightforward to show that:

$$\begin{aligned} E\left(y_t^* \mid I_{t+h}\right) &= \beta_0 + \beta_1 c_{1t} + \beta_2 c_{2t} + \beta_3 c_{3t} + u_t, \\ u_t &= \omega_t + \xi_t + \xi_{t+1} \cdots + \xi_{t+h}, \end{aligned}$$

where u_t is unforecastable given information at time $t-1$, i.e., $E\left(u_t \mid I_{t-1}\right) = 0$, has a "forward" $MA\left(h\right)$ structure, and is correlated with c_{it}, $i = 1, 2, 3$.

In order to estimate β_1, β_2 and β_3 consistently, we need to use an estimation method designed for estimation of a single structural equation with a limited dependent variable. All of such methods use instrumental variables. In our case, obvious instrumental variables would be the z_t variables (i.e. lags of coincident and leading variables). 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) due to its relative simplicity. Because we ignore the dynamic structure of u_t in constructing the likelihood function, i.e., the model is "dynamically incomplete" in the sense of Wooldridge (1994), autocorrelation-robust standard errors have to be used; see the Appendix for details.

Using the empirical results that will be presented fully in the next section, we assume that the four coincident series can be explained by three significant basis cycles

---

[6]This threshold value cannot be identified separately from the constant term in equation (1) from the data. Therefore, without loss of generality, we assume that this critical value is zero (in other words, we let the threshold value be absorbed in the constant term).

$\{c_{1t}, c_{2t}, c_{3t}\}$. 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[7]:

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

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:

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

which shows that it is a linear combination of the coincident series x_t. Similarly, if we replace c_{1t}, c_{2t}, c_{3t} with their best predictors $\gamma_1' z_t, \gamma_2' z_t, \gamma_3' z_t$ in the above formula, we obtain our leading index that is a linear combination of the leading series z_t.

In summary, our complete statistical model is the following:

$$
\begin{aligned}
\text{NBER}_t &= \begin{cases} 1 & \text{if } E(y_t^* \mid I_{t+h}) < 0 \\ 0 & \text{otherwise.} \end{cases} \\
E(y_t^* \mid I_{t+h}) &= \psi_0 + \psi' \underset{4\times 1}{x_t} + u_t \\
x_t &= \underset{4\times m}{\Pi} \underset{m\times 1}{z_t} + \varepsilon_t,
\end{aligned}
\tag{3}
$$

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

2.3 Directed specification tests for our coincident index

In our econometric model in (3), we have assumed that y_t^* and x_t have the same correlation pattern with z_t (which implies that z_t can be used as instruments for x_t, or that u_t and z_t are uncorrelated), that the errors are jointly normal, and that Π has less than full rank, specifically rank 3. There are also other assumptions about the choice of variables in x_t and z_t. After obtaining our coincident index, it is possible to test these underlying assumptions. However, we only test our model against specific directions. The reason is that we are putting forward an econometric model that we claim to be more appropriate than the existing models which lead to a coincident index constructed from the same four coincident variables. Therefore, as an alternative to the specification in (3), we do not consider other variables in x_t or z_t because that will not fit within the

---

[7]The negative sign, which is a slight difference from the textbook presentation of probit models, is a result of our assuming that the binary variable NBER_t is equal to 1 when y_t^* is *less than* zero.

objectives of this paper. The specific direction that we test our model against is implied in the following question. Given our coincident index, is there any information in alternative coincident indices based on the same coincident variables that helps explain the business cycle state of the economy? Natural candidates of alternative indices are TCB's (Dept. of Commerce) and Stock and Watson experimental 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 the result of the first comprehensive research project on constructing a coincident index based on a statistical model; see Appendix A.1 for more details on both indices.

Let $index_{1t}$ denote our index and $index_{2t}$ denote one of the two alternative indices. Our specification tests is based on a test of significance of the coefficient of $index_{2t}$ in the linear probability model

$$\text{NBER}_t = \theta_0 + \theta_1 index_{1t} + \theta_2 index_{2t} + e_t, \tag{4}$$

where the error term is allowed to be correlated with the right hand side variables and z_t are used as instrumental variables. We use a linear probability model rather than a probit model to make the test free of particular type of distributional assumptions[8]. Since the linearity assumption in equation (4) is too simplistic, we add higher powers of $index_{1t}$ and $index_{2t}$ to the right-hand side of equation (4),

$$\text{NBER}_t = \theta_0 + \theta_1 index_{1t} + \theta_1' index_{1t}^2 + \theta_1'' index_{1t}^3 + \theta_2 index_{2t} + \theta_2' index_{2t}^2 + \theta_2'' index_{2t}^3 + e_t. \tag{5}$$

Again, the right hand side variables are allowed to be correlated with the errors and z_t are used as instrumental variables. The specification test for our model is a test of the null hypothesis of $\theta_2 = \theta_2' = \theta_2'' = 0$ in equation (5). Of course, linear probability models are heteroskedastic, and, for reasons explained in the previous section, the errors may also be serially correlated. Therefore, we use a robust estimate of the covariance matrix to do hypothesis testing.

If the alternative coincident indices were constructed on the basis of the same information set as our index, the above specification tests could be interpreted as tests of our index *encompassing* the alternative indices (see Mizon 1984). However, since the NBER recession indicator is not used in the construction of either of the two alternative indices, it would be technically incorrect to conclude that our index encompasses those alternatives when we fail to reject $\theta_2 = \theta_2' = \theta_2'' = 0$ in equation (5). What we can conclude when we fail to reject $\theta_2 = \theta_2' = \theta_2'' = 0$ is that no linear combination of our index with the alternative indices provides a proxy for the unobserved business cycle state of the economy that is significantly superior to our index.

---

[8] Alternatively, one can take the probit specification as the correct specification under the null, and design the test along the lines of the so-called "artificial regression" approach described in Davidson and McKinnon (1993, pp. 523-528).

An alternative test for coincident indices is to measure how useful each of them is in describing the actual peaks and troughs of U.S. economic activity. Because all three indices aim at describing the current state of the economy, we can verify their relative success in this task by measuring the peaks and troughs of economic activity implied by their time-series behavior, comparing the results with what we observe in terms of U.S. peaks and troughs. The NBER Dating Committee's decisions are used to determine the latter[9].

There are several ways of determining peaks and troughs in a given time series. By far, the most used procedure in business-cycle analysis is the Bry and Boschan (1971) algorithm. Recently, there has been a revival of interest in this algorithm for business cycle research (see Watson, 1994, and Harding and Pagan, 2002). In the second test performed here, for all three indices (IVCI, TCB and XCI), we extract their respective peaks and troughs using the Bry and Boschan algorithm[10]. Time periods between peaks and troughs are labelled as recessions and assigned a value of one. Otherwise, periods are assigned a value of zero. For a group of coincident indices these dummy variables can be compared with the actual NBER Dating Committee's dummy using a quadratic loss function. Loss function results may be interpreted as a measurement of the percentage of time periods that a given index misclassifies the state of the economy.

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

---

[9]We thank Mark Watson for suggesting this exercise to us. Indeed, he suggested as well that we could go a step further, choosing the weights of coincident series that would best fit the NBER's committee peaks and troughs. We leave that for future research.

[10]We used the Gauss code downloadable from Mark Watson's Home-Page.

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 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[11].

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:

$$
\begin{bmatrix} c_{1t} \\ c_{2t} \\ c_{3t} \end{bmatrix} = \begin{bmatrix} 1.03 & 0.31 & 19.44 & -0.68 \\ -1.68 & 1.12 & 1.12 & 4.64 \\ -0.27 & 7.78 & -13.46 & -2.33 \end{bmatrix} \times \begin{bmatrix} \Delta \ln I_t \\ \Delta \ln Y_t \\ \Delta \ln N_t \\ \Delta \ln S_t \end{bmatrix} \quad (6)
$$

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": using the NBER information in computing the coincident index

The structural model in (3) enables us to incorporate the information in the NBER recession indicator in constructing a coincident index of economic activity. It also incorporates the information resulted from the canonical correlation analysis, namely that there are only three significant basis cycles in the four coincident series. We use the two stage conditional maximum likelihood (2SCML) estimator proposed by Rivers and

---

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

Vuong (1988) to obtain instrumental variable estimates for the coefficients of each basis cycle.

The 2SCML estimates are presented in Table 4. After rewriting the basis cycles as linear combinations of the coincident series and normalizing the weights to add up to unity, we obtain our index, which we call the instrumental-variable coincident index ($IVCI_t$):

$$\Delta IVCI_t = 0.02 \times \Delta \ln I_t + 0.13 \times \Delta \ln Y_t + 0.80 \times \Delta \ln N_t + 0.05 \times \Delta \ln S_t. \quad (7)$$

Equation (7) shows that most of the weight is given to employment, and that employment and industrial production together get 93% of the weight. A plot of this index is presented in Figure 4. This is not surprising given our previous analysis of Figure 1, since these two series have a more pronounced coherence with the NBER recession indicator. It also agrees with the latest memo of the Business Cycle Dating Committee (Hall et al. 2002, p. 9) where they state "employment is probably the single most reliable indicator [of recessions]". It is interesting to compare our index with alternative indices in the literature. The corresponding weights that are used by the Conference Board to calculate the coincident index[12] are $(0.28, 0.13, 0.48, 0.11)$. The striking difference between our weights and those of the TCB index is that income (I_t) is weighed much more heavily in the TCB index than in ours, and employment (N_t) is weighed more heavily in our index than in theirs.

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. It must be emphasized that this is only a one step-ahead leading index. In order to create several step-ahead leading indices, our model must be enlarged to include forecasting equations for the leading variables in z_t. Since best forecasting equations for some of the leading variables (such as interest rates) are nonlinear, we believe that proper several step-ahead leading indices cannot be linear. However, the construction of such indices is beyond the scope of the present paper.

3.3 Comparisons with Existing Coincident Indices

We perform the specification tests described in Section 2.3 regarding the TCB index and the experimental index[13] (XCI) proposed by Stock and Watson (1989). The results of estimating equations (4) and (5), once with the TCB index and once with the XCI index as the alternative index, are presented in Table 5. This table also shows the p-values

---

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

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

of the null hypothesis of "given our index, the alternative index is insignificant" in each equation. It can be seen from this table that the coefficients of our index are significant, whereas there is no evidence that the coefficients of the other two indices are significantly different from zero. The results of these tests clearly indicate that, controlling for our index, there is no useful information in either the TCB or the XCI indices in explaining the business cycle state of the economy.

We have labelled this a specification test since our coincident index models directly the state of the economy, as decided by the NBER Business Cycle Dating Committee, whereas the other two indices do not. To implement a more "neutral" test for these three coincident indices, we decided to measure how useful each of them is in describing the actual peaks and troughs of U.S. economic activity. As discussed at the end of Section 2.3, we verify their relative success in this task by measuring the peaks and troughs of economic activity implied by their time-series behavior, comparing the results with peaks and troughs implied by the NBER Dating Committee's decisions.

We use the Bry and Boschan (1971) algorithm to extract peaks and troughs of all three indices, and confront the results with the actual NBER Dating Committee's decisions using a quadratic loss function. The results show that our index, IVCI, has the smallest loss function overall: 0.028 versus 0.032 for the next index, and 0.041 for the last one. Hence, IVCI has a 2.8% chance of erroneously classifying the state of the economy in sample, which is better than the performance of other existing coincident indices.

The reason for the performance of IVCI can be analyzed by looking closely at specific boom and recession episodes. Our index makes its largest error in missing the onset of the 1973 recession from 12/1973 through 7/1974. However, IVCI outperforms the TCB and XCI indices in several occasions where these two indices give a false alarm of a recession. The single most important of such episodes occurred during the period 8/1980 through 7/1981. It is worth noticing that XCI produces the largest number of false alarms of a recession among the three indices. A possible explanation for the favorable behavior of IVCI is the fact that our index places a larger weight on employment relative to the other two. Because there are costs in hiring and firing employees, employment may not move immediately as soon as a recession starts, but, when it does, a recession has most probably already started. Because of that, and contrary to the TCB and the XCI indices, IVCI rarely indicates that a recession is occurring when it is not. Indeed, for the whole sample, IVCI only gave one false alarm of a recession in July, 1990. The fact that our index does not produce false alarms as often as the TCB and XCI indices do, makes its overall performance superior to that of the other two.

13

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. The 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. In our model, a structural equation relates the unobserved state of the economy to the cyclical components of the coincident series. We use a two stage conditional maximum likelihood method to use the information in the NBER recession indicator about the unobserved state of the economy in order to estimate the parameters of this structural equation. 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 index: the coincident index is a linear combination of the four coincident series proposed by Burns and Mitchell that has a common cycle with an unobserved state variable which is consistent with the deliberations of the NBER Business Cycle Dating Committee. It is noteworthy that our coincident index places the largest weight on employment (80%), which is in agreement with the latest opinion of the NBER Business Cycle Dating Committee (Hall et al. 2002, p. 9) that "employment is probably the single most reliable indicator [of recessions]".

Our methodology also conveniently produces a one-step 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. With specification tests against particular alternatives, we conclude that there is no gain in combining our index with either of the two currently popular coincident indices, namely the TCB and the XCI coincident indices. This means that given our index, there is no useful information in the other two indices about the state of the economy. Although technically we cannot conclude that our index encompasses the TCB and XCI indices, because those two indices do not use the information in the NBER dates in their construction, the specification test results delineate the important question that motivated our paper, i.e. why do TCB and XCI indices ignore this vital piece of information in their

construction?

In countries where there are no institutions similar to the NBER Business Cycle Dating Committee, simple rules such as two quarters of negative growth in the GDP or the quarterly version of Bry and Boschan (1971) algorithm applied to the quarterly GDP are used to identify recessions. A useful extension of the present paper will be to use our structural framework to identify the coincident index as the common cycle between the monthly coincident variables and the quarterly recession indicator or the quarterly GDP series. This extension is left for future research.

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.

Bry, G. and C. Boschan (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, New York: NBER.

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.

Hall, R., Feldstein, M., Bernanke, B., Frankel, J., Gordon, R. and Zarnowitz, V. (2002), "The NBER's Business-Cycle Dating Procedure", downloadable from http://www.nber.org/cycles/recessions.pdf.

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.

Harding, D. and A.R. Pagan (2002), "Dissecting the Cycle: A Methodological Investigation", *Journal of Monetary Economics*, 49, 365-381.

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.

Mizon, G.E. (1984), "The Encompassing Approach in Econometrics", Chapter 6 in D.F. Hendry and K.F. Wallis (eds) *Econometrics and Quantitative Economics*, Oxford: Blackwell.

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.

Watson, M.W. (1994), "Business-Cycle Durations and Postwar Stabilization of the U.S. Economy," *American Economic Review*, 84, 24-46.

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:

$$
\begin{aligned}
x_t &= \beta + \gamma(L)\Delta XCI_t + u_t, \\
\phi(L)\Delta XCI_t &= \delta + \eta_t, \\
D(L)u_t &= \epsilon_t,
\end{aligned}
\tag{8}
$$

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 $E\left[\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}'\right] = \mathrm{diag}(\sigma_\eta^2, \sigma_{\epsilon_1}^2, ..., \sigma_{\epsilon_4}^2)$, and $D(L) = \mathrm{diag}[d_{ii}(L)]$, which makes innovations mutually uncorrelated.

The model (8) 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 (8), 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:

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

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 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 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 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:

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + B_1 w_{t-1} + \cdots + B_p w_{t-p} + \varepsilon_t, \tag{9}$$

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 $z_t = (x'_{t-1}, \cdots, x'_{t-p}, w'_{t-1}, \cdots, w'_{t-p})'$. 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 Two stage conditional maximum likelihood estimation

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^*:

$$E(y_t^* \mid I_{t+h}) = \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + u_t \tag{10}$$

$$\text{NBER}_t = \begin{cases} 1 & \text{if } E(y_t^* \mid I_{t+h}) < 0 \\ 0 & \text{otherwise.} \end{cases}.$$

The possible correlation between $c_{1t}, \cdots, c_{kt}$ and the errors u_t is modelled as follows,

$$c_{it} = \lambda_i \left(\gamma_i' z_t \right) + v_{it}, \quad i = 1, \cdots, k \tag{11}$$

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim N \left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{vu}' \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix} \right)$$

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 (10), we obtain,

$$E(y_t^* \mid I_{t+h}) = \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta + \eta_t \tag{12}$$

$$\text{NBER}_t = \begin{cases} 1 & \text{if } \eta_t < -(\beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta) \\ 0 & \text{if } \eta_t \geq -(\beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta) \end{cases}.$$

Notice that, by construction, all the regressors in (12) 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 (12) 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:

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

where the appropriate formula for the asymptotic covariance matrix V is given in Rivers and Vuong(1988, p. 354).

The error term η_t in (12), as explained in the text, 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).

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: Two Stage Conditional Maximum Likelihood Estimates

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

Table 5: Specification Test Results

Dependent variable is NBER	index$_1$ is ΔIVCI index$_2$ is $\Delta \log(\text{TCB})$		index$_1$ is ΔIVCI index$_2$ is $\Delta \log(\text{XCI})$	
Regressor	$Coeff.$ (Std. Err.)	$Coeff.$ (Std. Err.)	$Coeff.$ (Std. Err.)	$Coeff.$ (Std. Err.)
index$_1$	-0.110 (0.047)	-0.130 (0.086)	-0.075 (0.020)	-0.112 (0.044)
index$_1^2$	-	-0.002 (0.007)	-	-0.003 (0.006)
index$_1^3$	-	0.001 (0.001)	-	0.001 (0.001)
index$_2$	0.242 (0.555)	-0.023 (1.016)	-0.119 (0.163)	-0.091 (0.348)
index$_2^2$	-	1.103 (1.115)	-	0.507 (0.355)
index$_2^3$	-	-1.053 (1.861)	-	-0.191 (0.231)
Constant	0.377 (0.044)	0.305 (0.049)	0.372 (0.042)	0.284 (0.054)
P-value for "index$_2$ is insignificant"	0.663	0.787	0.468	0.467

In both tables, all equations are estimated using 2 lags of coincident and leading variables as instruments. The standard errors and p-values are calculated using the Newey-West estimator of the covariance matrix.
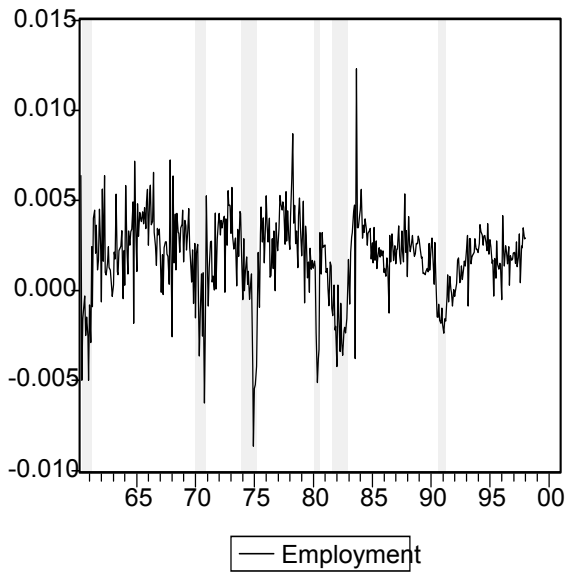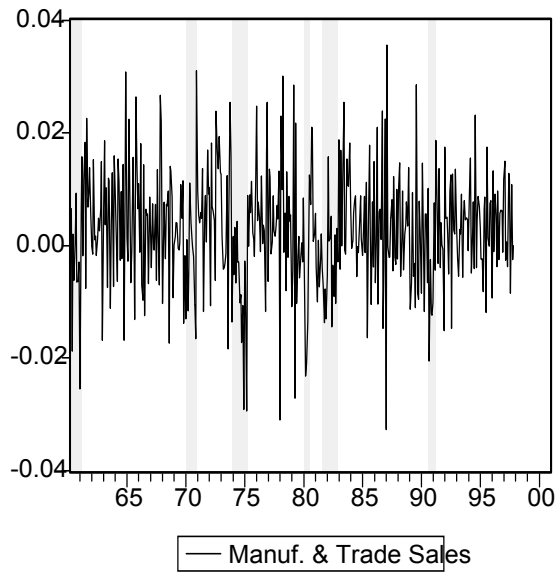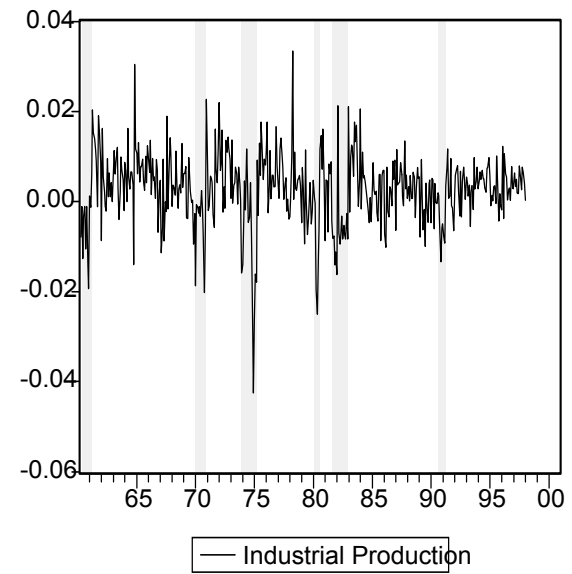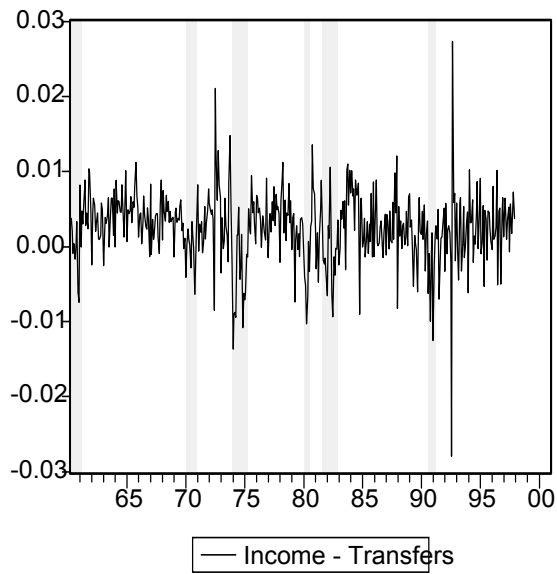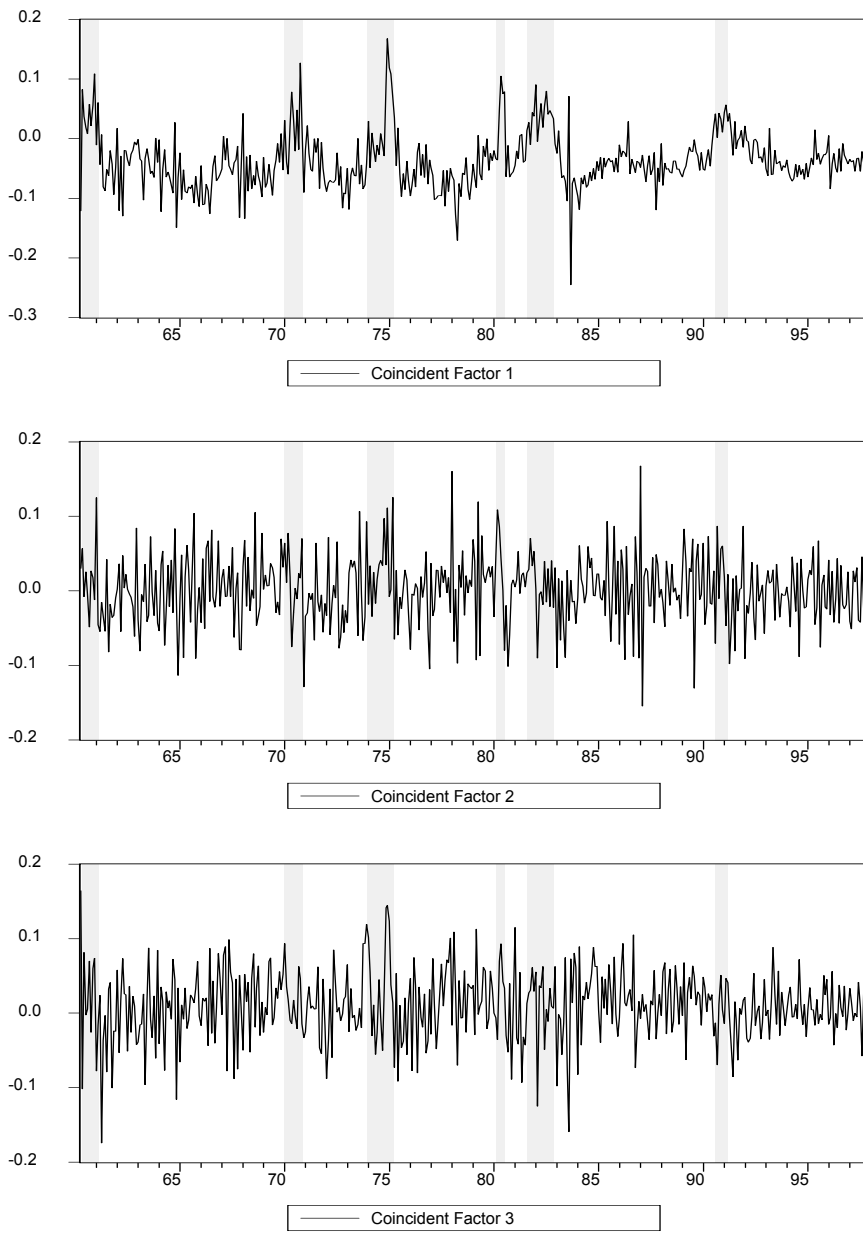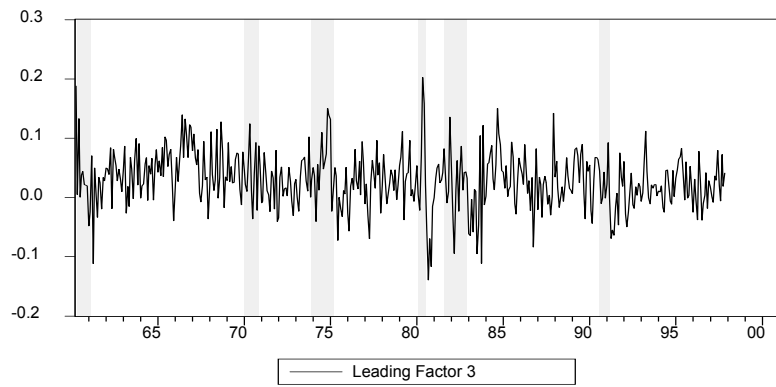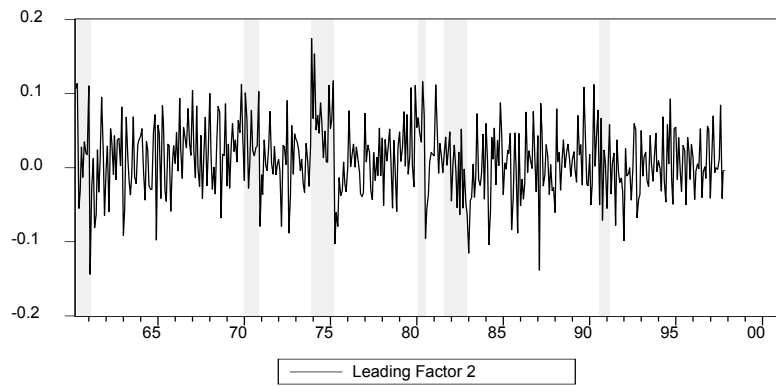
Figure 1: The Coincident Series
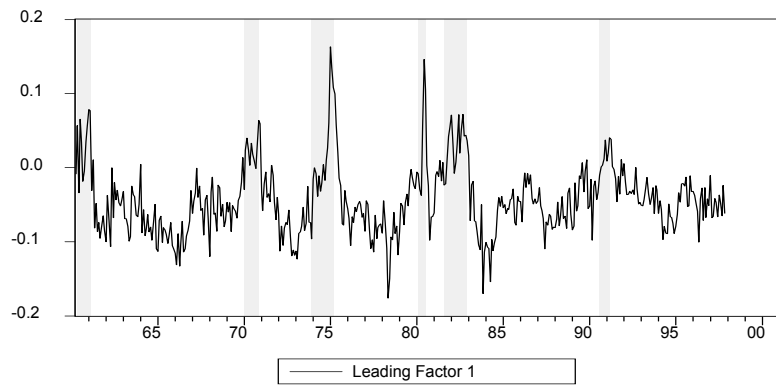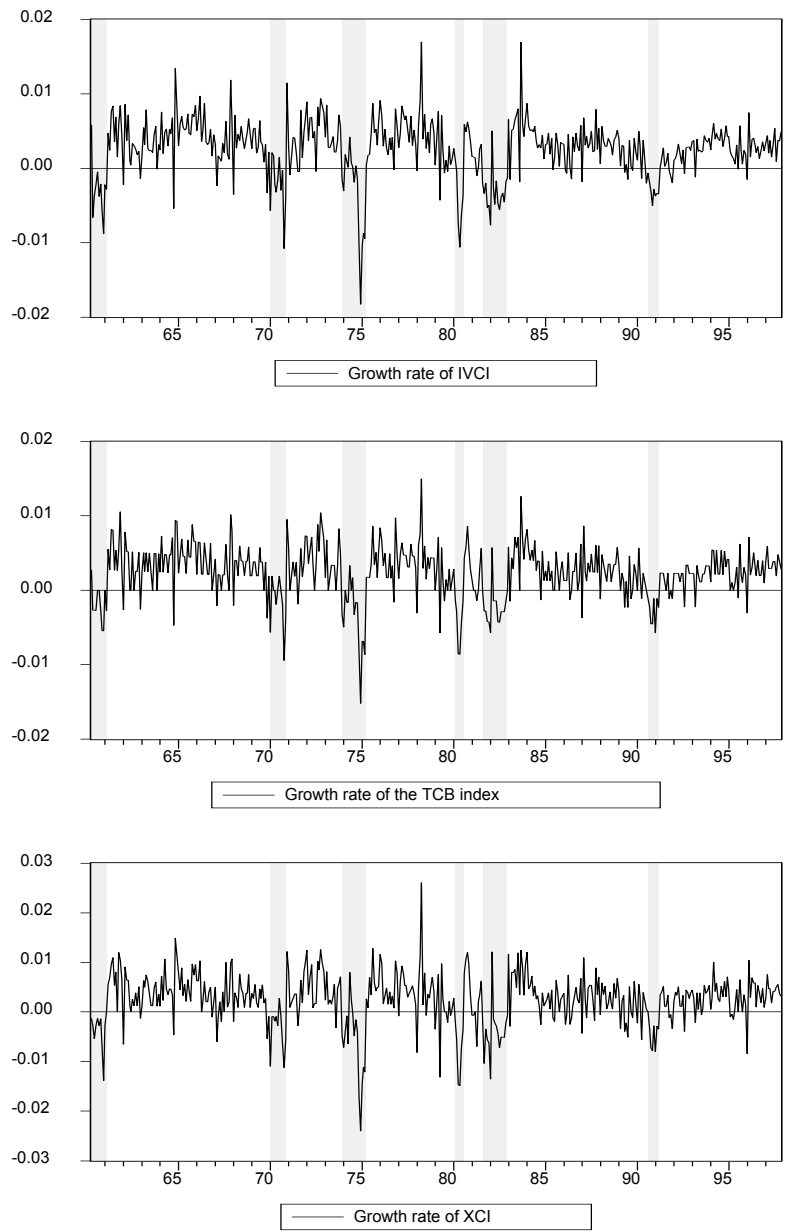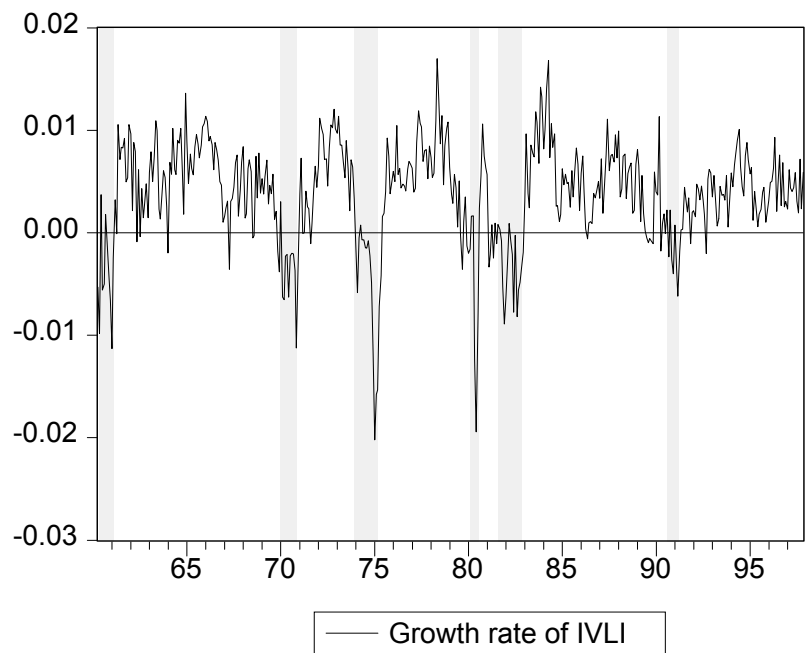
25

Figure 2: Basis Cycles

Figure 3: Leading Factors

Figure 4: Coincident Indices

28

Figure 5: The Leading Index