ÁREA DE INTESSE: TEORIA ECONÔMICA E MÉTODOS QUANTITATIVOS

TÍTULO DO TRABALHO:
FORECASTING BRAZILIAN GDP GROWTH RATE WITH DIFFUSION INDEX MODELS

AUTORES:

ROBERTO TATIWA FERREIRA. MESTRE E DOUTOR EM ECONOMIA PELO CAEN. PROFESSOR DO DEPTO. DE ECONOMIA DA UNIVERSIDADE FEDERAL DO PARÁ.

IVAN CASTELAR. PÓS-DOUTORADO PELA UNIVERSIDADE DE ABERDEEN (GB). PROFESSOR TITULAR DO DEPTO. DE ECONOMIA APLICADA E DO CURSO DE PÓS-GRADUAÇÃO EM ECONOMIA DA UNIVERSIDADE FEDERAL DO CEARÁ.

RESUMO

Este trabalho usa modelos de difusão como ferramenta para prever a taxa de crescimento do PIB brasileiro usando dados trimestrais de 1975 a 2001. Modelos de difusão permitem a recomendação macroeconômica de se usar um grande número de fatores para modelar o comportamento do PIB. Através do uso de componente principais esta massa inicial de fatores é substancialmente reduzida, diminuindo o número de parâmetros e circunscrevendo as predições a uma dimensão extremamente simples de operacionalizar. Os resultados obtidos são extremamente encorajadores quando comparados com previsões geradas por um modelo auto-regressivo.
Palavras chave: Modelos de difusão, previsão, taxa de crescimento do PIB, Brasil.

ABSTRACT

This work uses diffusion index models as the main tool to forecast Brazilian GDP growth rate on the basis of quarterly data running from 1975 to 2001. Diffusion Index models allow to include a large number of factors affecting the behavior of GDP growth rate. Through the use of principal components these factors can be substantially reduced, decreasing the number of parameters to be estimated, and circumscribing prediction to a manageable dimension. The results obtained were extremely encouraging when compared to prediction generated by an autoregressive model.
Key words: Diffusion index models, forecast, GDP growth rate, Brasil.

INTRODUCTION

Forecasting, from the point of view of econometrics, is the process to select and estimate a model in order to make statements about the future. The period in the future to be forecast can vary from short to long time. When short period is the target, one can consider technology constant and to predict some values is the main objective. In the long term the variations in the technology must be forecast and its effects to the model should be considered (Granger, 1980).

Recent lessons from economic forecasting practice has shown that the lack of parsimony is an important cause of forecast failure. This should be expected because the more coefficients there are in a model, the more is the uncertainty about the estimated parameters. Not only this means that some variables, which could give important information about the series to be predicted, would likely be out of the model, but also that lags of the included variables would be restricted.

Factor models for time series have been used to allow the construction of large number of cross-sections in macroeconomic forecasting models. The main idea is that all the information included in a large number of variables could be captured by a few numbers of common factors aamong them. At least two distinct literatures have been using this method. One of these branches is represented by the dynamic factor models (Geweke 1977; Sargent and Sims 1977; Geweke and Singleton 1981; Engle and Watson 1981; Stock and Watson 1989,1991; Quah and Sargent 1993; Kim and Nelson 1998). The common trace in these studies is the effort to estimate the unobservable common factors among some macroeconomic variables, relying in the use of MLE, Kalman filter or both.

The other factor model approach is represented by diffusion indexes models (Connor and Korajczyk 1993; Geweke and Zhou 1996; Forni and Reichlin 1996, 1998; Stock and Watson 1998, 2002), which uses principal components to estimate these common factors.

The main objective of this work is to forecast the Brazilian GDP growth rate. Some authors have been studying different models to forecast the variable in question Moreira, Fior\^{e}ncio and Lopes (1996) used a VAR, VEC, BVAR and BVEC. Moreira and Amendola (1998) used a bayesian vector autoregressive model of lead variables and a dynamic Bayesian model that extracts trend, seasonal and cyclical patterns to the same purpose. Chauvet, Lima and Vasquez (2002) used a Markov switching model to forecast.

In this study the diffusion index (DI) model was used to forecast Brazilian GDP growth rate and these predictions were compared to linear AR forecasts. DI forecasts were made using two kinds of data sets. In the first one, factors were estimated from the current values of 72 predictors. The second data set was constructed allowing for lags[1]. Quarterly data were used from 1975.Q1 to 2003.Q3. One step a head forecasts were produced in a simulated real time design.

[1] Sets with one up to three lags were applied of these predictors

Besides this introduction this study has four more sections. The first one, explains the data used in this work and gives a first glance of the variable to be forecast. The second section, as usual, contains a review of the theoretical background of the work. Subjects like latent variables and factor models, the estimation process and forecast environment used in this study are presented. The third section contains the main results of the forecasting experiment. Conclusions and the main remarks are presented in the last section.

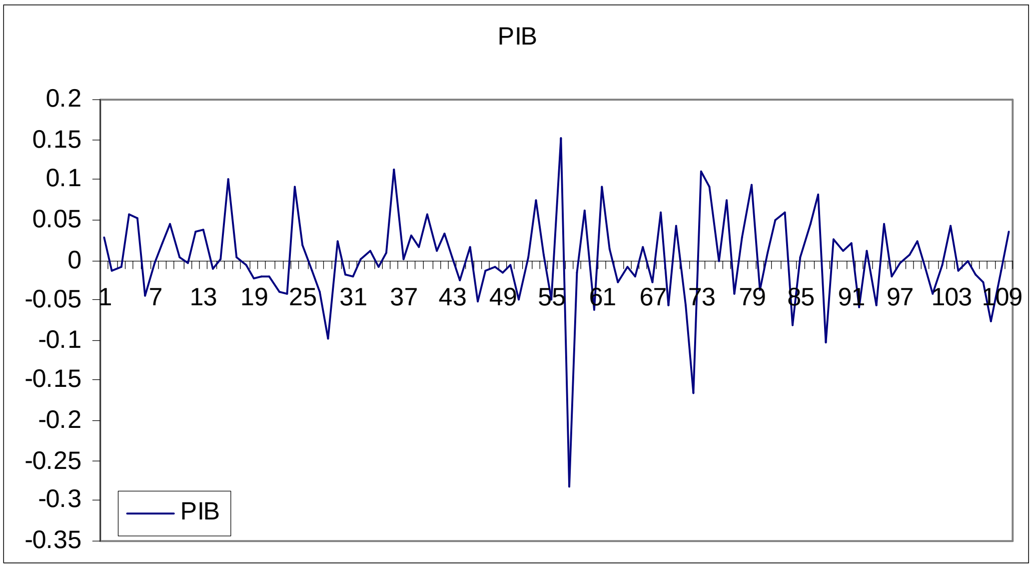
THE DATA

The quarterly sample data used in this study cover the major Brazilian macroeconomic series available from 1975.Q1 to 2003.Q3. In this study the time series to be forecast is the growth rate of Brazilian GDP[2]. There are two periods for forecast horizons. Traditional out-of-sample predictions are produced for 2002:1 to 2003:3. The 2003:4 up to 2004:3 available data were used to simulate *ex-ante* forecast.

The explanatory variables (x_t), that served to compute the diffusion index used in this work, are composed by a total of 72 national and international macroeconomic variables, including economic activity indicators of industrialized countries. These 72 series have been analyzed for unit-roots and seasonal patterns, with some usual tools such as plots and correlograms of the series, and ADF tests for unit root processes. All the nonnegative series were expressed in logs, except for the percentage scaled ones. Nominal variables in R\$ (Brazilian currency) were deflated. Seasonal adjustments were made based on the Census X-11 procedure. Moreover, first and second differences were taken to achieve stationarity when needed. After those transformations the sample started at 1975.Q4. A list of the variables and it's transformations are presented in Table A.1 in Appendix II.

The next plot give a first insight about the variable to be forecast in this study. Figure 1, shows that there are some peaks that are candidates for structural breaks occurred in the series under observation. In the beginning of 90's a price stabilization plan is the responsible for the major shift. After that, another stabilization plan called Real in 1994 can explain the expressive changes observed in the plot. In 1998 the Brazilian economy was hit by an international financial crisis and by the end of 2003, presidential election and the expectation about new directions in economic policy pushed the economy to a recession once more.

[2] $y_t = Ln(gdp_t / gdp_{t-1})$



THEORETICAL ASPECTS

DIFFUSION INDEX MODEL

A consensual point among economic models is that a good model of business cycles must reproduce some stylized facts. Burns and Mitchell (1946) present a statistical description of the cycle phenomenum. They argue that during an economic cycle there is a comovement between macroeconomic aggregate variables. Economists agree that a good business cycle model must reproduce this comovement among output, trade, exchange rate, employment, inflation, money aggregates and interest rate. But there is no agreement about what set of explanatory variables should be used to explain or forecast economic cycle.

The Diffusion Index (DI) model and its application to forecast output, following Stock and Watson(1998, 2002) is used to elaborate parsimonius models that capture the mentioned comovement. Diffusion Index model is indeed a factor model, and according to Bartholomew and Knott (1999), both are models with latent variables. This means that some variables are unobservable. Let *f* represent *r* of those variables and *x* to be *k* observable or *manifest* ones, with $r<k$. Following the ideas presented in Rummel(1970), the commom factor analysis model expresses the data matrix $X_{(Txk)}$ as a linear combination of unknown linearly independent vectors, usually called as common factors, plus a unique factor. For $i=1,...,k$ this can be represented as:

$$\begin{bmatrix} x_{1i} \\ \vdots \\ x_{Ti} \end{bmatrix} = l_{i1} \begin{bmatrix} f_{11} \\ \vdots \\ f_{T1} \end{bmatrix} + \dots + l_{ir} \begin{bmatrix} f_{1r} \\ \vdots \\ f_{Tr} \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix} \qquad (3.1)$$

Where, each individual *f* above is a factor score; each column vector with these factor scores is a common factor; each l is the factor load, and each *e* represents the above mentioned unique factor. The usual definition of common factors is that they are linear functions (of unknown) variables contributing to the common variance of the whole set of variables.

On the other hand, the unique factor contributes only to the variance of the variable that it is linked for. The unique factor is ussualy split in two components: specific variance and random errors.

Another way to express the system presented in eq(3.1) is for a given time period *T=t*, that system can be rewritten as,

$$
\begin{gathered}
x_{1t} = l_{11} f_1 + \ldots + l_{1r} f_r + e_{1t} \\
\vdots \\
x_{kt} = l_{k1} f_1 + \ldots + l_{kr} f_r + e_{kt}
\end{gathered}
$$

or simply,

$$x_t = \Lambda F_t + e_t \qquad (3.2)$$

Where, $x_t = [x_{1t}, \ldots, x_{kt}]'$ is a *(kx1)* vector, Λ is a *(kxr)* matrix of factor loadings, $F_t = [f_1, \ldots, f_r]'$ is a *(rx1)* vector, $e_t = [e_{1t}, \ldots, e_{kt}]'$ is a *(kx1)* vector of errors component, and *r<k*. Assuming that these common parts are not correlated with the unique part; and that those unique parts are not correlated across time, and that F'F=FF'=Ir, it is easy to show that[3] $\text{plim}(1/n)XX' = \Sigma = \Lambda\Lambda' + y$, and thus that $\text{var}(x_i) = \sum_{j=1}^{r} l_{ij}^2 + y_i$.

Now one can notice that not only all *k* variables in *X* matrix are represented by a linear combination of the *r* common factors plus a unique factor, producing a smaller subset of variables, but also that these *r* common factors and its factor loadings are sufficient to explain a common variance structure of all *k* variables, which by no means can be used to capture any possible commovements between these variables.

Let y_{t+1} be the series to be forecast, the class of linear models used in this work, are of the form

$$y_{t+1} = c + a(L) y_t + b(L) x_t + e_{t+1} \qquad (3.3)$$

for t=1,...,T and $a(L)$ and $b(L)$ are polynomials in the lag operator of dimension q_1 and q_2. There are $(q_1+q_2)xk$ parameters in (3.3). In applications where k is large, the

[3] Assuming that the data is well behaved to apply the Khinchine's weak law of large numbers

estimation of those parameters could be very imprecise. Furthermore, for prediction purpose parsimony may yield better Minimum Square Forecast Error. This is well described in a quotation by Clements and Hendry (1998):

"Policy analysis will often require a relatively detailed characterization of the channels of influence of the policy variables on the behavioral variables in the macroeconometric model, while a good forecasting performance may only be obtained from a model containing fewer parameters. Thus, the proprietors of large-scale models who routinely forecast and undertake policy analysis may find they require different models for each of these exercises".

Assuming that $E(e_{t+1} \mid F_t, y_t, x_t, F_{t-1}, y_{t-1}, x_{t-1,...}) = 0$, this implies that $E(e_{t+1} \mid F_t, y_t, x_t, F_{t-1}, y_{t-1}, x_{t-1,...}) = 0$ depends only on F_t. Assuming also that (y_{t+1}, x_t) has a dynamic factor representation, with $\bar{r}(\bar{r} < k)$ dynamic factors f_t, Stock and Watson(2002) redefined (3.3) as

$$y_{t+1} = c + a(L)y_t + b(L)f_t + e_{t+1} \qquad (3.4)$$

$$x_t = l(L)f_t + e_t \qquad (3.5)$$

where $l(L) = I + B_1(L) + ... + B_p(L)^p$, each B_i is a $(kx\bar{r})$ matrix and f_t is a $(\bar{r}x1)$ vector of factors. Thus, a factor model can replace the large amount of information contained in those k variables by a smaller group of *r* factors. Also, modeling all lag polynomials as having finite order of at most q, Stock and Watson developed a static representation of eq(3.5). The system in its time invariant representation is presented below.

$$y_{t+1} = c + ay_t + b'F_t + e_{t+1} \qquad (3.6)$$

$$x_t = \Lambda(L)f_t + e_t \qquad (3.7)$$

where $a = (a_0,..., a_q)'$, $F_t = (f'_t,..., f'_{t-q})'$ is a *(rx1)* vector with $r \leq (q+1)\bar{r}$, $\Lambda_i = (l_{i0},..., l_{iq})$ and $b = (b_0,..., b_q)'$. If the usual infinite lag assumption were applied, then this static representation of a dynamic factor model would have infinitely many factors. Furthermore, the main advantage of the last representation is to allow the estimation of factors by principal components, which has some advantage over dynamic factor model obtained through maximun likelihood estimation (MLE). First, principal components are simpler to calculate and allow for a bigger set of variables than MLE. Second, factors estimated by principal components are consistent as the number of variables goes to infinity, even for a fixed time period of observations for the series, and this is a good characteristic for empirical work when there is a reasonable number of variables, but just a few observations of them.

ESTIMATION, TESTING, FORECASTING AND COMBINING FORECASTS

Estimation procedure of DI Model

The estimation[4] procedure for the autoregressive diffusion index model represented by (3.6) and (3.7) is composed of two steps. First, the exact number of factor is unknown. Thus, under the hypothesis of the existence of *n(n<k)* common factors, the observed data x_t are used to estimate these factors. The static formulation of a dynamic factor model presented in (3.6) and (3.7) allows the use of principal components technique to estimate the unobservable common factors. Since principal components are very sensitive to data scaling, standardized values of x_t were used. The factors estimates $\hat{F}_t$ are the eigenvectors associated with the *n* largest eigenvalues of the standardized *(TxT)* matrix $k^{-1}\sum_{i=1}^{k} x_i x_i^{'}$, where $\underline{\mathrm{x}_{\mathrm{i}}} = (\mathrm{x}_{\mathrm{i1}},...,\mathrm{x}_{\mathrm{iT}})$ is a *(Tx1)* vector. Appendix I shows further details of the estimation procedure by principal components.

Thus, these factors estimates have the properties of the eigenvalue-eigenvector problem in principal components. This means that the first factor is the eigenvector associated with the largest eigenvalue, and it can be understood as a linear combination of observed data that explains the largest part of the variance of the data. Following this pattern, the second factor is the eigenvector associated with the second largest eigenvalue and represents a linear combination of the data which best explains the part of the variance that is not explained by the first factor, and so the other factors. Moreover, another main characteristic of the principal component solution is the rotation that guarantees that each of these factors will be linearly independent of the others, avoiding, therefore, any degree of multicolinearity that may exist between the regressors.

In the second step, y_{t+1} is regressed onto a constant, $\hat{F}_t$ and y_t to obtain estimates of $\hat{c}$, $\hat{a}$ and $\hat{b}$. This two step estimation method was adopted in Stock and Watson (1998,2002)[5].

Three types of panel sets were tried. The first panel set was made up of the current values of the 72 macroeconomic variables. The second and the third sets allowed for one and two lags, respectively, of these series. Thus, in the second stacked panel the number of columns of x_t were 144, and in the third this number jumped to 216 series.

Forecasting

The forecasting environment used in this work is based in a common practice nowadays - simulated real-time design forecasts. The simulated real-time forecasting environment, has influenced the estimation procedure also. Predictions were made in a recursive fashion. For the DI model after each forecast, the sample was updated and the model was

---

[4] A Gauss program was use to estimate the DI model, and to produce forecasts.

[5] Stock and Watson (1998) shows that the estimated factors are uniformly consistent, and that these estimates are consistent even when there is a time variation in Λ. Moreover, they also have shown that if r is unknow and even if $m \geq r$ the efficient forecast MSE can be achieved.

re-estimated, BIC\ was again computed, and another round of forecasts were produced. Thus, as the forecast period begins at 2002.Q1 the models were estimated from 1975.Q4 up to 2001.Q4 and the first period forecast was computed. Then, actual values at 2002.Q1 of these variables were included in the estimation sample, and the model and BIC for the DI model were re-estimated from 1975.Q4 up to 2002.Q1 and a forecast to $y_{\{2002:Q2\}}$ was generated. This step was repeated until the forecast of $y_{\{2003:Q3\}}$ was produced.

The general equation used for DI models, to make one step a head forecasts, is:

$$\hat{y}_{T+1|T} = \hat{c}_h + \sum_{i=1}^{q_1} \hat{a}_i y_{T-i+1} + \sum_{j=1}^{q_2} \hat{b}_j \hat{F}_{T-j+1} \qquad (3.8)$$

Where, $y_{t+1} < \ln\left(\frac{y_{t+1}}{y_t}\right)$ and $y_t < \ln\left(\frac{y_t}{y_{t-1}}\right)$. Variations of (3.8) were used to forecast. As in Stock and Watson (2002), the DI model uses only the current factor to forecast. DI-AR model is the DI model plus lags of the dependent variable $[1 \le q_1 \le 3]$. Another DI\ forecasts based on these two variations were tried. The DI-Lag allowed lags on the factors $[1 \le q_2 \le 3]$ and DI-AR-Lag models which used current and lags of the factors and lags of the dependent variable. Moreover, results of these models, where the number of factors and lags were chosen by Bayesian Information Criterion(BIC), are presented as DI-BIC, DIAR-BIC, DILAG-BIC and DIARLAG-BIC, respectively.

The number of factors in a model depends if the model has lagged factors or not. Models with lags on the latent variables used from one to three factors, while models with just current factors used up to five of these factors.

The autoregressive models (AR) were used as a benchmark for DI models' performance, and they were estimated making all $\hat{b} = 0$ in (3.8) and allowing for lags $[1 \le q_1 \le 3]$ to be set by its forecast performance.

In the next section the practical problems and results of the estimation and forecasting procedures will be presented. Also a comparison of forecast efficiency for all the models is calculated. This comparison is made up of ratios of Mean Square Forecast Error (MSFE) and plots of realized values against the predicted ones.

EMPIRICAL RESULTS

The efficiency measure of the different prediction mechanisms used in this study was the ratio of the Mean Square Forecast Error (MSFE) of AR(1)[6], to the others DI models. Table 1 shows these one step ahead forecasts ratio comparisons. The DI\ forecasts were better than the ones from AR(1) model except for DI-AR and for the DI-AR-Lag forecasts. One can see that the simplest DI model, with just one factor, could improve

---

[6] The AR(1) was the AR(p) for p=1,2,3, which produced the best forecasts among them.

almost 35% on AR(1) forecasts. Moreover, the model selection by BIC in the case of pure DI models without the autoregressive part has the same forecast efficiency as the unique fixed factor DI model. Also, allowing for factor lags does not improve on the fixed DI model.

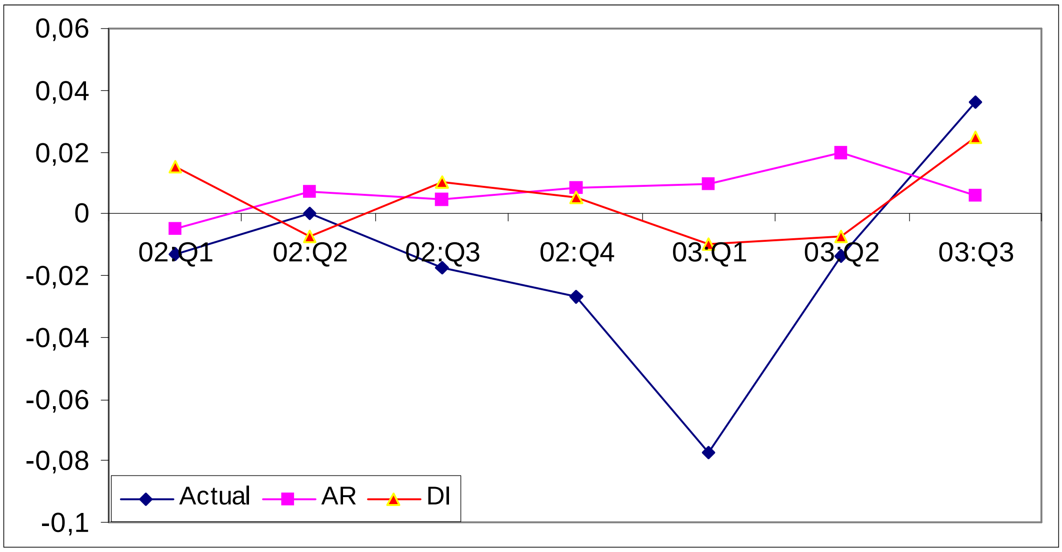
After that, two stacked panels were used to estimate the factors loadings. They include one and two lags of all the series contained in the unstacked model, respectively. The results of stacked data were not better than the results of the unstacked panel. Indeed, some of the models did worse with stacked data. A next step was to verify if a binary panel data would predict better. Thus, the positive values of the unstacked panel were set equal to one, and the negative values were set equal to zero. The results of this procedure were very similar to the original unstacked panel.

The result that only a small set of factors could be used to forecast is in tune with other recent similar studies, such as Stock and Watson (1998) and Brisson and Campbell (2003), for example. Indeed, the forecasts generated by DI models with one, two or three factors are so similar that their plots are indistinguishable; i.e., the plots become a thick line.

Based on that, all the analysis from now on will be concentrated on the fixed DI model with only one factor, because it is more parsimonious and it was chosen by BIC criterion. Figure 3 plots actual and forecast values of the AR(1) model (FAR) and of the fixed DI model (FDI) with only one factor for the growth of GDP. One can see that not only the DI model forecasts are closer to actual values, but also that it predicts changes of direction more accurately than the AR model. If the large shift at 2003.Q1, beginning in 2002.Q3 due to presidential election and the market's negative expectations about the upcoming economic policy, were included in the model, these forecasts probably would have had a better performance.

Table 1: One Step Ahead Forecasts of DI Models: 2002.Q1 to 2003.Q3

	Models	**Models**
	DI	**DI-AR**
num. fac.		
r=1	0.65	1.00
r=2	0.65	1.08
r=3	0.64	1.08
r=4	0.81	1.78
r=5	0.82	1.84
Num. fac.	**DI-LAG**	**DI-AR-LAG**
$q_2=1$	0.65	1.50
$q_2=2$	0.65	1.08
$q_2=3$	0.64	1.08
BIC	**DI-BIC**	**DIAR-BIC**
	0.65	1.00
BIC	**DILAG-BIC**	**DIARLAG-BIC**
	0.65	1.50



Intercept Corrections

Two intercept corrections (IC) techniques were tried. The first one consisted of adding the past forecast error to the actual forecast values. The second correction was to sum up the last sample error to actual forecast values. Neither of these IC approaches produced satisfactory results. Table 6 present pairwise comparisons between models in terms of their MSFE ratios. The first column shows the variable whose MSFE is used as the denominator of the ratio. The names with ic added means the models with the first intercept correction technique.

Table 6 – MSFE Ratios Comparison

Denominator	Ar	Di	Aric	Diic
Ar	1	0.65	2.48	6.51
Di	1.56	1	3.87	10.15

Table 6 shows that all the MSFE ratios of models with intercept correction are bigger than one, when compared to their respective MSFE without correction. The forecast performance of AR with IC is 148% worse than the AR without IC. This value is 915%, when this same analysis is done to the DI.

CONCLUDING REMARKS

In order to forecast the GDP growth rate, macroeconomic theory would suggest the use of a large set of financial, monetary, and others real and nominal variables to be included in a model capable to mimic some stylized facts of business cycles, such as the comovements among a set of variables, as pointed out in the first part of this work.

From the point of view of economic forecasting practice, parsimonious models have a great advantage in terms of forecast performance compared to large econometric theory based models.

This work used diffusion index models (DI) to forecast quarterly Brazilian GDP growth rate. A DI model is basically a static representation of an unobservable dynamic factor model. Both models may be used to capture the comovements between variables and to reduce, at the same time, the number of parameters in the model used to forecast.

The most important motivation behind the choice of such a static model is that the factors may be estimated by the solution of an eigenvalue and eigenvector problem similar to the problem found in principal component technique, while the dynamic factor model is estimated by MLE.

The principal component approach allows the number of variables to be bigger than the number of observations. Besides that, it also produces a factor estimator that only needs an increasing number of cross-sections to be consistent. These precious features of principal components are extremely important when one is facing a short time-series data set, as is the case in this study.

Quarterly data from 1975:1 up to 2003:3 of Brazilian GDP and another 72 macro variables representing the external sector, and the nominal and real side of the economy were used to compute the diffusion index. The estimation period ended in 2001:4 and forecasts were made from 2002:1 to 2003:3 in a recursive environment.

The results in terms of forecast performance was very encouraging. The linear DI model with only one factor model improved 35% on an autoregressive (AR) model when their MSFE were compared.

More studies using diffusion index models or other kind of indices are important because their predictions usually outperform most of time-series models, with the advantage of utilizing other exogenous variables. Thus, research efforts in this field may produce the conveyance of theoretical models with the practice of forecast in the economic science.

Finally an QTR{it}{ex-ante} forecast environment was simulated with a few sample data points. The results show that it may be better to generate predictions with forecast diffusion indices than with simple dynamic ARMA models. This is another point that deserves more study, in order to produce more useful DI models to forecast, when the one-step-ahead is the wanted forecast horizon, and only one lag of this index is used to forecast.

BIBLIOGRAPHY

BARTHOLOMEW, D.J. AND KNOTT, M. (1999): *Latent Variable Models and Factor Analysis*. New York: Oxford University Press Inc.

BURNS, A. AND W. MITCHELL (1946): *Measuring Business Cycles.* New york: National Bureau of Economic Research.

CHAUVET, M. (1998) "An Econometric Characterization of Business Cycles Dinamics with Factor Structure and Regime Switching". *International Economic Review*39(4).

______.(2001): "A Monthly Indicator of Brazilian GDP". *Brazilian Economic Journal* (Revista Brasileira de Economia) 21.

______. (2002): "The Brazilian Business and Growth Cycles". *Brazilian Economic Journal* (Revista Brasileira de Economia) 56(1).

CHAUVET, M., LIMA E.C.R. AND VASQUEZ, B. (2002) " Forecasting Brazilian Output in the Presence of Breaks: A Comparison of Linear and Nonlinear Models". Working Paper 2002-28 Series of The Federal Reserve Bank of Atlanta.

CLEMENTS, M.P. AND HENDRY, D.F. (1998): *Forecasting Economic Time Series.* United Kingdom: Cambridge University Press.

______.(2001): *Forecasting Non-Stationary Economic Time Series.* MIT Press.

CONNOR, G. AND KORAJCZYK, R.(1993): " A Test for the Number of Factors in an Approximate Factor Model". *Journal of Finance*, 48, 4,.

ENGLE, R.F. AND WATSON, M.W. (1981): "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates". *Journal of the American Statistical Association*, 76, 376, 774-781.

FORNI, M., HALLIN, M., LIPPI, M. AND REICHLIN, L. (2000): "The Generalized Dynamic Factor Model: Identification and Estimation". *The Review of Economics and Statistics*, 82,4, 540-552.

______.(1998): "Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle". *Review of Economic Studies*, 65, 453-474.

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

GEWEKE, J. AND SIGLETON, K. J. (1981): "Maximum Likelihood Confirmatory Factor Analysis of Economic Time Series". *International Economic Review*, 22, 1, 37-54.

GEWEKE, J. AND G. ZHOU (1996): "Measuring the Price Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557-587.

MOREIRA, A.R.B. AND AMENDOLA, E. (1998): " Comparação de Modelos de Previsão para o PIB e o Produto da Indústria". IPEA - Textos para Discussão 613.

MOREIRA, A.R.B., FIORENCIO, A. AND LOPES, H.F. (1996): " Um Modelo de Previsão do PIB, Inflaão e Meios de Pagamento". IPEA - Textos para Discussão 446.

RUMMEL, R.J. (1970): *Applied Factor Analysis*. Evanston: Northwestern University Press.

STOCK, J. H. AND WATSON, M.W. (1989): "New Indexes of Coincident and Leading Economic Indicators". *NBER Macroeconomics Annual*, 351-393.

______.(1991): "A Probability Model of the Coincident Economic Indicators" in Leading Economic Indicators: New Approaches and Forecasting Records, eds. K. Lahiri and G.H. Moore, New York: Cambridge University Press, 63-85.

______.(1998): " Diffusion Indexes". working paper 6702, National Bureau of Economic Research.

______.(2002): "Macroeconomic Forecasting Using Diffusion Indexes". *Journal of Business & Economic Statistics*,20, 2, 147-162.

APPENDIX I

Assume that *r* is unknown and there is *n(n<k)* common factors. Let $\underline{x_i} = (x_{i1},\ldots,x_{iT})'$ *{i=1,...,k}* to be a *(Tx1)* vector, $F=(F_1,\ldots,F_n)$ a *(Txn)* matrix, $P_F=F(F'F)^{-1}F'$ and l_{it} a *(nx1)* vector. Suppose that $l_{it} = l_{i0}$ and *F* is a unknown nonrandom matrix. The estimator for (l_{i0}, F) proposed by Stock and Watson (1998) minimize the following objective function:

$$V(l_{10},\ldots.l_{k0},F) = k^{-1}\sum_{i=1}^{k}(\underline{x_i} - F\,l_{i0})'(\underline{x_i} - F\,l_{i0}) \qquad (6.1)$$

Let $(\tilde{l}_{i0},\tilde{F})$ be the minimizers of $V(l_{i0},F)$, the first order condition with respect to l_{i0} implies that

$$\tilde{l}_{i0} = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\underline{x_i} \qquad (6.2)$$

Substituting (6.2) into (6.1), the result is the concentrate objective function

$$V(l_{10},....l_{k0},F) = k^{-1}\sum_{i=1}^{k}(\underline{x_i} - P_F\,\underline{x_i})'(\underline{x_i} - P_F\,\underline{x_i}) \qquad (6.3)$$

$$= k^{-1}\sum_{i=1}^{k}\underline{x_i}'(I_T - P_F)\underline{x_i}$$

Considering the normalization $F'F=I_n$ and that $\sum_{i=1}^{k} e_i' e_i = tr(e'e) = tr\left(k^{-1}\sum_{i=1}^{k}\underline{x_i}'(I_T - P_F)\underline{x_i}\right)$, minimizing eq(6.3) is equivalent to maximizing[7]

$$tr\left(F'\left(k^{-1}\sum_{i=1}^{k}\underline{x_i}\,\underline{x'_i}\right)F\right) \quad \text{subject to } F'F=I_n \qquad (6.4)$$

Where, $k^{-1}\sum_{i=1}^{k}\underline{x_i}\,\underline{x'_i}$ is a (TxT) symmetric matrix. Thus, the optimization problem becomes,

$$\underset{F,\,l}{Max} = \sum_{j=1}^{n} F_j' M F_j + \sum_{j=1}^{n} l_j (1 - F_j' F_j)$$

The first order condition to this problem is $[M - l_j I_T]F_j = 0$, where is l_j the eigenvalue of M and F_j is its associated eigenvector. Thus, $\sum_{j=1}^{n} F_j' M F_j = \sum_{j=1}^{n} l_j$. Adding the information that $n<T$, in order to maximize $\sum_{j=1}^{n} F_j' M F_j$ one must collect the first n largest eigenvalues and its associated eigenvectors of the *(TxT)* matrix $k^{-1}\sum_{i=1}^{k}\underline{x_i}\,\underline{x'_i}$. Thus, F is just the *(Txn)* matrix made with these associated eigenvectors.

APPENDIX II

Index of Hours Worked In Ind. Prod. of The State of Sao Paulo	*	4
Index of Industrial Production - Consumer Goods	*	4
Index of Industrial Production - Intermediate Goods	*	4
Index of Industrial Production - Capital Goods	*	4
Index of Industrial Production - Nondurable Consumer Goods	*	4
Index of Industrial Production - Durable Consumer Goods	*	4
Index of Industrial Production –Mining	*	4

[7] The result *tr(ABCD)=tr(CDAB)* was used to achieve eq(6.4)

Index of Industrial Production – Pharmaceuticals	*	4
Index of Industrial Production – General	*	4
Index of Industrial Production – Mechanics	*	4
Index of Ind. Production - Electrical and Communications Equip.	*	4
Index of Industrial Production –Metallurgy	*	4
Index of Industrial Production -Transport Equi.	*	4
Index of Industrial Production -Food Products	*	4
Index of Industrial Production -Paper and Cardboard	*	4
Index of Industrial Production –Plastics	*	4
Index of Industrial Production –Chemicals	*	4
Index of Industrial Production -	*	4
Index of Industrial Production –Textiles	*	4
Index of Ind. Prod.-Clothing, Footwear and Leather Goods	*	4
Capacity Utilization Rate-Industry-Capital Goods	*	7
Capacity Utilization Rate-Industry-Intermediate Goods	*	7
Capacity Utilization Rate-Industry-Material construction	*	7
Capacity Utilization Rate-Industry-Mean	*	7
Brazilian Direct Investment	**	0
Direct Investment	**	0
Foreign Direct Investment	*	0
Foreign Portfolio Investment	*	0
Portfolio Investment	*	0
Interest Rate-Bank Deposit Certificate (CDB)	*	1
Interest Rate Credit Operations to Short Term Private Capital	*	1
SELIC Interest Rate (Monetary Policy)	*	1
Loans of Financial System to Private Sector	*	3
Loans of Financial System to Private Sector-Habitation	*	3
M0-Monetary Aggregate	*	4
M1-Monetary Aggregate	*	4
Internal Debt	**	3
Federal Internal Mobiliary debt	*	3
Financial Execution of National Treasury Debt	*	3
Financial Execution of National Treasury Credit	*	3
Cost of Living Index of Sao Paulo	*	6
General Price Index Domestic Supply	*	2,6
INCC Price Index	*	2,6
Index of Nominal of The Retail Trade in Sao Paulo-Industry	*	4
Index of Employed People in Ind. Prod. of State of Sao Paulo	*	2
GDP of Brazil	*	5
Exports	*	2
Imports	*	4
Overall Balance of Payment Results	*	2
Exchange Rate (R$/US$)	*	3
International Reserve	*	2
IBOVESPA-Index of Stock Market-Brazil	***	3
Mundial Exports	*	2

Mundial Imports	*	4
Exports of Industrialized Countries	*	2
Imports of Industrialized Countries	*	4
GDP of Canada	*	6
GDP of China	*	4
GDP of Korea	*	4
GDP of Spain	*	4
GDP of France	*	4
GDP of Germany	*	2
GDP of Italy	*	6
GDP of Japan	*	6
GDP of United Kingdom	*	6
GDP of USA	*	4
USA Interest Rate-Federal Funds-3-month	*	1
USA Interest Rate-Treasury Maturities-10-years	*	1
USA Interest Rate-Treasury Maturities-3-years	*	1
USA Interest Rate-Prime-3-month	*	1
USA Interest Rate-Treasury Bills-3-month	*	1
USA Interest Rate-Treasury Bills-6-month	*	1

Where,

(*) Data from Ipeadata;

(**) Data from Central Bank of Brazil;

(***) Data from Economatica;

[0] Growth Rates;

[1]First Difference (1diff);

[2] Ln+1diff;

[3]Ln+Deflating+1diff;

[4] Ln+Seas. Adj.+1diff;

[5] Ln+deflating+seas.adj+1diff;

[6] Ln+Second Difference (2diff);

[7] $\Delta Ln(\frac{X_t}{100-X_t})$.