FUNDAÇÃO
GETULIO VARGAS

EPGE
Escola de Pós-Graduação
em Economia

Ensaios Econômicos

Escola de
Pós-Graduação
em Economia
da Fundação
Getulio Vargas

N° 492 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

Agosto de 2003

URL: http://hdl.handle.net/10438/544

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; 492)

Inclui bibliografia.

CDD-330

**Nº 492** ISSN 0104-8910

## *The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity*

***JoãoVictor Issler***
***Farshid Vahid***

*Agosto de 2003*

# 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

First Version: May 2002
Revised: July 2003

[1] **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, Clive Granger, Don Harding, Alain Hecq, Soren Johansen, Carlos Martins-Filho, Helio Migon, Ajax Moreira, Adrian Pagan, Mark Watson, Arnold Zellner and two anonymous referees, 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.

[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 compare the performance of our index with 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 the 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 inferences 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-cycle[1], to construct a coincident 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 model, hoping that its factor 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 the 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 the 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) and the collections of papers in Lahiri and Moore (1993) and 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.

Other related studies in the business cycle literature are Birchenhall et al (1999), who propose a logistic rule to classify the states of the economy, Zellner and Min (1999), who emphasize the role of leading indicators in the prediction of turning points, and Watson (1994), who examines basic business cycle statistics of the pre and post war US.

The structure of the rest of the paper is as follows. In Section 2 we present the basic ingredients of our methodology. Section 3 presents our empirical results 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[3]. This 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.

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-correlation 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 combina-

[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's 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-correlation analysis studies channels of *linear* dependence between $x$ and $z$ does not necessarily imply that it will only be 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

tions 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 combinations 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 $R^2$s 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 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. The likelihood ratio test statistic for the null hypothesis that there are $k$ significant cycles (i.e., there are $4 - k$ zero canonical correlations) is:

$$LR = -T \sum_{i=k+1}^{4} \ln(1 - \lambda_i^2)$$

which has an asymptotic $\chi^2$ distribution with $(4 - k)(m - k)$ degrees of freedom (see Anderson, 1984). It is customary to use $(T - m)$ instead of $T$ in the above statistic to improve its finite sample performance. If the null is not rejected, then the linear combinations corresponding to the statistically insignificant canonical correlations cannot be predicted from the past 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 combinations 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.

and further references.

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 is 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 will become 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 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.

[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.

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 $\beta_1, \beta_2$ and $\beta_3$ 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 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 $\omega_t$ 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 $\xi_{t+i}$ 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 $\beta_1, \beta_2$ and $\beta_3$ consistently, we need to use an estimation method designed for estimation of a single structural equation with a limited dependent variable. All 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.

[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).

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.

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 $\{c_{1t}, c_{2t}, c_{3t}\}$. Denoting the NBER indicator by $\text{NBER}_t$, the first stage of the 2SCML estimation procedure 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]:

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

where $\Phi$ is the standard normal cumulative distribution function. The estimates of $\beta_1$, $\beta_2$, and $\beta_3$ from the second stage probit will be the 2SCML estimates. The standard errors of the estimated parameters have to be modified according to the procedure in Rivers and Vuong (1988, page 354). In addition, 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.

Our coincident index, which we label the "instrumental variable coincident index" ($IVCI$), is then given by:

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

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 $\lambda_1 \gamma_1' z_t$, $\lambda_2 \gamma_2' z_t$, $\lambda_3 \gamma_3' z_t$ in the above formula, we obtain our leading index as a linear combination of the leading series $z_t$, that is:

$$\Delta IVLI_t = E_{t-1}\left(\widehat{\beta_1} c_{1t} + \widehat{\beta_2} c_{2t} + \widehat{\beta_3} c_{3t}\right) = \left(\widehat{\beta_1} \lambda_1 \gamma_1' + \widehat{\beta_2} \lambda_2 \gamma_2' + \widehat{\beta_3} \lambda_3 \gamma_3'\right) z_t. \quad (3)$$

In summary, our complete statistical model is the following:

[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 $\text{NBER}_t$ is equal to 1 when $y_t^*$ is *less than* zero.

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

where $u_t$ may be correlated with $\varepsilon_t$, $u_t$ and $\varepsilon_t$ are jointly normal, and $\Pi$ has rank 3.

### 2.3 Directed specification tests for our coincident index

In our econometric model in (4), 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 $\Pi$ 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 (4), we do not consider other variables in $x_t$ or $z_t$ because that will not fit within the 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 for alternative indices are the TCB (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 test is based on a test of significance of the coefficient of $index_{2t}$ in the linear probability model

$$\text{NBER}_t = \theta_0 + \theta_1 index_{1t} + \theta_2 index_{2t} + e_t, \tag{5}$$

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 types of distributional assumptions[8].

[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). We report the results of these tests also in Section 3.

Since the linearity assumption in equation (5) is too simplistic, we add higher powers of $index_{1t}$ and $index_{2t}$ to the right-hand side of equation (5),

$$\text{NBER}_t = \theta_0+\theta_1 index_{1t}+\theta_1' index_{1t}^2+\theta_1'' index_{1t}^3+\theta_2 index_{2t}+\theta_2' index_{2t}^2+\theta_2'' index_{2t}^3+e_t. \quad (6)$$

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 (6). 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 (6). 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.

# 3 Calculating coincident- and leading-indicator indices

## 3.1 Identification of the basis cycles

Our analysis is based on monthly data from 1960:01 to 1997:11. The four coincident series, "Income" ($I_t$), "Industrial Production" ($Y_t$), "Employment" ($N_t$) and "Sales" ($S_t$), are defined in Table 1 and their growth rates 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 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[9].

The Akaike Information Criterion chose a VAR of order 2. However, the LM test on the residuals of a VAR(2) showed signs of significant serial correlation in the errors. Since the canonical-correlation test is valid only if the model is dynamically well specified[10], we increased the lag length until there were no signs of serial correlation left in the residuals of the system. This led to a VAR of order 4. Conditional on a $VAR(4)$ 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 four lags of ($\Delta \ln I_t$, $\Delta \ln Y_t$, $\Delta \ln N_t$, $\Delta \ln S_t$) and four 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:

$$\begin{bmatrix} c_{1t} \\ c_{2t} \\ c_{3t} \end{bmatrix} = \begin{bmatrix} 0.45 & -0.05 & 20.90 & -0.52 \\ 1.43 & -0.69 & 6.72 & -4.78 \\ -0.87 & -7.82 & 14.56 & 2.13 \end{bmatrix} \times \begin{bmatrix} \Delta \ln I_t \\ \Delta \ln Y_t \\ \Delta \ln N_t \\ \Delta \ln S_t \end{bmatrix} \tag{7}$$

A correlation matrix of the three coincident and the three 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 (4) enables us to incorporate the information in the NBER recession indicator in constructing a coincident index of economic activity. It also incorporates the information resulting 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 Vuong (1988) to obtain instrumental variable estimates for the coefficients of each basis cycle.

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

[10] We thank the referee for reminding us of the sensitivity of canonical correlation test to serial correlation.

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 (standard errors are given in parentheses):

$$\Delta IVCI_t = \underset{(0.01)}{0.00} \times \Delta \ln I_t + \underset{(0.06)}{0.10} \times \Delta \ln Y_t + \underset{(0.06)}{0.84} \times \Delta \ln N_t + \underset{(0.02)}{0.06} \times \Delta \ln S_t. \quad (8)$$

Equation (8) shows that most of the weight is given to employment, no weight is given to income, and that employment and industrial production together get 94% of the weight. A plot of this index is presented in Figure 2. 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 a 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[11] 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.

In order to determine the sensitivity of the weights to the choice of three basis cycles, we have also computed our index using one, two and four basis cycles. When we use only one factor, the index is almost entirely based on the employment variable (weight of employment is almost 100%). As we increase the number of basis cycles, the weight of employment decreases to 0.91, 0.84 and 0.71 in the two, three and four factor models respectively. A four factor model is the one where all linear combinations of variables, even those which have no cyclical information, are included in the construction of the index. This shows that a large part of the difference between our index and the other two indices is our decision to exclude the white noise linear combination of the series from the analysis.

Another aspect of sensitivity analysis concerns the stability of the structure over time. This includes the number of significant basis cycles found and the coefficients of these basis cycles in the probit analysis. Starting from the first half of the sample and recursively estimating our model adding one observation at a time, we found almost invariably three significant basis cycles at the 5% level of significance. This reinforces our choice of three basis cycles in the construction of our index. The recursively estimated weights in equation (8) however show some variation over time. In particular, the weight of income has declined from 0.25 to 0.00, which is the only case where recursive estimates leave the 95% confidence band of the initial estimates. This is consistent with the observation that income has not fallen substantially in recent recessions and hence

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

the NBER has reduced its weight in dating these recessions (see Hall et al, 2002). We leave the question of whether it would be useful to relax our assumption of a fixed weight coincident indicator for future research.

As a by-product of our analysis, we construct $\Delta IVLI_t$ as in (3). We investigate its predictive performance by running a probit regression where $\Delta IVLI_t$ is used to forecast the state of the economy. If we use a cuttoff of 0.5, the model correctly predicts 97% of expansion periods and 62% of recession periods.Despite its good performance in identifying the state of the economy, it must be emphasized that $\Delta IVLI_t$ is only a one-step-ahead leading index, and that constructing multi-step-ahead leading indices are beyond the scope of this paper, since it involves dealing with non-linearity and causality.

### 3.3 Comparisons with Existing Coincident Indices

We perform the specification tests described in Section 2.3 with regard to the TCB index and the experimental index[12] (XCI) proposed by Stock and Watson (1989). The results of estimating equations (5) and (6), 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 of the null hypothesis of "given one index, the alternative index is insignificant" in each equation. It can be seen from this table that in a linear probability model, the coefficient of our index is significant, whereas there is no evidence that the coefficients of the other two indices are significantly different from zero. When we add the squares and cubes of indices to test equations, there is no evidence that, given our index, XCI and TCB indices have any additional explanatory power. However, when we perform the reverse test, we conclude that given XCI, our index has significant explanatory power, but that given the TCB index it doesn't. These tests 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[13].

Our coincident index is designed to provide the best prediction of the state of the economy at time $t$ given the information available at time $t$. Our probit analysis produces estimates of recession probabilities at each time period, conditional on the information available at that time[14]. Figure 3 plots these predicted probabilities. Based on a cutoff point of 0.5, the model predicts recessions and expansions with accuracy of 77% and 98% respectively. A naive comparison of the correct predictions of three probit regressions with IVCI, TCB and the XCI indices as explanatory variables shows that IVCI is more accurate in predicting recessions (46 correct predictions versus 37 for TCB and 31 for XCI), while all three indices perform equally well in predicting expansions. Note that

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

[13] We also used the appropriately modified "artificial regression" test of Davidson and McKinnon (1993), and the results were qualitatively identical at the 5% significance level.

[14] These probabilities are similar to *filtered* probabilities of a Markov switching model.

this exercise uses only current information to predict the state of the economy.

This picture changes when we allow for "smoothing" using future information to detect turning points. We use the Bry and Boschan (1971) algorithm (a two-sided filter) to extract peaks and troughs (i.e. turning points) of the three indices. We observe that, for turning points alone, the TCB index gets 9 out of the 12 turning points correctly, while IVCI gets 7 and XCI gets 6. For each index, time periods between peaks and troughs are labelled as recessions and assigned a value of one. Otherwise, periods are assigned a value of zero. These dummy variables are compared with the actual NBER Dating Committee's dummy using a quadratic loss function. For the sample as a whole, TCB has a mean squared error of 0.9%, whereas IVCI and XCI have mean squared errors of 2.8% and 4.1% respectively. It is important to reiterate that our goal is to design an index that produces the best prediction of the state of economy based on the information available at time $t$, and not an index that produces the best prediction with the benefit of hindsight. In real time, the user of the coincident index will not have future information to "smooth" the estimate of the likelihood of a recession. If the objective was to construct an index, whose Bry-Boschan smoothed states coincided with the NBER indicator, one could estimate the weights of coincident series that would produce the smallest mean squared error in that dimension. Our search over a two digit grid on the simplex resulted in 185 sets of weights that produced only two wrong predictions out of 450 data periods, which is half of the wrong predictions of the TCB index.[15]

# 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 employ 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).

[15] We thank Mark Watson for suggesting this exercise to us. The Gauss code for the Bry-Boschan algorithm was downloaded from Mark Watson's Home-Page.

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 (84%), which is in agreement with the 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 the probability of recessions.

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 the Bry and Boschan (1971) algorithm applied to 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.

Birchenhall, C., H. Jensen, D. Osborn and P. Simpson (1999), "Predicting US business cycle regimes", *Journal of Business and Economic Statistics*, 17, 313-323.

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/november2001/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.

Zellner, A. and C. Min (1999), "Forecasting Turning Points in Countries' Output Growth Rates", *Journal of Econometrics*, 88, 203 - 206.

# A Appendices

## 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*, $\Delta XCI_t$, which is a latent dynamic factor that represents (the change in) "the state of the economy." Denoting the growth rates of the coincident series as the vector $x_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})'$ , their proposed statistical model is as follows:

$$\begin{aligned} x_t &= \beta + \gamma(L)\Delta XCI_t + u_t, \\ \phi(L)\Delta XCI_t &= \delta + \eta_t, \\ D(L)u_t &= \epsilon_t, \end{aligned} \tag{9}$$

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

The model (9) assumes that there is a single source of comovement among the growth rates of the coincident series – $\Delta XCI_t$. 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 $\Delta XCI_t$ in (9), representing recession and non-recession regimes.

### A.2 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 $y_t^*$:

$$\begin{aligned} E(y_t^* \quad | \quad I_{t+h}) &= \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + u_t \qquad (10) \\ \text{NBER}_t \quad &= \begin{cases} 1 & \text{if } E(y_t^* \mid I_{t+h}) < 0 \\ 0 & \text{otherwise.} \end{cases} . \end{aligned}$$

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

$$\begin{aligned} c_{it} &= \lambda_i \left(\gamma_i' z_t\right) + v_{it}, \qquad i = 1, \cdots, k \qquad (11) \\ \begin{pmatrix} u_t \\ v_t \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{vu}' \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix}\right) \end{aligned}$$

where the $v_{it}$, $i = 1, \cdots, k$, are collected into a $k$-vector $v_t$, $\lambda_i$ and $\gamma_i' z_t$ for $i = 1, \cdots, k$ come from the canonical-correlation analysis, $\Sigma_{vv}$ is a $k \times k$ diagonal variance-covariance matrix of $v_t$, and $\sigma_{vu}$ 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 $\eta_t$ is independent of $v_t$. Substituting for $u_t$ in equation (10), we obtain,

$$\begin{aligned} E(y_t^* \quad | \quad I_{t+h}) &= \beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta + \eta_t \\ \text{NBER}_t &= \begin{cases} 1 & \text{if } \eta_t < -\left(\beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta\right) \\ 0 & \text{if } \eta_t \geq -\left(\beta_0 + \beta_1 c_{1t} + \cdots + \beta_k c_{kt} + v_t'\delta\right) \end{cases}. \end{aligned} \tag{12}$$

Notice that, by construction, all the regressors in (12) are uncorrelated with the error term $\eta_t$. 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 $\Sigma_{vv}$.

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 $\theta$, if $\eta_t$ is i.i.d., the following central-limit theorem holds:

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

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

The error term $\eta_t$ 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 | $\lambda_j^2$ *and all smaller* $\lambda_j^2 = 0$ |
|---|---|---|
| $\lambda_j^2$ | | P-Values (df corrected test) |
| 0.5365 | 208 | 0.0000 |
| 0.3370 | 153 | 0.0000 |
| 0.2484 | 100 | 0.0000 |
| 0.1360 | 49 | 0.1768 |

**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 | 0.60 | 1 | | | | | |
| Basis Cycle 2 | 0.15 | 0 | 1 | | | | |
| Basis Cycle 3 | 0.15 | 0 | 0 | 1 | | | |
| Leading Factor 1 | 0.63 | 0.73 | 0 | 0 | 1 | | |
| Leading Factor 2 | 0.14 | 0 | 0.58 | 0 | 0 | 1 | |
| Leading Factor 3 | 0.11 | 0 | 0 | 0.49 | 0 | 0 | 1 |

**Table 4: Two Stage Conditional Maximum Likelihood Estimates**

| Regressor | Est. Coeff. | Std. Error |
|---|---|---|
| $c_{1t}$ | 65.21 | 8.86 |
| $c_{2t}$ | 28.05 | 6.41 |
| $c_{3t}$ | 13.00 | 6.75 |
| Constant | -0.33 | 0.19 |
| p-value of overall significance | | <0.01 |
| McFadden R-Squared | | 0.71 |
| % overall correct prediction | | 94.89 |

**Table 5: Specification Test Results**

| Dependent variable is NBER | index$_1$ is $\Delta$IVCI index$_2$ is $\Delta\log$(TCB) | | index$_1$ is $\Delta$IVCI index$_2$ is $\Delta\log$(XCI) | |
|---|---|---|---|---|
| Regressor | *Coeff.* (*Std. Err.*) | *Coeff.* (*Std. Err.*) | *Coeff.* (*Std. Err.*) | *Coeff.* (*Std. Err.*) |
| index$_1$ | $-0.98$ (0.46) | $-1.08$ (0.60) | $-0.90$ (0.23) | $-1.41$ (0.47) |
| index$_1^2$ | - | $-0.38$ (1.02) | - | $-0.33$ (0.72) |
| index$_1^3$ | - | 1.87 (1.84) | - | 1.96 (1.11) |
| index$_2$ | $-0.14$ (0.42) | $-0.52$ (0.55) | $-0.15$ (0.13) | $-0.06$ (0.26) |
| index$_2^2$ | - | 1.00 (0.78) | - | 0.35 (0.20) |
| index$_2^3$ | - | $-0.56$ (1.19) | - | $-0.16$ (0.15) |
| Constant | 0.37 (0.04) | 0.31 (0.05) | 0.36 (0.04) | 0.30 (0.05) |
| P-value for "index$_1$ is insignificant" | 0.03 | 0.31 | $< 0.001$ | $< 0.001$ |
| P-value for "index$_2$ is insignificant" | 0.74 | 0.35 | 0.24 | 0.26 |

In both tables, all equations are estimated using 4 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.

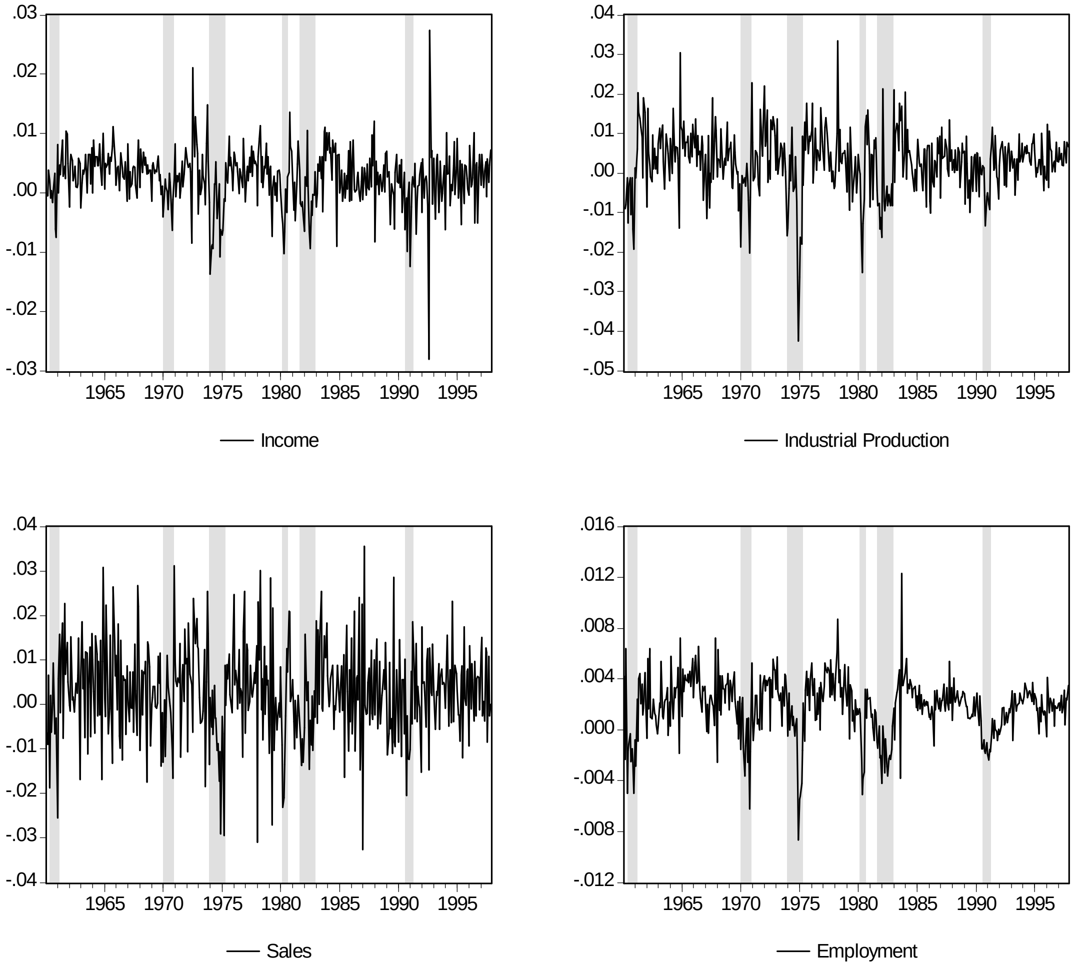


**Figure 1: The Coincident Series**

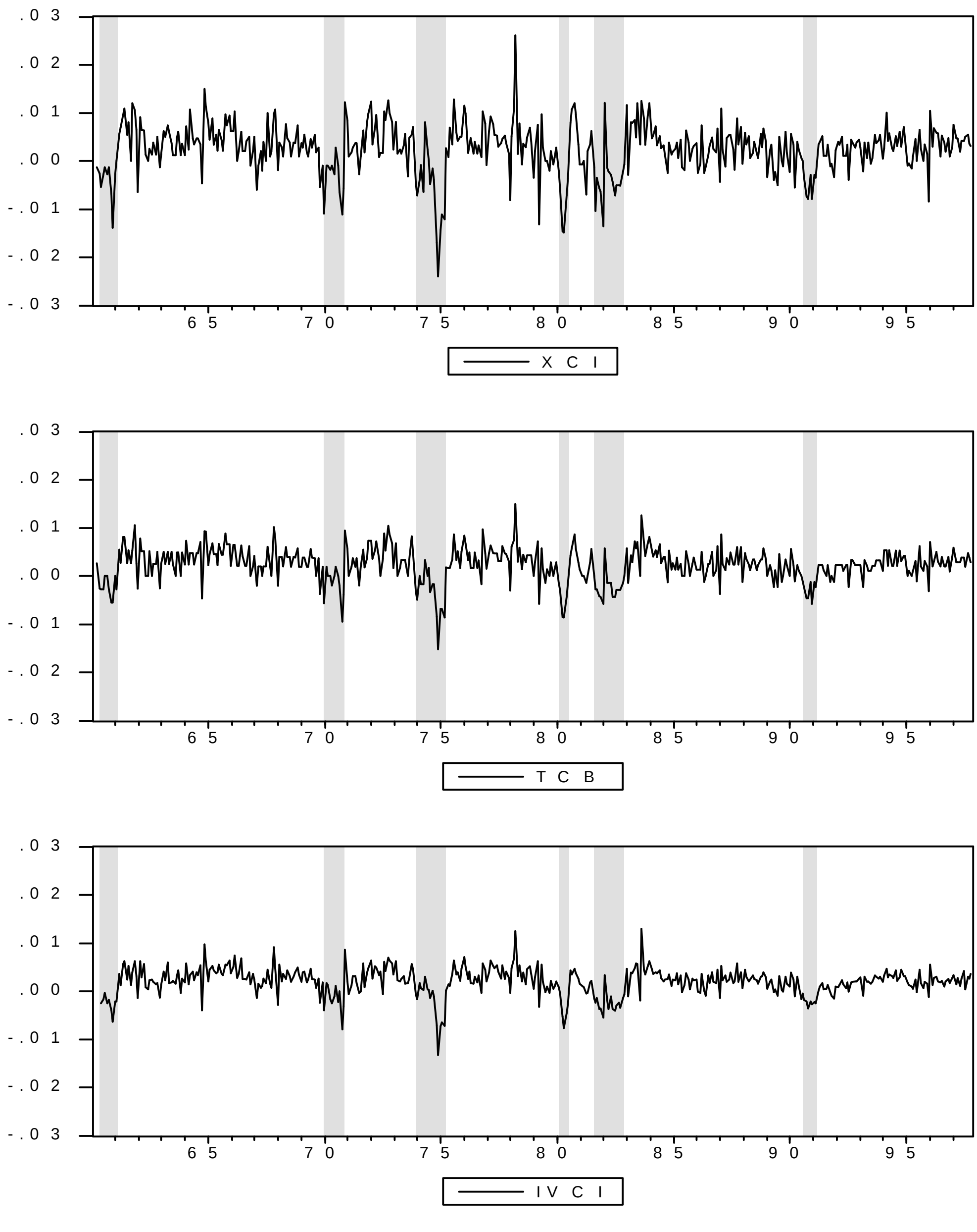


**Figure 2: Growth Rates of Coincident Indices**

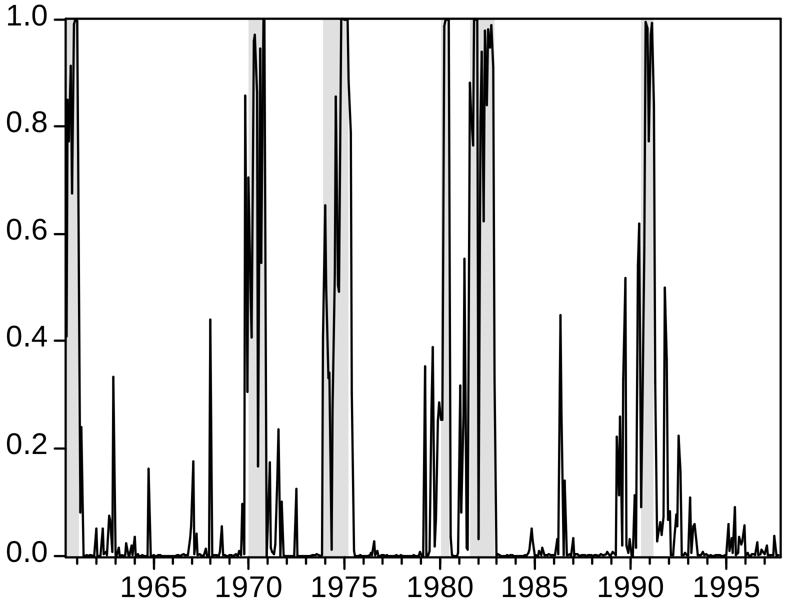


**Figure 3: Recession Probabilities – IVCI**

# ENSAIOS ECONÔMICOS DA EPGE

445. 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 – 31 págs.

446. SPECULATIVE ATTACKS ON DEBTS, DOLLARIZATION AND OPTIMUM CURRENCY AREAS – Aloisio Araujo; Marcia Leon – Maio de 2002 – 74 págs.

447. VINTAGE CAPITAL, DISTORTIONS AND DEVELOPMENT – Samuel de Abreu Pessoa; Rafael Rob – Maio de 2002 – 43 págs.

448. GOVERNMENT ACTIONS TO SUPPORT COFFEE PRODUCERS – AN INVESTIGATION OF POSSIBLE MEASURES FROM THE EUROPEAN UNION – Germán Calfat; Renato G. Flôres Jr. – Junho de 2002 – 30 págs.

449. A MODEL OF CAPITAL ACCUMULATION AND RENT-SEEKING – Paulo Barelli; Samuel de Abreu Pessôa – Junho de 2002 – 50 págs.

450. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY – João Victor Issler; Farshid Vahid – Junho de 2002 – 29 págs.

451. LENDER LIABILITY IN THE CONSUMER CREDIT MARKET – Elisabetta Iossa; Giuliana Palumbo – Agosto de 2002 – 20 págs.

452. DECISION RULES AND INFORMATION PROVISION: MONITORING VERSUS MANIPULATION - Elisabetta Iossa; Giuliana Palumbo – Agosto de 2002 – 37 págs.

453. ON CERTAIN GEOMETRIC ASPECTS OF PORTFOLIO OPTIMISATION WITH HIGHER MOMENTS – Gustavo M. de Athayde; Renato G. flôres Jr. – Setembro de 2002 – 21 págs.

454. MENSURANDO A PRODUÇÃO CIENTÍFICA INTERNACIONAL EM ECONOMIA DE PESQUISADORES E DEPARTAMENTOS BRASILEIROS - João Victor Issler; Tatiana Caldas de Lima Aché Pillar – Setembro de 2002 – 39 págs.

455. FOREIGN DIRECT INVESTMENT SPILLOVERS: ADDITIONAL LESSONS FROM A COUNTRY STUDY - Renato G. Flores Jr; Maria Paula Fontoura; Rogério Guerra Santos – Setembro de 2002 – 30 págs.

456. A CONTRACTIVE METHOD FOR COMPUTING THE STATIONARY SOLUTION OF THE EULER EQUATION - Wilfredo L. Maldonado; Humberto Moreira – Setembro de 2002 – 14 págs.

457. TRADE LIBERALIZATION AND THE EVOLUTION OF SKILL EARNINGS DIFFERENTIALS IN BRAZIL - Gustavo Gonzaga; Naércio Menezes Filho; Cristina Terra – Setembro de 2002 – 31 págs.

458. DESEMPENHO DE ESTIMADORES DE VOLATILIDADE NA BOLSA DE VALORES DE SÃO PAULO - Bernardo de Sá Mota; Marcelo Fernandes – Outubro de 2002 – 37 págs.

459. FOREIGN FUNDING TO AN EMERGING MARKET: THE MONETARY PREMIUM THEORY AND THE BRAZILIAN CASE, 1991-1998 - Carlos Hamilton V. Araújo; Renato G. Flores Jr. – Outubro de 2002 – 46 págs.

460. REFORMA PREVIDENCIÁRIA: EM BUSCA DE INCENTIVOS PARA ATRAIR O TRABALHADOR AUTÔNOMO - Samantha Taam Dart; Marcelo Côrtes Neri; Flavio Menezes – Novembro de 2002 – 28 págs.

461. DECENT WORK AND THE INFORMAL SECTOR IN BRAZIL – Marcelo Côrtes Neri – Novembro de 2002 – 115 págs.

462. POLÍTICA DE COTAS E INCLUSÃO TRABALHISTA DAS PESSOAS COM DEFICIÊNCIA - Marcelo Côrtes Neri; Alexandre Pinto de Carvalho; Hessia Guilhermo Costilla – Novembro de 2002 – 67 págs.

463. SELETIVIDADE E MEDIDAS DE QUALIDADE DA EDUCAÇÃO BRASILEIRA 1995-2001 - Marcelo Côrtes Neri; Alexandre Pinto de Carvalho – Novembro de 2002 – 331 págs.

464. BRAZILIAN MACROECONOMICS WITH A HUMAN FACE: METROPOLITAN CRISIS, POVERTY AND SOCIAL TARGETS – Marcelo Côrtes Neri – Novembro de 2002 – 61 págs.

465. POBREZA, ATIVOS E SAÚDE NO BRASIL - Marcelo Côrtes Neri; Wagner L. Soares – Dezembro de 2002 – 29 págs.

466. INFLAÇÃO E FLEXIBILIDADE SALARIAL - Marcelo Côrtes Neri; Maurício Pinheiro – Dezembro de 2002 – 16 págs.

467. DISTRIBUTIVE EFFECTTS OF BRAZILIAN STRUCTURAL REFORMS - Marcelo Côrtes Neri; José Márcio Camargo – Dezembro de 2002 – 38 págs.

468. O TEMPO DAS CRIANÇAS - Marcelo Côrtes Neri; Daniela Costa – Dezembro de 2002 – 14 págs.

469. EMPLOYMENT AND PRODUCTIVITY IN BRAZIL IN THE NINETIES - José Márcio Camargo; Marcelo Côrtes Neri; Maurício Cortez Reis – Dezembro de 2002 – 32 págs.

470. THE ALIASING EFFECT, THE FEJER KERNEL AND TEMPORALLY AGGREGATED LONG MEMORY PROCESSES - Leonardo R. Souza – Janeiro de 2003 – 32 págs.

471. CUSTO DE CICLO ECONÔMICO NO BRASIL EM UM MODELO COM RESTRIÇÃO A CRÉDITO - Bárbara Vasconcelos Boavista da Cunha; Pedro Cavalcanti Ferreira – Janeiro de 2003 – 21 págs.

472. THE COSTS OF EDUCATION, LONGEVITY AND THE POVERTY OF NATIONS - Pedro Cavalcanti Ferreira; Samuel de Abreu Pessoa – Janeiro de 2003 – 31 págs.

473. A GENERALIZATION OF JUDD'S METHOD OF OUT-STEADY-STATE COMPARISONS IN PERFECT FORESIGHT MODELS - Paulo Barelli; Samuel de Abreu Pessoa – Fevereiro de 2003 – 7 págs.

474. AS LEIS DA FALÊNCIA: UMA ABORDAGEM ECONÔMICA - Aloísio Pessoa de Araújo – Fevereiro de 2003 – 25 págs.

475. THE LONG-RUN ECONOMIC IMPACT OF AIDS - Pedro Cavalcanti G. Ferreira; Samuel de Abreu Pessoa – Fevereiro de 2003 – 30 págs.

476. A MONETARY MECHANISM FOR SHARING CAPITAL: DIAMOND AND DYBVIG MEET KIYOTAKI AND WRIGHT – Ricardo de O. Cavalcanti – Fevereiro de 2003 – 16 págs.

477. INADA CONDITIONS IMPLY THAT PRODUCTION FUNCTION MUST BE ASYMPTOTICALLY COBB-DOUGLAS - Paulo Barelli; Samuel de Abreu Pessoa – Março de 2003 – 4 págs.

478. TEMPORAL AGGREGATION AND BANDWIDTH SELECTION IN ESTIMATING LONG MEMORY - Leonardo R. Souza - Março de 2003 – 19 págs.

479. A NOTE ON COLE AND STOCKMAN - Paulo Barelli; Samuel de Abreu Pessoa – Abril de 2003 – 8 págs.

480. A HIPÓTESE DAS EXPECTATIVAS NA ESTRUTURA A TERMO DE JUROS NO BRASIL: UMA APLICAÇÃO DE MODELOS DE VALOR PRESENTE - Alexandre Maia Correia Lima; João Victor Issler – Maio de 2003 – 30 págs.

481. ON THE WELFARE COSTS OF BUSINESS CYCLES IN THE 20TH CENTURY - João Victor Issler; Afonso Arinos de Mello Franco; Osmani Teixeira de Carvalho Guillén – Maio de 2003 – 29 págs.

482. RETORNOS ANORMAIS E ESTRATÉGIAS CONTRÁRIAS - Marco Antonio Bonomo; Ivana Dall'Agnol – Junho de 2003 – 27 págs.

483. EVOLUÇÃO DA PRODUTIVIDADE TOTAL DOS FATORES NA ECONOMIA BRASILEIRA: UMA ANÁLISE COMPARATIVA - Victor Gomes; Samuel de Abreu Pessoa;Fernando A . Veloso – Junho de 2003 – 45 págs.

484. MIGRAÇÃO, SELEÇÃO E DIFERENÇAS REGIONAIS DE RENDA NO BRASIL - Enestor da Rosa dos Santos Junior; Naércio Menezes Filho; Pedro Cavalcanti Ferreira – Junho de 2003 – 23 págs.

485. THE RISK PREMIUM ON BRAZILIAN GOVERNMENT DEBT, 1996-2002 - André Soares Loureiro; Fernando de Holanda Barbosa - Junho de 2003 – 16 págs.

486. FORECASTING ELECTRICITY DEMAND USING GENERALIZED LONG MEMORY - Lacir Jorge Soares; Leonardo Rocha Souza – Junho de 2003 – 22 págs.

487. USING IRREGULARLY SPACED RETURNS TO ESTIMATE MULTI-FACTOR MODELS: APPLICATION TO BRAZILIAN EQUITY DATA - Álvaro Veiga; Leonardo Rocha Souza – Junho de 2003 – 26 págs.

488. BOUNDS FOR THE PROBABILITY DISTRIBUTION FUNCTION OF THE LINEAR ACD PROCESS – Marcelo Fernandes – Julho de 2003 – 10 págs.

489. CONVEX COMBINATIONS OF LONG MEMORY ESTIMATES FROM DIFFERENT SAMPLING RATES - Leonardo R. Souza; Jeremy Smith; Reinaldo C. Souza – Julho de 2003 – 20 págs.

490. IDADE, INCAPACIDADE E A INFLAÇÃO DO NÚMERO DE PESSOAS COM DEFICIÊNCIA - Marcelo Neri ; Wagner Soares – Julho de 2003 – 54 págs.

491. FORECASTING ELECTRICITY LOAD DEMAND: ANALYSIS OF THE 2001 RATIONING PERIOD IN BRAZIL - Leonardo Rocha Souza; Lacir Jorge Soares – Julho de 2003 – 27 págs.

492. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY - JoãoVictor Issler; Farshid Vahid – Agosto de 2003 – 26 págs.