A Common-Feature Model for Coincident Index of Brazilian Economic Activity

Claudia Fontoura Rodrigues - IBRE/FGV

João Victor Issler - Graduate School of Economics - EPGE, Getulio Vargas Foundation.

Hilton Hostalácio Notini - Graduate School of Economics - EPGE,FGV; ANAC.

Área 3 - Macroeconomia, Economia Monetária e Finanças

Resumo

Em economias modernas tem sido essencial o uso de indicadores capazes de resumir o estado da economia. Tal índice deve agregar indicadores capazes de refletir um cenário econômico amplo. Mais importante, esses indicadores precisam ser coincidentes ao ciclo de negócios da atividade economica. Em artigo anterior, construímos um indicador coincidente simples para a economia brasileira desconsiderando pontos de inflexão do ciclo. Isto porque não existia uma série oficial datando periodos recessives. No entanto, recentemente, o “Comitê de Datação de Ciclos Econômicos” (Codace) identificou os pontos de inflexão da atividade economica nacional em base mensal, por isso, neste artigo usamos esta nova informação e aplicamos a técnica de correlação canônica utilizada por Issler e Vahid (2006) unindo o indicador de recessão à variáveis coincidentes tradicionais. Queremos investigar o quanto esta nova informação agregada ao índice melhora sua performance, comparada a performance de datação do nosso índice anterior.

Palavras Chave: características comuns, correlação canônica, ciclos significantes, índice coincidente, indicador de recessão.

Abstract

In modern countries it has been essential the use of an index that summarizes the state of the economy. This index should aggregate some indicators in order to better reflect the broad economic scene. More importantly, these indicators need to be coincident with the reference cycle, also known as business cycle of the economic activity. In a previous article, we built a coincident index for the Brazilian economy disregarding the turning points of a reference business cycle and so did several other authors. This was the consequence of the lack of an official series representing recession periods. However, the Brazilian business-cycle dating committee, "Comitê de Datação de Ciclos Econômicos" (CODACE), recently identified the turning points of the national economic activity in a monthly basis; therefore in this article we took advantage of this new information and followed Issler and Vahid's (2006), linking the recession indicator to traditional coincident variables. We aimed to investigate how much this new information impacted on our previous coincident index. We used data from 1980 to 2009, and series related to employment, sales, income and industrial production. Our research shows that building a link between coincident cycles and an official recession indicator is essential to obtain a better aggregate index of economic activity.

Key Words: common-feature, canonical correlation, significant cycles, coincident index, recession indicator.

JEL Classification: C32, E32.

Introduction

 Monitoring economic activity is a task of great importance in modern economies. Government, private enterprises and individuals are concerned about the state of the economy and its prospects for strategic decision-making purposes, the difficulty being that the state of the economy is an unobservable or latent variable. In the U.S. the NBER (the National Bureau of Economic Research, created in 1920) by the Business Cycle Dating Committee has been responsible for announcing the turning points of the state of the economy. Although their announcements are the result of a consensus opinion among a panel of experts, they are released with a delay (usually of six months to a year) related to the turning points. The latter is not by chance, since the priority of the committee is accuracy. However this precision has a drawback: the deliberations are not immediately useful for current decisions. Ex-post, however, they may be useful for constructing coincident and leading indices of economics activity.

 In order to have some knowledge of the unobservable state of the economy the construction of coincident and leading indices of economic activity has been a common practice for a long time. Initially, the method of choice widely used in practice was based on the heuristic approach of The Conference Board (TCB). Later, this research has mainly focused on sophisticated econometric models trying to capture the main features of the four coincident variables that the NBER is said to follow (employment, income, industrial production, and sales) to estimate coincident and leading indices of economic activity, establish business-cycle turning points, as well as to estimate their respective probability of occurrence; see Stock and Watson (1988a, b, 1989, 1991, 1993a), Hamilton (1989).

 Arguably, these models miss a key variable that should be included in them -- the NBER decisions on U.S. turning points as determined by its business-cycle dating committee. Although this information is usually available with a considerable lag, there is no reason not to include it ex-post on econometric models and then use it in real-time analysis to infer the state of the economy. This point was forcefully made in Issler and Vahid (2006), who proposed an instrumental-variable coincident index which models NBER decisions using only current coincident series (employment, income, industrial production, and sales). Because the series and the decisions were both originally proposed and provided respectively by the NBER, we can think as this method as unveiling the "missing link" between the series and the decisions, i.e., estimating the weights implicit in NBER's decisions.

 For the U.S., there has been a recent trend to incorporate NBER dating-committee decisions into different econometric models of business cycles. Although some of these contributions were independent, they all recognize that one should not discard the informational content of these decisions in constructing econometric models; see Birchenhall et al. (1999), Dueker (2005), Issler and Vahid, and Chauvet and Hamilton (2006).

 In Brazil, up to very recently, we did not have a business-cycle dating committee investigating the phases of the Brazilian business cycle. This committee, called "Comitê de Datação dos Ciclos Econômicos", or CODACE, had its first deliberation in 2009. A monthly chronology of Brazilian recessions became available in 2010, which motivates its current use jointly with a set of coincident and leading series.

 With a few exceptions, research on coincident and leading indices of economic activity in Brazil is fairly young and dates from the 2000's. Chauvet (2001) and Picchetti and Toledo (2002) use common-factor models to generate a monthly coincident indicator of economic activity. Chauvet (2002) uses a two-state Markov Chain characterizing a recession or an expansion to propose a chronology for Brazilian business cycles.

 On a much broader study, Duarte, Issler and Spacov (2004) evaluated three different methods for constructing composite coincident indices: The Conference Board's (TCB's); Spacov's (2000), and Issler and Vahid's (2006). Using quadratic loss, the business-cycle dating of these three methods were compared with that of a monthly proxy of Brazilian GDP, suggesting that the Brazilian coincident index should use the methodology put forth by TCB. Since in 2004 Brazil did not have a widely accepted chronology of recessions, these authors could not have relied on a dating-committee decisions for the state of the economy. Instead, they relied on proxies for these decisions, represented either by the turning points of GDP or by Harding and Pagan's (2006) concordance index for the coincident series.

 In this paper, we apply the common-feature technique proposed in Issler and Vahid (2006) to build a coincident index of economic activity for Brazil. There are two main ingredients in constructing this index. The first is a generally accepted business-cycle dating for recessions, which has recently been available after the Brazilian Business Cycle Dating Committee has been created -- CODACE. The second is the choice of a set of coincident and leading variables to serve as bases for the coincident index. As usual, the coincident variables are industrial production, income, employment and sales. The set of leading variables includes lags of the coincident series as well as series that seem to precede economic activity as represented by the coincident series. The common-feature technique identifies the significant canonical correlations, found among coincident and leading variables. Canonical correlations can be thought to be the strength by which the coincident series correlate with the past, represented by the group of leading series. Non-significant canonical correlations, associated with linear combinations of the coincident series, are discarded as noise. These two ingredients are put together when instrumental-variable Probit regressions are run using CODACE's decisions on linear combinations of the coincident series associated with significant canonical correlations, leading to a weighted average of the coincident series that can be used to track the state of the economy using optimally estimated weights.

 Issler and Vahid technique identifies, in the econometric sense, the coincident index by assuming that the coincident variables have a common cycle with the unobserved state of the economy, and that the CODACE business cycle dates signify the turning points in the unobserved state. Because of interest in constructing indices of business-cycle activity, the cyclical parts of the coincident series are used in this structural equation. This ensures that noise in the coincident series does not affect the final index. The instrumental-variable Probit regression is estimated 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 set.

 In putting together the database for this paper, we search for series that satisfy the following conditions: (a) to be observable at a monthly frequency for the period 1980-2010; (b) timely data releases, and having small revisions regarding final data figures. We also include business-tendency survey data in line with the recent research by Issler, Notini and Rodrigues (2008), which have been shown to be particularly suitable for business cycle monitoring and forecasting.

 Finally, we revisit part of the exercise in Duarte, Issler and Spacov, comparing the TCB method index proposed there, and also examined by Issler, Notini and Rodrigues, with the Issler and Vahid method used here for Brazil. The main comparison is in terms of the ability of both indices to date the state of the Brazilian economy; we used the quadratic loss measure to compare both indexes.

 This article is organized as follows. Section 2 presents the econometric theory in use, detailing both the canonical correlation and the probabilistic model. Section 3 describes all the steps we followed to obtain the Brazilian coincident index. Section 4 concludes our research.

The Econometric Theory

 As stated in the introduction section we assume that the coincident variables have a common cycle with the unobserved state of the economy; therefore we built our index using a linear combination of all the significant cycles identified in these coincident variables. To obtain the significant cycles we applied the canonical correlation procedure followed by a likelihood ratio test. To linearly combine the cycles we used the weights generated by a probabilistic model, which regressed the official recession indicator on the significant cycles. The next two subsections explain both procedures.

 We borrow the cyclical definiton in Issler and Vahid (2006) which states that a cyclical variable is one which can be linearly predicted from some set of variables defining a past information set. Because of this high dependence among variables (present and past values), determining an appropriate past information set is essential. A suitable one usually includes lags of both leading and coincident variables.

 Yet there are infinitely many linear combinations of the coincident variables that are predictable from a given information set which means there are infinitely many cycles. In order to restrict the number of the cycles, we found the canonical-correlation analysis a most appropriate method, as it determines a basis for the space of cycles.

 The canonical-correlation was introduced by Hotelling (1935, 1936), but it was Akaike (1976) who first used it in a multivariate time series context. He defined the canonical correlation as the `strength' of a channel that links two distinct sets. In this article the canonical correlation is used to identify the links between coincident variables and variables in a past information set. These links in fact are linear combinations of coincident variables most predictable from linear combinations of variables in the past information set; the strength of each link is measured by the squared canonical-correlation number which has a meaning similar to R² in the simple regression analysis.

 As most traditionally accepted the coincident series we used correspond to output, income, sales and employment; the series' specifications and transformations applied on them are described in the appendix. For the leading set we tested a group of 50 variables described in the appendix.

The Canonical Correlation Analysis

 Using a more suitable notation, we explain the inputs and outputs of the canonical correlation procedure.

 Consider that our group of coincident series (income, employment, sales and industrial output) constitutes a vector denoted by xt =(x1t,x2t,x3t,x4t)′ and that our list of m instruments (m≥4), the past information set, constitutes the vector zt. The canonical correlation procedure returns a set of four linearly independent combinations of the elements in xt, which we refer to as A(xt)=(α₁′xt,α₂′xt,α₃′xt,α₄′xt) that contains the most predictable linear combinations of xt given zt (the past information set). From our definition of cycles, this means that the elements in A(xt) are in fact cycles; also, from the linearly independent property of these elements they constitute a base of the space of cycles. For that reason A(xt) is called the set of "basis cycles".

 The order of the elements in A(xt) defines their degree of likelihood given zt. The α₁′xt cycle corresponds to the linear combination of xt that is most linearly predictable from zt, α₂′xt is the second most linearly predictable from zt and so on.

 Arising as a by-product of this analysis Γ(zt)=(γ₁′zt,γ₂′zt,γ₃′zt,γ₄′zt) consists of four linearly independent combinations of zt which have the highest squared canonical correlation with xt (that means γi′zt is the one with the highest correlation with αi′xt, for i=1,2,3,4). The squared canonical correlation, which we denote by λ=(λ₁²,λ₂²,λ₃²,λ₄²), corresponds to the R² of a simple regression between αi′xt and γi′zt. This statistic is useful to measure the significance of each cycle (or a group of them). If λi is small, the cycle related to it, may be eliminated from the group A(xt) without affecting its capacity of generating most of the actually predictable cycles, given zt; in this case the dimension of A(xt) can be decreased (discarding the non significant cycles). The significance test used to evaluate whether the smallest squared canonical correlation (or a group of them) is statistically equal to zero, was the likelihood ratio test (LR), with the null hypothesis being k significant cycles (i.e 4-k insignificant canonical correlations) exist. The LR formula is:

 LR=-T Σ⁴i=k+1ln(1-λi²)

 which has an asymptotic χ² distribution with (4-k)(m-k) degrees of freedom (Anderson, 1984); note that the number of instruments (m) in the information set directly influences the distribution of LR statistics. To improve the finite sample performance (T-m) is usually used instead of T. If the null is not rejected for any basis cycle k, then the other (4-k) basis cycles are deemed to be insignificant which means they cannot be predicted from the information set and therefore they can be dropped from A(xt). If this is the case, we conclude that all cyclical behavior in the four coincident series can be written in terms of k basis cycles and that reducing the cycle space (to a dimension less than 4) does not prevent its usefulness in capturing (or generating) possible cycles. More technically we can state that not all the space generated by the basis cycles is actually necessary to generate the most predictable linear combinations of xt given zt.

 Summing up, the canonical correlation analysis only changes the coordinates of xt to A(xt) which simultaneously brings Γ(zt). No information is added or thrown away in the process of moving xt to the new coordinates in A(xt) because this latter is the base of the space that generates xt. From A(xt) a likelihood ratio test can be applied to investigate if all its elements are in fact significant, non significant ones can be discarded without loss of information. In this way the canonical-correlation sheds light into a would-be coincident index of economic activity.

The Probabilistic Model: linking the significant common-cycles to the Codace's recession indicator

 After determining the significant cycles we needed to define a way to combine them in an index. We decided on a linear combination of them, whereby the weights of each came from a probabilistic model, one that related the information content in the recession indicator (provided by Codace) with the significant basis cycles.

 A straightforward but imperfect way to relate these variables (the recession indicator, and the significant cycles) would be a probit model having the recession indicator vector (defined here as a vector of 1's during recessions and 0's otherwise) as the dependent variable and the significant cycles as the independent ones. Although simple, this model loses some essential information, clearly stated in our main assumption:

 Assumption 1: A linear index exists of the cyclical parts of the coincident series which has the exact same correlation pattern with the past information as the unobserved state of the economy.

 Bear in mind that being a linear combination of the cyclical parts of the coincident series, means that the index itself ultimately is a linear combination of the coincident series (as the cycles themselves are linear combinations of the coincident series).

 Let yt\* denotes the unobserved state of the economy and {c1t,c2t} denotes the significant basis cycles of the coincident series at time t. From assumption 1 (an index, built as a linear combination of the cycles, and the state of the economy have the same correlation with the past information set) there must be a linear combination of yt\* and {c1t,c2t} which is unpredictable from the information set before time t. That is,

 E(yt∗-β₀-β₁c1t-β₂c2t|It-1)=0

 where It-1 is the information available at time t-1. If yt\* was observed, we could estimate the β′s directly by GMM. However yt\* is not observed. Our best knowlegde of it, is contained by the Codace recession indicator (denoted by CRI), which is equal to 1 (at time t) when the Committee at time t+h identifies a recession (at t) and 0 otherwise.

 CRIt= 1, if E(yt\*|It+h)<0

0, otherwise.

 From the equation above we obtain:

E(yt\*|It-1) =E(β₀+β₁c1t+β₂c2t|It-1)

 =β₀+β₁E(c1t|It-1)+β₂E(c2t|It-1)

 =β₀+β₁c1t+β₂c2t+ωt

 where E(ωt|It-1)=0,and obviously correlated with all cit, i=1,2. Because we can always write:

 E(yt\*|It+h)=E(yt\*|It-1)+ζt+ζt+1+...+ζt+h

 where ζt+i is the surprise associated to new information arriving in period t+i, it is straightforward to show that:

E(yt\*|It+h)=β₀+β₁c1t+β₂c2t+ut

ut = ωt+ζt+ζt+1+...+ζt+h,

 where ut is unforecastable given information at time t-1, i.e., E(ut|It-1)=0, has a "forward" MA(h) structure, and is correlated with cit,i=1,2. Because of this correlation, in order to estimate β₁ and β₂ consistently we needed to use instrumental variables. Our best instruments would be the zt variables (lags of leading, and coincident variables) already defined for the canonical analysis.

 Among the methods adequate for estimating coefficients of single equation with limited dependent variable in a simultaneous equation model we used the two-stage conditional maximum likelihood (2SCML) estimator proposed by Rivers and Vuong (1988) due to its relative simplicity.

 Because we assumed that the coincident variables could be fully explained by the significant basis cycles, the first stage of the 2SCML estimating procedure involves regressing {c1t,c2t} on the instruments zt and saving the residuals, which we denote as {v1t,v2t. In the second stage, both the basis cycles and the residuals from the first stage are included in a probit model:

 Pr(CRIt=1)=Φ(β₀+β₁c1t+β₂c2t+β₃v1t+β₄v2t)

 where Φ is the standard normal cumulative distribution function. The outcome of the probit estimate will be the final result of the 2SCML method. Rivers and Vuong (1988, p. 354) describes a way to calculate the standard errors of the estimates; in addition, as we ignored the dynamic structure of ut in constructing the likelihood function, autocorrelation-robust standard errors have to be used.

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

ΔIVCIt=β₁c1t+β₂c2t

 =β₁α₁′xt+β₂α₂′xt

 =(β₁α₁′+β₂α₂′)xt

 which shows that our index is in fact a linear combination of xt.

The Coincident Index

 Our IV-CI is built from the cycles (resulted from the canonical correlation analysis) weighted by a maximum likelihood procedure, which links recession periods to significant cycles.

 In order to obtain the IV-index we first needed to propose, along with the coincident series, a reasonable instrumental set. The latter should contain elements that anticipate the coincident variables as stated in the canonical-correlation theoretic section. We describe the steps taken to build our instrumental set in the following section. Then we describe the results obtained from the canonical correlation analysis and the probabilistic model. To assess the performance of the IV-index we finally built and index following the TCB methodology.

The Information Set

 In order to provide a reasonable information set for the canonical correlation analysis we needed to choose appropriate series to be part of it. For this purpose we followed some guidelines described by Stock and Watson (1989, 1993a) which can be summarized as: i) build a large list of national series quite related to economic activity that also represents a broader scene; these series should be released on a monthly basis and they should not take long to become available for public use. The set of variables considered in this phase is listed in the appendix and contains 50 series. ii) apply some necessary functions to make data suitable for our analysis (both correlation and causality tests); the steps in this item correspond to: deflating nominal variables by a Brazilian general price index (IGP-DI), seasonally adjusting the data that showed a seasonal pattern (the X-11 algorithm was used), taking log of all the series except the ones expressed in growth rates in order to decrease their variances, first-differentiating series that we could not reject the unit root assumption; iii) test the Granger causality hypothesis (in lags 3, 6, 12) considering a leading candidate and each time one of the four coincident variables. The candidate series for which we could not reject the non-causality assumption for at least 3 coincident variables were eliminated. The series that exhibited a high contemporaneous correlation with any coincident series were also discarded.

 In addition to the above we also computed a quadratic loss function or the quadratic probability score (QPS), a measure introduced by Diebold and Rudebusch (1989) to compare turning points of leading candidate series with turning points of the four coincident series. In order to perform this analysis, we first computed for all the leading and coincident variables their turning points by using the Monch and Uhlig (2005) dating procedure, having the turning points we built a recession indicator vector for each variable, which contained 1 during recession periods and 0 otherwise. The QPS related to any two variables is:

 QPSij=(1/T)(Si,t-k-Rj,t)²

 where S (R) is the recession indicator vector for a leading (coincident) variable identified by the subscript i (j); k is the optimal number of shifts in the leading recession vector that we should consider, in order to minimize the overall QPS number. A QPS of 0 indicates that recession periods of both variables are alike.

 Table 1 reports the information set we used; it corresponds to the series selected from a much more general leading set after all the preceding recomendations were followed.

Table 1: Instrumental set components

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Series | Transformation | Deflated | Seasonal Adjust | Source |
| Expected industrial production | Δln() | no | yes | IBRE/FGV |
| Expected job creation in manufacturing | Δln() | no | yes | IBRE/FGV |
| Expected domestic demand | Δln() | no | yes | IBRE/FGV |
| Expected global demand | Δln() | no | yes | IBRE/FGV |
| Financial market index | Δln() | yes | no | Bovespa |
| Manufacturing stocks level | Δln() | no | yes | IBRE/FGV |
| Monetary aggregate | Δln() | yes | no | Bacen |
| Hours worked in industry  | Δln() | no | yes | FIESP |

The Coincident Variables

 For the Brazilian coincident series we needed variables promptly released and reported on a monthly basis. Also, as we consider the TCB coincident variables reliable proxies for a country's economic activity (as they cover distinct sectors of the overall economic scene) we searched for national series that closely related to them. The TCB coincident series are: industrial production, employees on non-agricultural payrolls, manufacturing and trade sales, and personal income less transfer payments. The Brazilian series that closely approach to this specification are: general industrial production (Yt), income (It), occupied employed population (Nt) and corrugated paper (St) as a proxy for sales. All are compiled on a monthly basis and are promptly released. Table 2 summarizes some information about these series:

Table 2: Coincident Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Series | Transformation | Seasonally Adjusted | Source |
| Industrial Production | Δln() | yes | IBGE |
| Employees | Δln() | yes | IBGE |
| Sales | Δln() | yes | ABPO |
| Income | Δln() | yes | IBGE |

 Because our analysis extended from 1980 onward, we needed to input values for income and employment before 2000. In the referred year a methodological change was enforced on both series causing a discontinuity in their time paths. To handle this issue we used a backcast procedure, implemented through a Kalman filter; this was first proposed by Issler, Notini and Rodrigues (2008) in order to build a TCB coincident index. The resulting backcasted series as well as a complete description of all of the coincident variables and graphs of their rate of growths are detailed in the appendix.

 After backcasting the two incomplete series, we obtained the whole data set for the sample period we are interested in. We could then analyze series' individual and joint behavior. First, the Augmented Dickey-Fuller unit root test could not reject the null hypothesis (H0: series has a unit root) for any of the coincident variables; therefore to stationarize the series, we used first-differences of their logarithms in the canonical analysis. Next, the Johansen cointegration test, which investigates long run co-movements among variables, could not reject the hypothesis of no cointegration by both the trace and the maximum eigenvalue test at 5% significance level, which indicates no long run behavior is shared by the coincident series; this outcome allowed us to discard linear combinations of the coincident series (in level) from the information set.

The Instrumental Variable Index

 The instrumental variable coincident index (henceforth, IV-CI) is formed by a linear combination of the coincident variables which weights result from a canonical correlation analysis and a subsequent probit model.

 From the canonical correlation theory we know that a well defined conditional set is essential for the correlation analysis to identify common cycles. The eight leading series we used are listed in Table 1.

 After defining a base conditional set, we needed to determine the lag length that better helps to fit coincident variables path. For this reason we estimated some vector autoregressive models (VAR) of (lnYt,lnSt,lnNt,lnIt) each time with a distinct conditional set (only lag lengths of the series in it were subject to change). From the Akaike information criteria, a VAR of order 2 well suited our modeling needs. However, only a VAR of order 3 produced residuals with no serial correlation; therefore this latter lag was used in the subsequent canonical analysis. In brief, our conditional set was formed by up to three lags of both the coincident series as well as the leading ones.

 Applying the canonical correlation method we obtained the outcome reported in Table 3. The canonical correlation brought the basis for the space of coincident variables (also called the basis cycles). A subsequent likelihood ratio test (LR test) calculated the probability of the cycles being significant. From the values reported in Table 3, the null hypothesis of cycle i (and all subsequent cycles) being insignificant could be rejected at a 1% level (column 4) for all i, meaning that all the basis cycles were necessary to generate the common cycles identified in the coincident series.

 In this way the cyclical behavior of (ΔlnYt,ΔlnSt,ΔlnNt,ΔlnIt) could only be generated by the four orthogonal basis cycles (c1,c2,c3,c4). From Table 3 we also note, the more significant is the cycle, the higher is its squared canonical correlation; which also means the cycle is the highly predictable from the conditional set.

Table 3 - Canonical Correlation outcome

|  |  |  |  |
| --- | --- | --- | --- |
| Basis Cycles (ci) | Sq. Canonical Correlations (λi²) | Degrees of freedom | LR test p-value(λi² and all subsequent λi²=0) |
| c1 | 0.58 | 156 | 5.14e⁻⁰⁵⁷ |
| c2 | 0.43 | 114 | 1.44e⁻⁰²⁵ |
| c3 | 0.27 | 74 | 9.47e⁻⁰¹⁰ |
| c4 | 0.20 | 36 | 0.0003 |

As established in the theoretical section, basis cycles are linear combinations of coincident series and because of this they can be reported in a matrix notation:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | c1t |  |  |  | -0.52 | 0.44 | 1.82 | -0.6 |  |  |  | ΔlnYt |  |
|  | c2t |  | = |  | 0.880 | 0.55 | 0.54 | -0.5 |  |  |  | ΔlnSt |  |
|  | c3t |  |  |  | 1.51 | -2.3 | 0.48 | 2.63 |  |  |  | ΔlnNt |  |
|  | c4t |  |  |  | -0.72 | -0.6 | 0.13 | 2.50 |  |  |  | ΔlnIt |  |

 The IV-CI is a weighted average of the significant cycles, in our case, all of them. The weights were estimated by the two-stage maximum likelihood procedure (2SCML) proposed in 1988 by Rivers and Vuong. In the first stage, basis cycles are regressed on their corresponding leading factor, and the residuals of the regressions are saved. In the second step basis cycles and the residuals (of step 1) are used as regressors in a probit model that used the recession indicator of the Brazilian Dating Committee as the dependent variable.

 Table 5 reports the results of the 2SCML procedure; column 2 and 3 present the coefficients and standard deviations respectively of each series in the probit model. Note that for the two less significant cycles the coefficients of the probit are not significant at 1% level, which indicates that although they were significant to generate cycles, the cycles they generated were ignored by the Dating Committee when identifying recession periods. This may cause some difference between IV-CI turning points and the turning points stated by the committee.

Table 5 - Probit β-Coefficients

|  |  |  |
| --- | --- | --- |
| relative to | Estimates | Corrected std |
| constant | -0.34 | 0.07 |
| c1 | -2.02 | 1.70 |
| c2 | -7.62 | 2.02 |
| c3 | -2.49 | 2.52 |
| c4 | 0.25 | 2.84 |

 The coefficients from the probit along with cycles expressions defined in (<ref>ciclos</ref>) were used to ultimately define the IV-CI (<ref>ivci</ref>) as a combination of coincident series:

 ΔIVCIt=0.95Yt - 0.04St + 0.90Nt + 0.12It

 Note that although IV-CI also assigns a great weight for employment, industrial production is the most important series in it, showing that the Committee heavily considered this latter series when deciding for a turning point.

 Figure 1 plots the IV-CI. Shaded areas are recession periods identified by the Monch and Uhlig algorithm when the IV-CI is the input series. Pairs of dashed lines frame a recession period according to Codace, which stated a total of eight recessions from 1980 to 2009. We see that six out of the eight recessions were well captured by the IV-CI; the other two, although the index turning points did not exactly match Codace's recessions decision, they did not totally ignore them; for the two periods the index showed a severe decrease in economic activity in the beginning months and a high volatility during their time span.



Figure 1: Instrumental Variable Coincident Index for Business Cycles.

The TCB Method for Coincident Index

 The TCB method equally weights each coincident series, after controlling for the inverse of their standard deviation, in a simple linear combination. The formula is:

 Δln(CIt)=(1/4)[((Δln(It))/(σΔln(I)))+((Δln(Yt))/(σΔln(Y)))+((Δln(Nt))/(σΔln(N)))+((Δln(St))/(σΔln(S)))],

where σΔln(I), σΔln(Y), σΔln(N), and σΔln(S) are respectively the standard deviations of income, output, employment, and sales growth. It is straightforward to obtain the level series ln(CIt) or CIt once we have Δln(CIt) given some initial value.

 From our data set the following weights were calculated for each series:

 Δln(CIt)=0.67Δln(It)+0.63Δln(Yt)+0.92Δln(Nt)+0.39Δln(St),

 From the coefficients in the equation above we note that the rate of growth of employment (Nt) is the one that most contributes to the index formation, while the opposite is observed for the rate of growth of sales (St); this is in accordance to TCB formula which gives preference to less volatile variables in the index formation. Refer to the appendix for the graphs of each coincident series rate of growth.

 Figure 2 illustrates the graph of a TCB index. Shaded areas in the graph correspond to recession periods (which are periods framed by a peak and a trough or local maximum and minimum values) dated by the Monch and Uhlig algorithm; dashed lines frame recession periods identified by Codace. Note that the TCB index well identified recent recession periods (after 1994) however it missed or badly indicated earlier ones.



Figure 2: TCB-like index.

Comparing the Turning Points of the Indices

 The turning points of the IV-CI and TCB indices along with the ones from Codace's recession indicator are listed in Table 5. Columns 3 and 4, as well as 7 and 8, report the number of months that leaded (negative numbers) or lagged (positive numbers) the recession begining (peak) and ending (trough) according to the reference indicator provided by the committee.

Table 5: Dates of turning points

|  |  |  |
| --- | --- | --- |
| TCB-like index | IV index | Codace indicator |
| peaks | troughs |  |  | peaks | troughs |  |  | peaks | troughs |
| 1980:10 | 1981:09 | 0 |  | 1981:01 | 1981:09 | +3 |  | 1980:10 |  |
| 1982:07 | 1983:02 |  | 0 | 1982:04 | 1983:02 |  | 0 |  | 1983:02 |
|  |  |  |  | 1987:01 | 1988:11 | -1 | +1 | 1987:02 | 1988:10 |
| 1989:11 | 1990:04 | +5 |  | 1989:06 | 1990:04 | 0 |  | 1989:06 |  |
| 1991:07 | 1991:12 |  | 0 | 1991:08 | 1992:08 |  | +8 |  | 1991:12 |
| 1994:12 | 1995:07 | 0 | -2 | 1995:01 | 1995:08 | +1 | -1 | 1994:12 | 1995:09 |
| 1997:10 | 1999:02 | 0 | 0 | 1997:10 | 1999:02 | 0 | 0 | 1997:10 | 1999:02 |
| 2000:12 | 2001:09 | 0 | 0 | 2000:12 | 2001:07 | 0 | -2 | 2000:12 | 2001:09 |
| 2002:10 | 2003:06 | 0 | 0 | 2002:10 | 2003:06 | 0 | 0 | 2002:10 | 2003:06 |
| 2008:07 | 2009:01 | 0 | 0 | 2008:09 | 2009:01 | +2 | 0 | 2008:07 | 2009:01 |

 Besides the number of leading and lagging months, we calculated the QPS (<ref>qps</ref>) of each index relative to Codace recession turning points. From the figures in table 6, we can see that the IV-CI outperformed the TCB when the overall period was considered and when only the recession times were in the sample. In this latter case the IV's QPS is highly superior than the TCB's (smaller values indicates better index). During growth periods the performance of both indices were very similar. From these outcomes we conclude that the IV-CI is a much more accurate index than the TCB.

 This is a different conclusion from the one reached by Duarte, Issler and Spacov (2004). In their article, they found TCB a very satisfactory and a better index when compared to a canonical correlation one. Although they used the same method we applied here, they did not have a reference recession indicator to use in the probit phase (their data set was also smaller than ours). This lack of information caused the dissimilar results.

Table 6: Quadratic probability score

|  |  |  |
| --- | --- | --- |
|  | IV-CI | TCB |
| whole sample | 0.11 | 0.14 |
| recession periods | 0.25 | 0.41 |
| expansion periods | 0.04 | 0 |

Conclusion

 Considering Brazilian coincident variables to build coincident indices, the first challenge to overcome is to have consistent series from 1980 onward. Two traditional coincident variables (income and employment), because of a huge methodological change in 2002, present a discontinuity in this latter date. In order to have more stable series (from 1980 to 2002) we backcasted both of them using the same procedure adopted by Issler, Notini and Rodrigues (2008) - the Kalman Filter algorithm.

 From a complete data set of coincident variables, and another one formed by selective leading variables which works as an information set, we applied the canonical correlation analysis to identify significant common cycles among the coincident variables. Because none of the cycles could be discarded from the analysis (all of them were deemed significant by the likelihood ratio test) the canonical correlation had a smaller role just shifting the coordinates of the coincident series. The main contribuition of this paper was to incorporate the essential information contained in the recession indicator - provided by Codace (the Brazilian business-cycle dating committee) - to the coincident index. The resulting IV-CI index outperformed the TCB, a different outcome obtained by Duarte, Issler, Spacov (2004). The latter authors could not reach the same result because they did not have a recession indicator to add information to their cycles, and maybe because their sample was smaller.

 The use of a recession indicator as a reference series to better shape a canonical correlation index proved to be an essential step for the index construction as earlier versions of it, which ignored this information, could not build such a robust index.

Appendix

Coincident Series

 The coincident series we used are released by IBGE, The Brazilian Institute of Geography and Statistics, on a monthly basis. Each of them are described below. Shaded areas represent recession periods according to the IV-CI, dashed lines frame recession period according to Codace (a red (green) dashed line is the beginning (end) of a recession period).

 Industrial production: we used the quantum index of general industrial production from 1980 to 2009. The general aspect of this index (it encompasses different industrial sectors) is a desirable feature as we are interested in a broader scene of the industrial activity. Figure A1 depicts the series.



Figure A1: Industrial Production

 Employees: the series that we consider as the number of employees is occupied employed population (N\_{t}) and it has been tracked by IBGE since 2002. The less than ten years of observations is not a long enough time span for our recession analysis, thus applying the Kalman filter procedure we backcasted the series following Issler, Notini and Rodrigues (2008). The state-space equations for the Kalman filter were defined as:

 Nt\*=α₁Ot+εt

 Nt=ht Nt\*

 where Xt is the vector of exogenous, εt~N(0,1), Nt and Nt\* are the observed and unobserved number of employees (in fact, occupied employed population) respectively, and ht=1 from 2002 to 2010, otherwise it is equal to 0, reflecting the unavailability of the data the period. The Kalman filter outcome is the backcasted number of employees from 1980 to 2010, which is then appropriate for our study. Figure A2 presents the observed series (from 2002 onwards) and the backcasted one (from 1980 to 2002).



Figure A2: Employees

Income: the best series for income data is the real average income earned from the main activity of individuals 10 years old and over. This series had a methodology change in 2002. Because of the change some discontinuity is observed in its data; to handle this issue we backcasted the data from 1980 to 2002, using the Kalman filter and covariates proposed in an earlier paper by Issler, Notini, and Rodrigues (2008). The state-space equations for the Kalman filter are similar to the ones defined for employees, only the series in use were distinct.



Figure A3: Income

Sales: the series that describes trade was initially compiled by IBGE in 2000. It represents the volume of sales of the retail market. As we wanted a longer time span we searched for another option. A proxy variable recommended by Duarte, Issler, Spacov (2004) was the corrugated paper series provided by the ABPO - The Brazilian Corrugated Board Association - since 1980. We tested the cross correlation between these series (seasonally adjusted and in its log first difference form) from 2000 to 2009 and the sole significant correlation occurred contemporaneously, which strengthened the use of corrugated paper as a proxy for sales.



Figure A4: Sales (corrugated paper)

Coincident series' rate of growth from 1980 to 2009:



Figure A5: The rate of growth of each coincident variable.

Bibliography

Akaike, H., 1976. Canonical correlation analysis of time series and the use of an information criterion. In: Mehra, R.K., Lainiotis, D.G. (Eds.), System Idendification: Advances and Case Studies. Academic Press, New York, pp. 27-96.

Anderson, H. M., 1997. Transaction costs and nonlinear adjustment towards equilibrium in the US treasury bill market. Oxford Bulletin of Economics and Statistics 59, 465-484.

Anderson, T. W., 1984. An introduction to multivariate statistical analysis, second ed. Wiley, New York.

Barros, A. R., 1993. A periodization of business cycles in the Brazilian economy, 1856-1985, Revista Brasileira de Economia, 47 (1): 53-82, 1993.

Bry, G., Boschan, C., 1971. Cyclical Analysis of Time Series: selected procedures and computer programs. National Bureau of Economic Research, New York.

Burns, A. F., Mitchell, W.C., 1946. Measuring Business Cycles. National Bureau of Economic Research, New York.

Chauvet, M. 2001. A monthly indicator of Brazilian GDP, The Brazilian Review of Econometrics, 21, 1,1-15.

Chauvet, M. 2002. The Brazilian business and growth cycles, Revista Brasileira de Economia, 56, 1, 75-106.

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

Diebold, F.X. and Rudebusch, R., 1989. Scoring the Leading Indicators, Journal of Business, 62, 369 - 402.

Duarte, A., Issler, J.V., Spacov, A., 2004. Indicadores coincidentes de atividade econômica e uma cronologia de recessões para o Brasil.

Estrella, A., Mishkin, F.S., 1998. Predicting U.S recessions: financial variables as leading indicators. Review of Economics and Statistics 80, 45-61.

Hamilton, J. D., 1989. A new approach to the economic analysis of non stationary time series and the business cycle. Econometrica, 57, 357-384.

Harding D. and A. Pagan (2006), Synchronisation of cycles, Journal of. Econometrics, Vol 132 (1), pp. 59-79.

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., Notini, H., Rodrigues, C. F., 2008, Evaluating different approaches in constructing coincident and leading indices of economic activity for the Brazilian economy.

Issler, J. V., Vahid, F., 2006. The missing link: using NBER recession indicator to construct coincident and leading indices of economic activity. Journal of Econometrics132, 281-303.

Mariano, R., Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics, v. 18, p. 427-443.

Monch, E., Uhlig, H., 2005. Towards a monthly business cycle chronology for the Euro area. Journal of Business Cycle Measurement and Analysis 2 (1).

Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics 39, 347-366.

Stock, J., Watson, M., 1989. New indexes of coincident and leading economics indicators. NBER Macroeconomics Annual, p. 351-395.

Stock, J., Watson, M., 1993a. A procedure for predicting recessions with leading indicators: econometric issues and recent experience. In: Stock, J., Watson, M. (eds) New research on business cycles, indicators and forecasting. Chicago: University of Chicago Press.