FORECASTING THE REVENUE FROM VALUE ADDED TAXES IN THE STATE OF SANTA CATARINA, BRAZIL: THE GENERAL TO SPECIFIC APPROACH IN DYNAMIC REGRESSION

Eder Daniel Corvalão, M.Sc.

CIASC – Cia. de Informática e Automação de Santa Catarina

Rod. SC 404 – Km 4, Itacorubi, Florianópolis – SC. Cep 88034-000

Tel. (0 55 48) 231 1147 E-mail: corvalao@ciasc.gov.br

Robert Wayne Samohyl, Ph.D.

UFSC – Universidade Federal de Santa Catarina

Campus universitário – Trindade, Florianópolis – SC. Cep 88040-900

Tel (0 55 48) 331 7027 E-mail: samohyl@deps.ufsc.br

Guttemberg Hespanha Brasil, Dr.

UFES – Universidade Federal do Espírito Santo

Depto. Estatística, Avenida Fernando Ferrari, s/n – Goiabeiras, Vitória – ES. Cep 290600-900

Tel (0 55 27) 335 2481 E-mail: ghbrasil@terra.com.br

FORECASTING THE REVENUE FROM VALUE ADDED TAXES IN THE STATE OF SANTA CATARINA, BRAZIL: THE GENERAL TO SPECIFIC APPROACH IN DYNAMIC REGRESSION

Abstract

This paper verifies the possibility of improving the monthly forecasts of the Value Added Tax on Sales and Services (ICMS) collected in the State of Santa Catarina, Brazil. A model of dynamic regression is used based on the concepts of cointegration and error correction models utilizing the general to specific approach suggested by the London School of Economics (LSE). Several variables were selected for the final model: among the most important are industry sales, electric energy consumption and the number of consultations with the Credit Protection Service (SPC). In choosing variables, Granger's test of causality and long-run equations were utilized. The results obtained were very satisfactory for forecasts both inside and outside the sample period, indicating that the use of this model by the Budget Department of the State of Santa Catarina will provide more suitable values for the decision making process and should lead to improvements in fiscal planning.

Keywords: Forecasting, general to specific, regression, cointegration, value added taxes, Brazil

1. Introduction

Created in 1967, the Brazilian Value Added Tax on Sales (ICM) was presented as a national tax with intra and interstate tax rates fixed by the Brazilian Senate. With the constitutional reform of 1988, the ICM had its incidence base amplified, which incorporated all preexistent taxes on retail sales and services, thereafter being called the Value Added Tax on Sales and Services (ICMS), attributing competence and autonomy to each State to fix its own tax rates, representing practically 90% of the total of all State tax revenues.

In this context forecasts become an important part of the decision making process for the State treasury. The more accurate the forecasts the better will be the administration of available resources in the State of Santa Catarina. During many years, the forecasts elaborated by the Budget Department of the State of Santa Catarina (SEF-SC) were based on a simple moving average, generating estimates with a high margin of error. For the sake of comparing the accuracy of forecasts, in this paper we employed the mean absolute percentage error (MAPE). The value of the MAPE for the moving average forecasts surpassed a monthly average rate of 12% according to Corvalão (1999).

The employment of econometric models and consequently the elaboration of forecasts has been the target of many studies in the last two decades. Fildes (1985) lays out an extensive exposition of the theme and comments on the conclusions from Armstrong (1985) which compares the performance of forecasts of econometric and extrapolative models, strongly suggesting that, in general, results from econometric models tend to be better than extrapolative methods in both short- and long-term horizons. Hendry et al. (1984 p. 1043) argue that econometric models should in principal adjust better to the data than ARIMA or exponential univariate models, and when they do not then they are probably mis-specified.

Aside from the benefits for budget planning which are derived from accurate revenue forecasts, the other important point to stress is the probable gain to the State in terms of purchasing policies, due to the fact that payments by the State to its suppliers frequently run late and suffer monetary penalties, motivated by errors in forecasts. Technicians at the State treasury department estimate that costs quoted in public contracts are automatically adjusted higher by about 10% in order to compensate for the inevitable delays in payments to suppliers. With better planning, based on more accurate forecasts, and consequently an improvement in the capacity to honor its commitments, the cost associated with the risk of non-payment will decrease. It has been estimated that this risk premium is more than one million dollars per year.

2. Dynamic models and the general to specific approach

Econometric models represent systems of relationships between variables of interest, where the relationships are estimated starting from the available data. FILDES, (1985) makes the point that econometric models are only one of many different tools both quantitative and qualitative for characterizing the economy or any behavioral system for that matter.

What is necessary to make an econometric analysis? On the one hand we need well founded economic theory and sources of reliable statistical data, and on the other a methodology for selecting precisely the relevant candidates for explanatory variables and the application of statistical procedures to uncover the important relationships. The general to specific approach in econometrics follows this reasoning.

2.1 General to specific approach

Developed in accordance with the econometric tradition of LSE (*London School of Economics)*, starting from the research of Sargan in the 1960´s, Davidson, Mizon and principally through innumerous applications published by Hendry beginning in the 1970´s, applied econometrics has had an enormous influence on the forecasting profession. For more details, see Hendry et al. (1984) and Charemza and Deadman (1997).

The general to specific approach is an approach based on a process of successive reduction of a general econometric model, beginning with a general dynamic statistical model, which captures the essential characteristics of the data set. The starting point should be a dynamic model with many variables and with a high order of time lags. Tests are used to reduce the complexity of the model, eliminating variables that are statistically not significant, checking the validity of the reduction at each stage to guarantee the congruency of the model selected. The simplifications of the general model are conducted through a series of transformations and reductions.

During the successive simplifications of the model, the estimated models should attend the following criteria, suggested by Gilbert (1986):

- Coherency of the data, observing if the the model fulfills the basic supositions such as no autocorrelation in the residuals, homoscedasticity, and so on;
- validity of conditioning in the sense that explanatory variables should be strictly exogenous;
- Exibit constancy in parameters;
- criteria of admissibility, since the estimated coeficients should make sense. For example, that values obtained for elasticities not be extreme;
- consistency with theory, review of the signs, the magnitude of the coefficients, observing if the estimated values are congruent with the postulated theory and, in general, with economic theory.
- a model only will be considered adequate if it encompasses the results of the rival models.

The approach of Hendry is based on the concept of the data generating process (DGP), which represents a totally general view of the distribution of all variables (CUTHBERTSON et al., 1992). Modeling in econometrics for Hendry consists therefore in simplifying this DGP in such a manner that it has an estimatable form. The process of simplification is with the following steps: marginalization, conditioning, reparameterization, estimation and diagnosis; and Pagan (1990) suggests that the steps of the approach should be preceded by the analysis of the order of integration (and co-integration) of the variables studied.

2.2 Process of reduction

After the selection of the most general model (with many lags) research for an appropriate (parsimonious) model proceeds through a simplification of the general model based on the evidence contained in the data through the appropriate tests.

However, the number of lags to be incorporated in the model should be dealt with.

The strategy employed follows the steps in Samohyl (2000): The procedure includes initially a large number of lags in order to capture the richness of potential dynamic behavior among variables, followed by reducing the model gradually through a test process of restrictions on the parameters in the general model, and imposing the restrictions which can not be rejected in statistical terms. In the process of step by step reduction the criteria of Schwarz (SC) and Hannan-Quinn (HQ) are used. These scaled measures serve to choose between alternative models in the process of reduction, arriving at the most appropriate model.

3. Empirical analysis: one model of forecasting for the ICMS/SC

The monthly data for value added tax revenue in Santa Catarina were obtained from the State Treasury Department (SEF-SC) for the period December 1995 to December 2001. For the choice of variables, considering the elaboration of the econometric model, initially the composition of the largest economic sectors which participated in the formation of tax revenue was analyzed in the period from 1995 to 2001.

- INDU - Monthly sales of Santa Catarina industry, elaborated by the Federation of Industry in Santa Catarina, in 220 industries, aggregating all business sectors;
- ELET – Electric energy consumption, data from the Statewide Electrical Energy Distributor;

- GASO – Fuel consumption: automotive gasoline, data from the National Petroleum Agency (ANP)
- OLEO – Fuel consumption: diesel oil, data from the National Petroleum Agency (ANP);
- CIME - Apparent cement consumption (a small fraction of this cement is in inventories) in Santa Catarina, made available by the National Cement Industry Union (SNIC);
- SPC – Consultations with the credit protection service. The total quantity of consultations made by merchants to verify the solvency of clients indicate an increase or decrease of potential sales. Data made available by the Credit Protection Service of Florianópolis.
- INA – Indicator of the Level of Economic Activity combines physical production industry data from the State of São Paulo, hours worked and also sales elaborated by the Federation of Industry of São Paulo (FIESP).

Figures 1 and 2 present the evolution of the time series in the period of analysis. All the variables were transformed in natural logarithms seeking to make the time series more homogeneous, and using the letter L to identify the variables in the formulation of the model, thus, Lina refers to the variable of INA in logarithms.

3.1 Marginalization and conditioning

In the marginalization process, those variables in study which will participate in the model should be evaluated, verifying the significance of the estimated parameters and the magnitude of the coefficients. In an initial analysis it was verified that the variables: gasoline consumption (Lgaso), Diesel oil consumption (Loleo) and the activity level indicator(Lina) did not present the expected signs in their coefficients, given that it was expected that all the variables would participate positively in the increase of the collection of ICMS. The tests of integration of the series suggested the non-stationarity of the variables level.

Before discarding any of the variables the test of causality, in Granger's sense, can be used, which assures that all the explanatory variables which participated in the model are at least highly exogenous, a necessary condition for the model to be used for forecasts, whose results discarded the use of the variables: diesel oil consumption (Loleo) and apparent cement consumption (Lcime). After this analysis three variables remain: Lindu, Lelet and Lspc.

3.2 Results of the co-intergration test

Following the two-step procedure of Engle and Granger (1987) to verify the existence of cointegration among the series, the equilibrium equation (long-run) should be estimated. Initially, estimate the long-run equation in the variables level. With the utilization of the PcGive software, we conclude that the variables are co-integrated (same order of integration and stationary residuals), obtaining the long-run equation (1):

$$\boldsymbol{Licms_t = - 4{,}6918 + 0{,}62677\, Lindu_t + 0{,}66237 Lelet_t + 0{,}12463\, Lspc_t} \qquad (1)$$

3.3 Re-parametrization and estimation

According to Engle and Granger (1987), if the series are co-integrated, then it will be possible to re-parameterize the former model (1) to the model of the short–term incorporating an error correction mechanism (ECM). For the estimation of the short-term model, lagged variables were used and the residuals of the co-integrated equation were added, denoted here by ECM. Initially, by being monthly data, 12 lags for the explanatory variables and 3 lags for the error correction mechanism were employed.

In the process of reduction the method of elimination of variables was employed based on the significance of the statistics t and F, aiming to minimize the criteria of Schawrz (SC) and Hannan-Quinn (HQ). The most parsimonious model has its results exhibited in table 1. As required by the error correction mechanism, if equation 1 is a vector of valid co-integration, the sum of the coefficients of the error correction terms (ECM_{t-1} and ECM_{t-2}) is negative and significant. For the effects of forecasting, the short-run equation (2) below will be employed.

$$\boldsymbol{DLicms_t = 0{,}014874 - 0{,}35478DLicms_{t-2} + 0{,}15467DLicms_{t-7} - 0{,}27831DLindu_{t-12} - 0{,}68609DLelet_{t-6} + 0{,}34484DLelet_{t-10} + 0{,}087971DLspc_{t-5} - 0{,}20338DLspc_{t-12} - 0{,}62955ECM_{t-1} + 0{,}46128ECM_{t-2}} \quad (2)$$

The results of the model above show that, in the short-run, the variables: Industry revenues, electric energy consumption and the number of consultations with SPC explain 72% of the variations in the collection of ICMS in Santa Catarina (R^2 = 0.72, obtained in the formulation of the equation in PcGive).

Since more than one lag in the error correction mechanism was found, the coefficient of error corrections is the result of the sum of the coefficients (in this case: - 0.62955 and + 0.46128); with this, the value found for the coefficient of error correction was - 0.16827, being significant at the level of 1%, and presenting a compatible sign with theory. The error correction mechanism indicates that, on average, about 17% of the changes in the collection of ICMS in the current period are due to the short run disequilibrium in two subsequent periods. The values of the parameters of the explanatory variables are statistically significant and the value of the Durbin-Watson test (DW) discards the presence of autocorrelation in the residuals.

3.4 Specification tests

In accordance with the results obtained above , the model appears to be well specified and presents constant coefficients throughout the sampled units judging by the statistical tests, which do not indicate any specification problems , and in Figure 3, where the estimated out-of-sample forecasts appear to be always within the confidence interval of 95%.

Model 2 presents all regressors as statistically significant at levels of 1% or 5%. Figure 4 shows the residuals as a function of time and the correlogram of the residuals up to 13ª lag, and both graphs and tests point to a well specified model.

3.5 Forecasts and observed values.

In order to test the predictability of the model, the values of the last year of the dataset were compared to the predicted values. The Mean Absolute Percentage Error (MAPE) as a criterion of accuracy of the model was employed. Table 2 presents the observed values for tax revenues as well as the forecasts from the model estimated here and the ARIMA model presently in use by the State government.

- Model in use today - mean absolute percentage error = 4.63%
- Model 2, from this paper - mean absolute percentage error = 2.51%

The values obtained by the dynamic model with the incorporation of the ECM for the forecast within the sample had a better adjustment than the values obtained from the present model employed by the State government. One of the advantages of model 2 is the fact that all of the explanatory variables use at least one period lagged values facilitating one step ahead forecasts.

To verify the precision of model 2 and its eventual applicability for the State government´s planning process, estimates of the tax revenues for a period of 4 months beyond the sample period, January to April, were calculated. For the months of February, March and April in which the model demanded values outside of the range of the dataset for the error correction terms ECM_{t-1} and ECM_{t-2}, an ARIMA model was employed to project the respective values. In the case of the variable $Dlicms_{t-2}$ for the months of March and April the forecasts generated in this estimate were used.

Table 3 presents the observed values and forecasts for the first four months of the year (MAPE = 2.55%). From the results obtained in the first 4 months of the year, the excellent quality of the model can be verified. Two points should be emphasized: (1) that the one step ahead forecast for the month of January, when the values of all the exogenous variables for the model were known, resulted in an error of only 0.17%, and (2) the model is actually quite simple requiring only 3 explanatory variables (consumption of electric energy, industry revenues and consultations with SPC).

4. Final considerations

Following the econometric methodology developed by the London School of Economics, also known as the general to specific approach, a model for the long-term was constructed and afterwards a model was estimated for the short-term with the incorporation of an error correction mechanism, to be employed in the forecast of tax revenues in the Brazilian State of Santa Catarina.

The forecasts for the State of Santa Catarina from the model presented here produced results significantly better than those obtained with the model presently employed (ARIMA) by the State government, according to a percentage error criterion.

Short-term forecasts, up to one year, have been implemented in an electronic spreadsheet using the regression equation developed here which includes the automatic updating of the dataset and a CUSUM procedure for tracking results and alerting to the need for periodic revisions of the equation when relatively large errors appear.

5. Acknowledgments

This project was supported by the Núcleo de Normalização e Qualimetria (NNQ – Laboratory for Normalization and Quality Metrics) in the graduate program in Industrial Engineering at the Federal University of S. Catarina http://www.qualimetria.ufsc.br, and by the Companhia de Informática e Automação (Company for Informatics and Automation) of the State of Santa Catarina (CIASC) – http://www.ciasc.gov.br.

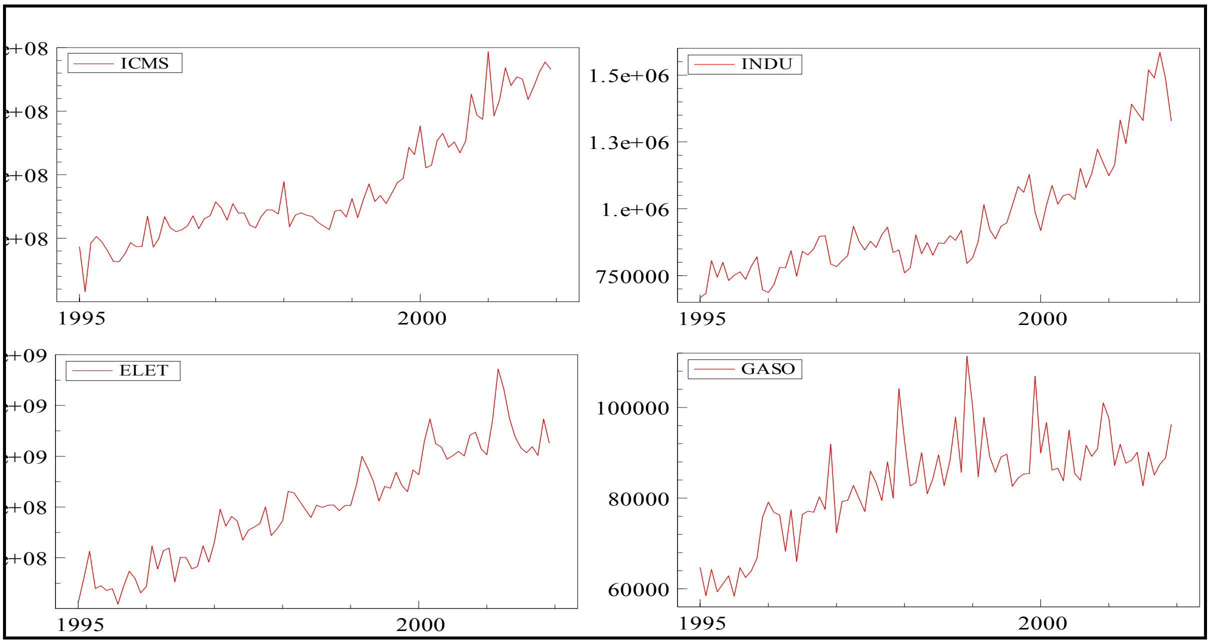

Figure. 1 – ICMS, INDU, ELET, GASO. Period: Jan.1995 to Dec.2001

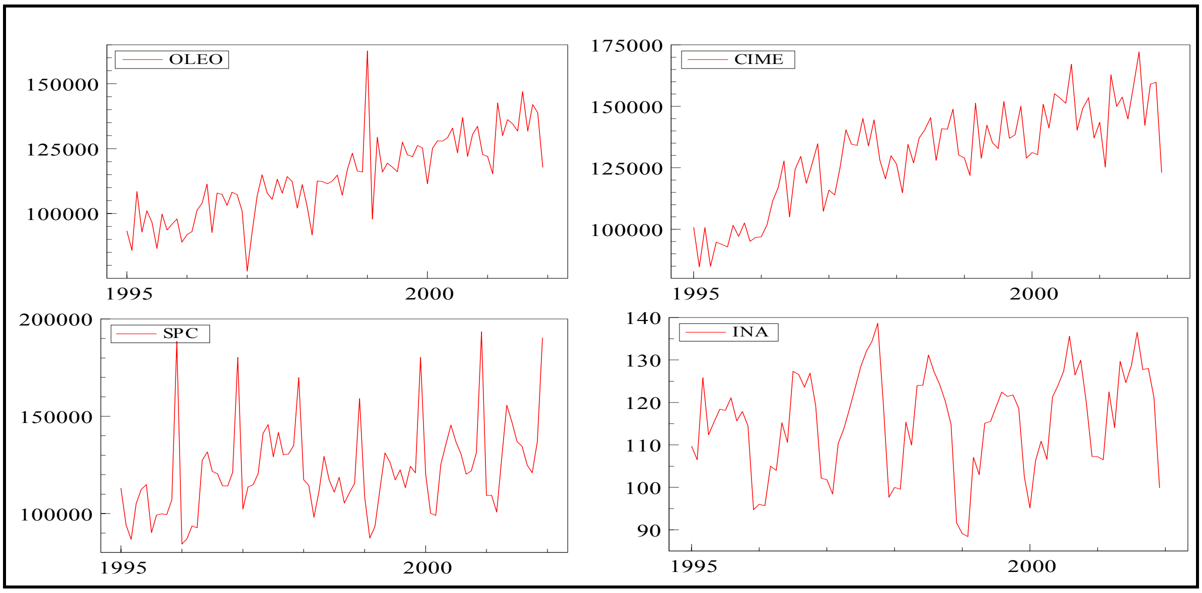

Figure. 2 – OLEO, CIME, SPC, INA. Períod: Jan.1995 to Dec.2001

Variables	Coeficient	Standard Error	t-value	t-prob	Part-R^2
Constant	0,014	0,005	2,970	0,004	0,126
DLicms_2	-0,354	0,087	-4,049	0,000	0,211
DLicms_7	0,154	0,070	2,207	0,031	0,073
DLindu_12	-0,278	0,083	-3,331	0,001	0,153
DLelet_6	-0,686	0,163	-4,189	0,000	0,223
DLelet_10	0,344	0,154	2,225	0,029	0,075
DLspc_5	0,087	0,024	3,529	0,000	0,169
DLspc_12	-0,203	0,030	-6,769	0,000	0,428
ECM_1	-0,629	0,062	-10,076	0,000	0,624
ECM_2	0,461	0,079	5,827	0,000	0,357

DW = 2,24 R^2 = 0,723 RESET F(1,60) = 12198 [0,273]

ARCH 5 F(5,51) = 0,154 [0,977] AR 1-5 F(5,56) = 1,546 [0,190]

NORM $\chi^2_{(1)}$ F(18,42) = 0,437 [0,969]

Obs.: Results from the software PcGive 8.0.

1) The test of normality is based on Doornik e Hansen (1994).

Table 1 – Modelling Dlicms by OLS

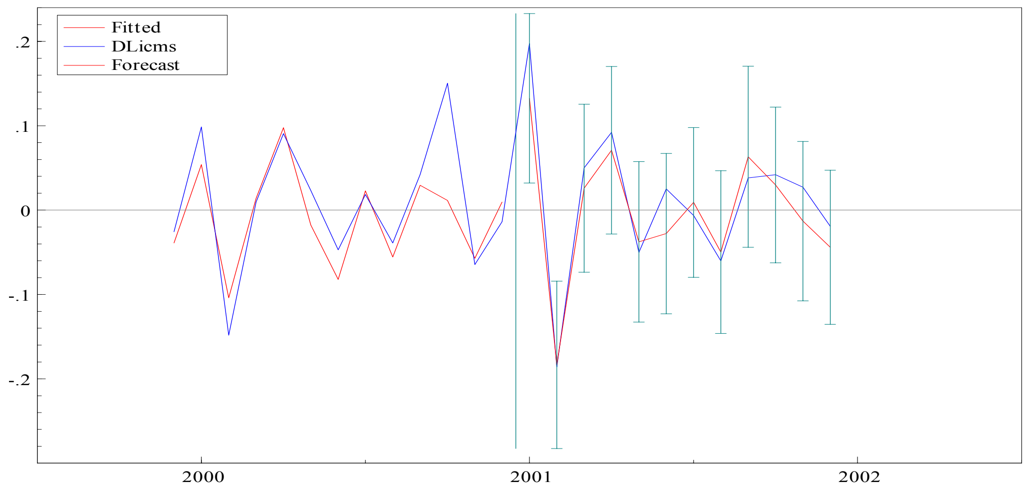

Figure 3 – Ajustament of the series Dlicms and forecasts for the year of 2001

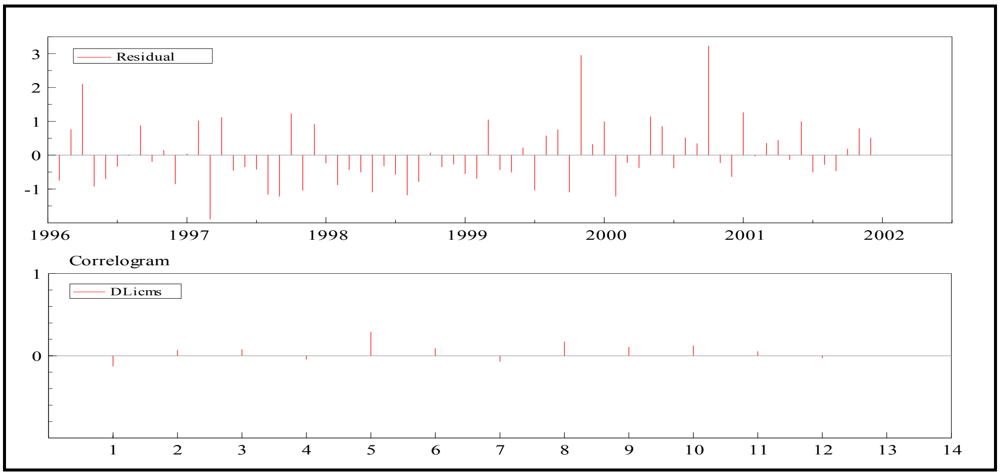

Figure. 4 – Residual analysis

Months	ICMS	Arima(1, 0, 2)[1]	Model 2
January	296,779,501.00	249,443,082.00	278,158,874.70
February	246,408,183.20	251,634,140.53	247,045,725.30
March	259,093,614.40	254,175,231.24	252,886,669.60
April	284,117,942.40	256,741,982.79	278,146,135.30
May	270,207,300.00	259,334,654.30	273,633,314.50
June	277,075,336.50	261,953,507.53	262,798,280.10
July	275,307,716.70	264,598,806.87	279,624,002.40
August	259,249,117.20	267,270,819.38	261,957,359.90
September	269,344,018.60	269,969,814.82	276,180,749.80
October	280,925,481.00	272,696,065.66	277,493,758.80
November	288,671,522.70	275,449,847.16	277,288,128.90
December	283,210,218.10	278,231,437.32	276,223,754.50
MAPE[2]		4.630	2.519

Obs: [1] Values obtained from the software eViews

[2] Calculations done in MS-Excel

Table 2 – ICMS observed and forecasted values for year 2001

Months	ICMS	Model 2
January	330,183,237.09	329,608,859.10
February	289,134,774.00	279,510,515.53
March	277,327,570.99	282,169,664.21
April	329,107,569.98	312,735,007.10
EPAM[1]		2.555

Obs: Calculations done in MS-Excel

Table 3 – Tax revenues, observed and forecasted values.

References

CHAREMZA, W.W. e DEADMAN, D.F. (1997). *New Directions in Econometric Practice*, Edward Elgar, Cheltenham.

CORVALÃO, E. D. (1999). *Estudo comparativo de modelos de previsão aplicados à arrecadação do ICMS no Estado de Santa Catarina*. Universidade Federal de Santa Catarina, Under-graduate Dissertation, Department of Administration.

CUTHBERTSON, K., HALL, S. G., TAYLOR, M.P. (1992). *Applied econometric techniques*, Ann Arbor: University of Michigan Press.

ENGLE, R.F, e GRANGER, C.W.J. (1987). Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, v. 55, 251-276.

FILDES, Robert. (1985) Quantitative Forecasting – the State of the Art: Econometric Models, *Journal of the Operational Research Society*, v. 36, n. 7, 549-580.

GILBERT, C. L. (1986). Professor Hendry's econometric methodology. *Oxford Bulletin of Economics and Statistics*, v. 48, 283-307.

HENDRY, DAVID F., PAGAN, A.R. e SARGAN, J.D. (1984). Dynamic Specifications, Ch. 8 in Z. Griliches e M.D. Intrilligator (eds.), *The Handbook of Econometrics*, North Holland, Amsterdam.

PAGAN, A. (1990). Three Econometric Methodologies: A Critical Appraisal, em Granger, C.WJ. (ed.), *Modeling economic series*, Oxford University Press.

SAMOHYL, R. W. (2000) Regressão dinâmica. Mimeographed, in Portuguese.

BIOGRAPHICAL SKETCHES:

ROBERT WAYNE SAMOHYL, PH.D:

Professor Bob is from Houston, Texas, Ph.D. in Economics from Rice University, and has lived in Brazil permanently since 1978. He is full professor and recent Chairman of the Industrial Engineering Department, Federal University of Santa Catarina, Florianopolis, Southern Brazil, a region known for a thriving software industry and wonderful beaches. He is recent Vice President of the Brazilian Operations Research Society. He has developed courses for the internet in Statistics, Econometrics and Forecasting, and taught by video-conferencing. He is presently coordinating a government project for implementing forecasting techniques in electrical energy companies. His most recent publication is "Multivariate feed back control: an application in a productive process." Computers and Industrial Engineering (Elsevier), 46 (2004) 837-850.

EDER DANIEL CORVALÃO, M.SC.

Eder Daniel CORVALÃO is a systems developer at Center of Informatics and Automation in Santa Catarina (CIASC), specialized in web development. He received a B.S. degree in Management Science and M.S. degree in Industrial Engineering at the Federal University of Santa Catarina (UFSC), Brazil, and is currently a doctoral student in Industrial Engineering. His research interests are in the areas of systems development, time series modeling and forecasting, statistical pattern recognition and clustering.

GUTTEMBERG HESPANHA BRASIL, DR.

Gutemberg H. Brasil has a MSc. degree in industrial engineering, and his doctorate in Bayesian forecasting from the Catholic University (PUC) in Rio de Janeiro, Brazil. Currently he is a Professor in Statistics at the State University of Espírito Santo (UFES), Brazil. His research interests include time series analysis, econometrics, operations research and Bayesian forecasting applied to areas such as economics and marketing. He has published in the Journal of Forecasting, Estadística, and Portuguese and Latin-American Journals.