Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

Forecasting the Collection of the State Value Added Tax (Icms) in Santa Catarina: the General to Specific Approach in Regression Analysis

Eder Daniel Corvalão[a]

Robert Wayne Samohyl[a]
[a] Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil

Gutemberg Hespanha Brasil[b]
[b] Universidade Federal do Espírito Santo (UFES), Vitória, Brazil

Abstract

In this paper was verified the possibility of improving the monthly forecasts of the Value Added Tax on Merchandise and Services (ICMS in Portuguese: *Imposto sobre Circulação de Mercadorias e Serviços*) collected by the State of Santa Catarina, Brazil. Dynamic regression will be used based on the concepts of cointegration and error correction utilizing the general to specific approach suggested by the London School of Economics (LSE). Different data series were selected and analyzed for the final model industry profit, consumption of electric energy and other energy sources, and cement, and business consultations to the Credit Service Protection Agency (SPC). In the process of the choice of the variables, Granger's tests of causality and the analysis of long-run equations were used. 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 improvement in budget planning.

Keywords: *Forecasting, General to specific, Value added taxes*

Introduction

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

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

In this context the utilization of forecasts is a necessary part of the decision making process. The more accurate the forecasts the better will be the administration of available resources. In the past, forecasts elaborated by the Treasury Department of the State of Santa Catarina (SEF-SC), were based on simple methods of moving averages, generating estimates with a high margin of error. For comparing methods and time periods in this work was employed the MAPE (mean absolute percentage error). The value of MAPE surpassed 12% for the annual forecasts in the 1990's according to Corvalão (1999).

The employment of econometric models based on regression time series analysis for the elaboration of forecasts has been the target of many studies in the last decades. For a brief introduction to time series regression see Samohyl (2009) chapter 15. Fildes (1985, p.28) brings an extensive exposition of the theme and comment the conclusions of work by Armstrong (1985) which compares the performance of the forecasts of regression and extrapolative models: "put together, their results show that the econometric methods are more accurate than the extrapolative methods, whether for the short or long-term". According to Hendry *et al.* (1984 p. 1043): "econometric models which do not fit better than univariate time-series have at least mis-specified dynamics, and if they do not forecast 'better' must be highly suspect for policy analysis".

The procedures suggested here utilize the concepts of cointegration and error correction mechanism. While these procedures are common tools to Econometricians and Statisticians, their applications in Engineering are still very rare.

Aside from the benefits for budget planning which are derived from accurate revenue forecasts, the other important point to stress is the probable gain that the State will have in the price of future purchases. Payments from the State to its suppliers frequently run late motivated by errors in the collection forecasts. The technicians within the state treasury estimate that the prices offered in public contract bidding are higher by about 10% due to the predicted delay in state reimbursements. 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 is expected to decrease.

Dynamic Models and the General to Specific Approach

Econometric regression models usually incorporate systems of relationships between variables, where the relationships are estimated starting from the available data. "Econometric models are only one of several different forms to characterize the economy or a behavioral system" (FILDES, 1985).

What are the necessary ingredients to make an econometric regression analysis?

a) Theory;

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

b) statistical data;

c) some method which permits expressing the theory in terms of the statistical data;

d) a methodology which states how to apply the method of estimation to the statistical data, and how to evaluate if the procedure was successful.

General to Specific Approach

Developed in accordance with the econometric tradition of the LSE (London School of Economics), starting with Sargan in the 1960´s, and later by Davidson, Mizon and principally through innumerous articles published by Hendry beginning in the 1970's, applied econometrics regression analysis has obtained success in diminishing forecast error in many fields of study. For more details on the method and its history, see Hendry *et al.* (1984) and Charemza and Deadman (1997).

The LSE approach is based on a process of successive reductions applied to a general econometric model, beginning with a generalized dynamic statistical model which captures the essential characteristics of the data group. The starting point is a generalized dynamic model with a high lag order in all variables. Tests are used to reduce the model complexity, eliminating variables whose coefficients are statistically insignificant, and checking the validity of the reduction process 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 simplifications, the estimated models should attend the following criteria, designated by Gilbert (1986):

a) Coherency of the data, observing if the model fulfills the basic statistical suppositions such as non autocorrelation in the residuals, homoscedasticity, *etc.*;

b) validity of conditioning between dependent and independent variables, during the simplifications of the explanatory variables which should present exogenous characteristics;

c) exhibit parameter constancy;

d) criteria of admissibility, since the estimated values should make sense. For example, that values obtained for the elasticity are not extreme;

e) consistency with theory, review of signs and magnitude of estimated coefficients, observing if the estimated values are congruent with the postulated theory;

f) a model will be considered adequate only if it encompasses the results of all rival models.

The approach of Hendry is based on the concept of a data generating process

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

(DGP), "which represents a totally general view of the distribution of all variables" (CUTHBERTSON *et al.,* 1992). Econometrics modeling for Hendry consists in simplifying this DGP in such a manner that it has an estimable form. "Rigorously tested models, which adequately describe the available data, encompass previous findings and were derived from well based theories would greatly enhance any claim to be scientific" (Hendry 1980).

Summarizing, the reduction process uses the following steps: marginalization, conditioning, re-parameterization, estimation and diagnosis. In later work, Pagan (1990) suggests that the approach should be accompanied by the analysis of the order of integration (and co-integration) of the variables studied.

Cointegration and Error Correction Mechanism

The cointegration between two variables indicates that their means move together, maintaining an equilibrium relationship in the long run. If the variables that make up the model are cointegrated, then they are described by a stable long-run relationship.

A set of variables is in long-run equilibrium when:

$$\beta_1 X_{1t} + \beta_2 X_{1t} + \ldots\ldots + \beta_n X_{nt} = 0 \tag{1}$$

or as a vector:

$$\beta' X_t = 0 \tag{2}$$

the deviation e_t from long-run equilibrium can be expressed as:

$$\varepsilon_t = \beta' X_t \tag{3}$$

If the equilibrium relationship is significant, the error must be stationary. The components of the X_t vector are called cointegrated in order d, b and can be denoted by $X_t \sim CI(d,b)$, if:

1) All components of X_t are integrated by order d, where d are consecutive differences of a non-stationary variable will bring it to stationarity.
2) There exists a vector β whose linear combination βX_t is integrated at order (d – b), where b > 0. The β vector is called a cointegrating vector.

If cointegration is found in the series, the loss of information in long-run equilibrium can be recaptured by including an error correction term in the regression equation, a procedure which will be applied to the ICMS regression model in the next section. The Error Correction Mechanism (ECM), which denotes a proportion of the imbalance of one period, is included in the regression model and acts as a correction for the next period.

The existence of a cointegrating vector allow the estimated ECM to be

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

used in the short-run equation, which in turn permits the estimation of short-run forecasts that are consistent with the long-run equations derived from theory. In other words, as Engle and Granger (1987) have shown, if two time series are *I (1)* and are cointegrated, there is an ECM between them such that:

$$DYt = Yt - Yt\text{-}1 = \beta\ DXt - \gamma\ (Yt - Xt\text{-}1) + \varepsilon t \qquad (4)$$

where ε_t is $N(0,\ s^2)$. Engle and Granger (1987) have suggested a two-stage procedure for estimating the cointegration equation and the error correction mechanism. This specification is easily augmented to two or more independent variables. In the first stage was tested for the integration order of the various time series along with, and for the presence of cointegration among the variables. It was also estimated the regression equation for long-run equilibrium, the regression equation with no lags. In the second stage once the existence of cointegration among the time series is confirmed, the residuals of the cointegration regression, which by definition are stationary *I (0)*, are used as a component of the error correction model. All the variables expressed as ECM are stationary *I (0).* Given these conditions, the error correction model is a valid description of the dynamics of short-run.

According to Engle, Granger and Hallman (1989): "The incorporation of ECM has been a factor that significantly improved the properties of forecasting".

Process of Reduction

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

The procedure includes a great number of lags which include the full dynamics of the model, followed by reducing the model gradually through a test process of restrictions in 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 Schwarz criteria (SC) and Hannan-Quinn (HQ) is used. These scaled measures serve to choose between other alternative models in the process of reduction, arriving at the most recommended model or models.

Test for Poor Model Specification

After each reduction, several tests are used to check for specification problems. In this study was used the software *PcGive* in version 8.0 that provides the following tests:

- Autocorrelation (AR 1-5) - test for residual serial correlation through an auxiliary regression of residuals in the original variables and lagged residuals. Under the null hypothesis of no autocorrelation, the value $F_{calculated} < F_{critical}$ in usual levels of significance;
- Autoregressive conditional heteroscedasticity (ARCH) - checks if the

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

squared residuals ε^2_t of model 4, are dependent on $\varepsilon^2_{t-1}, \varepsilon^2_{t-2}, \ldots, \varepsilon^2_{t-k}$ residuals. A null hypothesis of independent squared residuals, will not be rejected when the value $F_{calculated} < F_{critical}$ for appropriate levels of significance;

- The test for normality in the residual distribution is distributed as a chi-square with two degrees of freedom. The test of normality is based on Doornik and Hansen (1994).
- Poor-specification in the functional form (RESET). Under the null hypothesis there is no error in the specification of functional form. The null is not rejected if $F_{calculated} < F_{critical}$ for usual levels of significance.

For more details of the tests involved and their use in the software, see Doornik and Hendry (1994).

Encompassing

To encompass rival models means that the rival models do not contain any information that could be useful to improve the model actually chosen. Each empirical model that is appropriate (congruent) as a candidate for selection must be a reduction of the same DGP.

To select the most appropriate model, researchers should run tests to find the more parsimonious encompassing model. “A model *M1* can be said to cover another model *M2* if it can explain the results of the latest model” (CUTHBERTSON *et al.*, 1992).

The concept of “encompassing” provides a basis for a progressive strategy, where any new comprehensive model contributes something new to the explanation of the phenomenon. “Encompassing ensures not only that a model based on those insights adds to the existing knowledge about the phenomenon being modeled, but also that it does not neglect existing knowledge” (ERICSSON *et al.,* 1990).

Empirical Analysis: Forecasting ICMS/SC

The ICMS tax collection data for Santa Catarina were obtained from the State Treasury Department (SEF-SC), for the period December 1995 to December 2001. For the choice of the variables of interest, in line with the elaboration of the model, initially the composition of the largest economic sectors were analyzed.

INDU - Monthly revenues of all state industry, elaborated by the Federation of Industry in Santa Catarina (Federação das Indústrias do Estado de Santa Catarina FIESC), in 220 industries, aggregating all sectors;

ELET – Electric energy consumption, data: Electrical Energy Company of Santa Catarina (Centrais Elétricas de Santa Catarina CELESC);

GASO – Fuel consumption: gasoline, data: National Petroleum Agency (ANP)

OLEO – Fuel consumption: diesel oil, data: National Petroleum Agency (ANP);

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

CIME - Cement consumption (including 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 indicates an increase or decrease in potential sales. Data made available by the Credit Protection Service of Florianópolis.

INA – Activity Level Indicator combines physical production industry data of the state of São Paulo, hours worked and also sales elaborated by: Federation of Industry in the State of Sao Paulo (FIESP).

Figures 1 and 2 present the evolution of the series in the period of analysis. All the variables were transformed in 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.

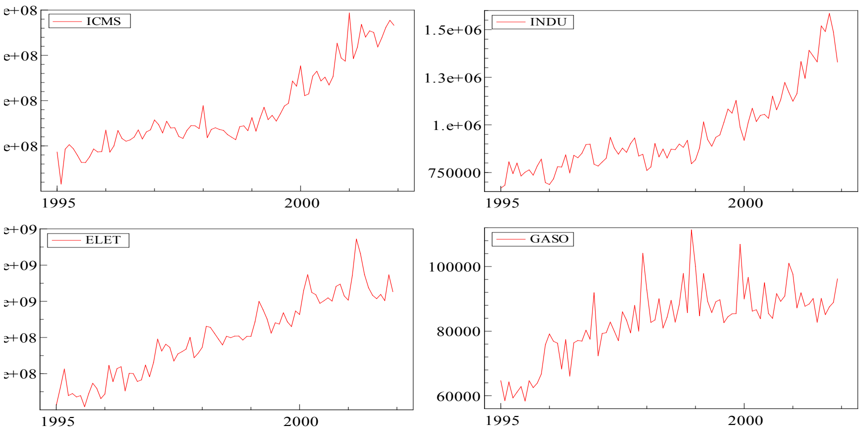

Figure 1 – Behavior of the series ICMS, INDU, ELET, GASO. Period: Jan.1995 to Dec.2001

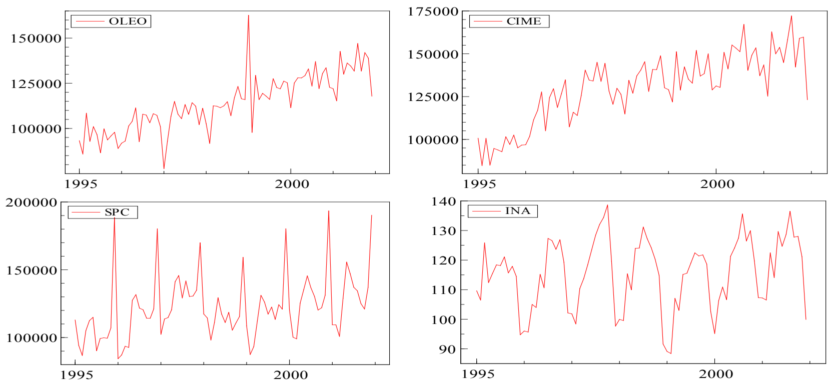

Figure 2 – Behavior of the series OLEO, CIME, SPC, INA. Períod: Jan.1995 to Dec.2001

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

Testing for Orders of Integration

Because the order of integration is critical in general to specific approach, the results of augmented Dickey-Fuller (ADF) unit roots tests are shown in Table 1.

The results from the unit root tests show that cannot be rejected the null hypothesis that the variables are integrated of degree one for all variables. Using this first differencing, Figures 3 and 4 presents the evolution of the series in the period of analysis. *D* indicates first-order difference.

Table 1 – Augmented Dickey-Fuller (ADF) unit root tests on level and first difference variables1.

Level	Test Statictic	Lag Order[2]	t-Prob
Licms	$T_{adf} = 1{,}5648$	1	0,0000
Lindu	$T_{adf} = 2{,}2435$	12	0,0038
Lelet	$T_{adf} = 4{,}4892$	10	0,0001
Lgaso	$T_{ad\,f} = 0{,}3865$	1	0,0000
Loleo	$T_{ad\,f} = 0{,}5679$	1	0,0011
Lcime	$T_{ad\,f} = 0{,}6257$	1	0,0156
Lspc	$T_{ad\,f} = 1{,}8637$	11	0,0000
Lina	$T_{ad\,f} = 0{,}6546$	12	0,0000
	critical values: 5% = -1.945 1% = -2.595		
First difference	**Test Statistic**	**Lag Order**	**t-Prob**
DLicms	$T_{adf} = -10{,}781$ **	1	0,0005
DLindu	$T_{adf} = -7{,}0663$ **	3	0,0060
DLelet	$T_{adf} = -6{,}9447$ **	10	0,0007
DLgaso	$T_{ad\,f} = -5{,}8659$ **	1	0,0000
DLoleo	$T_{ad\,f} = -9{,}8902$ **	1	0,0069
DLcime	$T_{ad\,f} = -9{,}2862$ **	1	0,0147
DLspc	$T_{ad\,f} = -7{,}4674$ **	10	0,0000
DLna	$T_{ad\,f} = -3{,}6440$ **	11	0,0000
	critical values: 5% = -2.904 1% = -3.527		

Notes: [1] Test made with a constant presence.

[2] The lag order choosen was based on Doornik and Hendry (1994) strategy.

[*] significant at the 5% level.

[**] significant at the 1% level.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

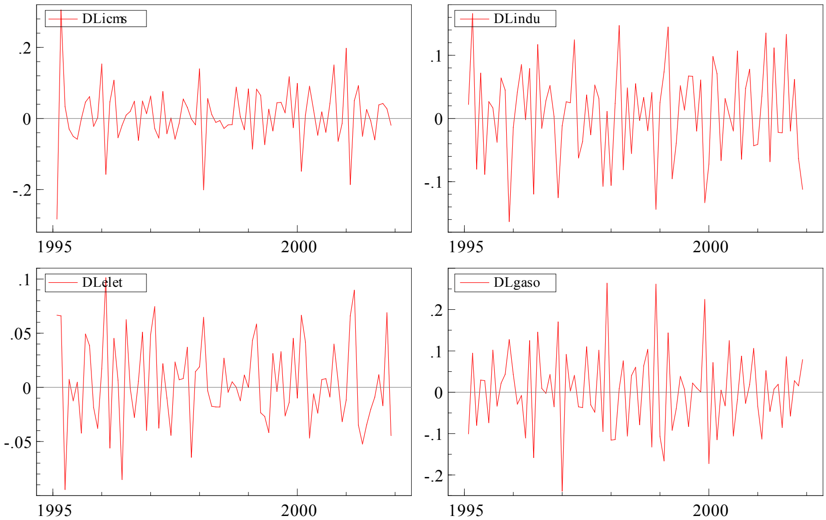

Figure 3 – Behavior of the first difference of series DLicms, DLindu, DLelet, DLgaso

Period: jan/1995 a dec/2001

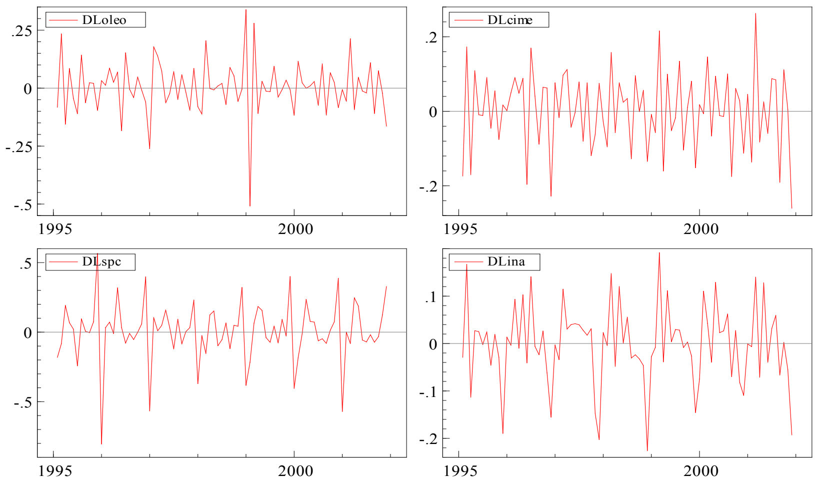

Figure 4 – Behavior of first difference of series DLoleo, DLcime, DLspc, DLina -

Period: jan/1995 a dec/2001

Marginalization and Conditioning

In the marginalization process those variables which will participate in the model should be evaluated, verifying the estimated parameter significance and the

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

magnitude of the coefficients. In general the modeling is estimated by Ordinary Least Squares (OLS) method and Table 2 presents the results.

Table 2 - Modeling *Licms* by OLS

Variables	Coefficient	Standard error	t-value	t-prob	Part-R^2
Constant	-0,49790	2,7055	-0,184	0,8545	0,0004
Lindu	0,83019	0,10615	7,821	0,0000	0,4459
Lelet	0,44283	0,19750	2,242	0,0279	0,0620
Lgaso	-0,042742	0,14109	0,303	0,7628	0,0012
Loleo	-0,067584	0,13623	-0,496	0,6212	0,0032
Lcime	0,072383	0,13885	0,521	0,6037	0,0036
Lspc	0,11539	0,058898	1,959	0,0538	0,0481
Lina	-0,40849	0,13075	-3,124	0,0025	0,1138

Note: These results are obtained with *PcGive* software

In an initial analysis it was verified that the variables: gasoline consumption (*Lgaso*) - the natural logarithm of gasoline consumption, 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.

Before discarding any of the variables, Granger's test of causality 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 able to be used for forecasts. The results discarded the use of the variables: diesel oil consumption (*Loleo*) and cement consumption (*Lcime*). The results are shown if Table 3.

Table 3 - Results of Granger causality test

Variables	Lags	Statistics
Lindu => Licms	7	F(8,62) = 3.625 [0.0016] **
Lelet => Licms	8	F(9,60) = 2.6783 [0.0110] *
Lgaso => Licms	8	F(9,60) = 2.3639 [0.0233] *
Loleo => Licms	7	F(8,62) = 0.99272 [0.4506]
Lcime => Licms	8	F(9,60) = 0.64462 [0.7544]
Lspc => Licms	8	F(9,60) = 6.4237 [0.0000] **
Lina => Licms	7	F(8,62) = 3.4561 [0.0024] **

Note: Numbers in parenthesis are F statistics, shows the direction of causality.

The remaining variables after coefficient's significance test and Granger analysis are *Lindu*, *Lelet* and *Lspc*.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

Results of Cointegration Tests

The tests of integration of the series analyzed suggested non stationarity of variable levels and following the two-step procedure of Engle and Granger to verify the existence of cointegration among the series, the equilibrium equation (long-run) should be estimated. Initially, estimate the long-run equation in variable level, the model is summarized in Table 4. It can be concluded that the variables are co-integrated (same order of integration and stationary residuals),

Table 4 - Modeling *Licms* by OLS

Variables	Coefficient	Std-Error	t-value	t-prob	Part-R^2
Constant	-4,6918	2,1035	-2,230	0,0285	0,0585
Lindu	0,62677	0,086562	7,241	0,0000	0,3959
Lelet	0,66237	0,14218	4,659	0,0000	0,2134
Lspc	0,12463	0,053665	2,322	0,0228	0,0632

Note: These results are obtained with PcGive software

and the long-run equation (5) was obtained:

$$Licms_t = -\,4{,}6918 + 0{,}62677\ Lindu_t + 0{,}66237 Lelet_t + 0{,}12463\ Lspc_t \qquad (5)$$

Re-parameterization 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 (5) to a short–term model incorporating an error correction mechanism (ECM). For the estimation of the short-term model, lagged variables were used and the residuals of the cointegration 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 reduction process the method of elimination of variables was employed based on the significance of the statistics *t* and *F*, minimizing the Schawrz criteria (SC) and Hannan-Quinn (HQ). The most parsimonious model has its results in Table 1, where letter *D* denotes the first-order difference (i.e., $Dy_t = y_t - y_{t-1}$). As required by the error correction mechanism, if equation 1 is a vector of valid co-integration, the sum of the coefficient 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 (6) below, will be employed, which repeats the results of Table 5.

BJ Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

Table 5 – Modeling *Dlicms* (the first difference of the logarithms of icms) by OLS.

Variables	Coefficient	Standard Error	t-value	t-prob	Part-R^2
Constant	0,014874	0,005800	2,970	0,0043	0,1263
DLicms_2	-0,35478	0,087623	-4,049	0,0001	0,2118
DLicms_7	0,15467	0,070090	2,207	0,0311	0,0739
DLindu_12	-0,27831	0,083545	-3,331	0,0015	0,1539
DLelet_6	-0,68609	0,16380	-4,189	0,0001	0,2234
DLelet_10	0,34484	0,15499	2,225	0,0298	0,0751
DLspc_5	0,087971	0,024927	3,529	0,0008	0,1696
DLspc_12	-0,20338	0,030047	-6,769	0,0000	0,4289
ECM_1	-0,62955	0,062480	-10,076	0,0000	0,6247
ECM_2	0,46128	0,079157	5,827	0,0000	0,3576

DW = 2,24 R^2 = 0,723425 RESET F(1,60) = 12198 [0,2738]

ARCH 5 F(5,51) = 0,1544 [0,9778] AR 1-5 F(5,56) = 1,5467 [0,1903]

NORM $\chi^2_{(1)}$ F(18,42) = 0,43708 [0,9699]

Note: Results from the software *PcGive* 8.0.

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

The model results 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 ($R^2 = 0.72$).

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 of the error corrections (in this case: - 0.62955 and + 0.46128). The value found for the coefficient of error corrections was - 0.16827, being significant and presenting a compatible sign with economic theory. Such parameter indicates that, on average, 16.827% of the changes in the collection of ICMS in the current period are due to the alterations in the expectations of an increase of ICMS in the two subsequent periods. The values of the parameters of the explicative variables are statistically significant and the value of the Durbin-Watson test (DW) discards the presence of residual correlation.

Specification Tests

The model appears to be well specified and presents constant coefficients throughout the sample period judging from the statistical tests, which do not indicate any specification problems. In Figure 5, where the estimated forecasts are always within the prediction interval of 95%.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

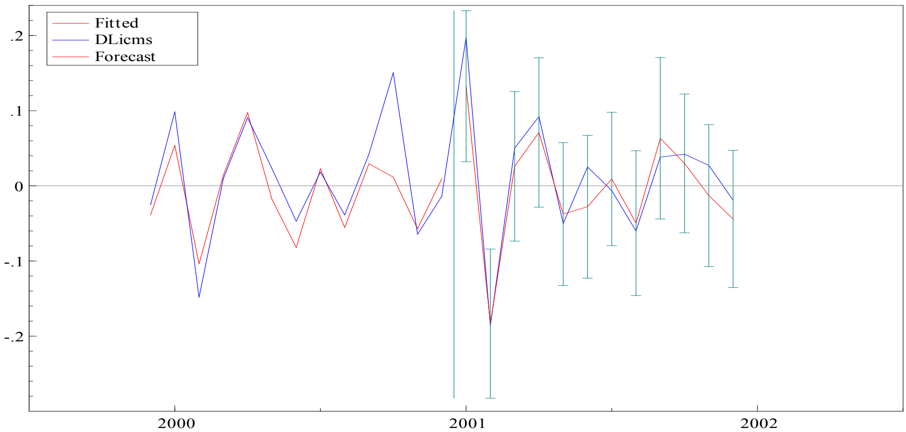

Figure 5 – Adjustment of the series *Dlicms* and forecasts for the year of 2001

The model in equation 6 presents all of the statistically significant regressors. Figure 6 shows the residuals as a function of time and the correlogram of the residuals up to lag 13, and both the graphs and tests point to the adequate specification of the model.

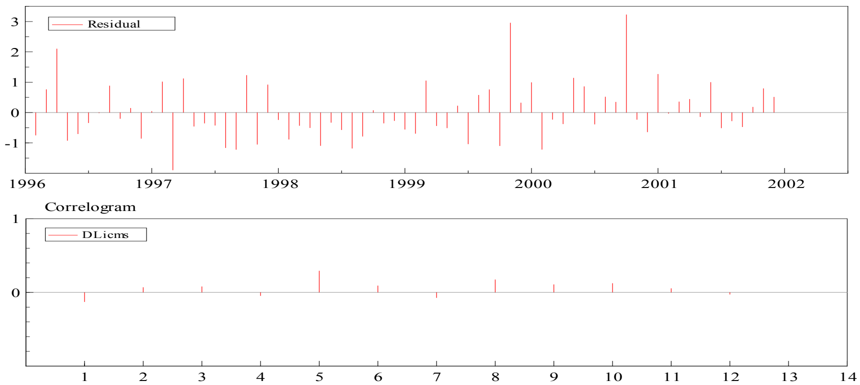

Figure 6 – Residual analysis

Forecasts and Comparative with the Actual Model

For the effects of comparison of the model analyzed, the observed values of the year 2001 were used. The Mean Absolute Percentage Error (MAPE) as a criterion of accuracy of the models was employed. Table 6 shows the observed values for the collection of ICMS as well as the forecasts with the model used presently.

a) Actual model - mean absolute percentage error (MAPE)= 4.63%

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

b) Equation 6 - mean absolute percentage error (MAPE) = 2.51%

The values obtained by the dynamic model with the incorporation of error correction for the forecast within the sample interval had a better adjustment, according to the criteria chosen, than the values obtained from the actual model employed by the Budget Department. One of the advantages of Equation 6 is the fact that none of the explanatory variables use contemporary values, facilitating the calculation of the one step ahead forecasts.

Table 6 – ICMS observed and forecast values for year 2001, old procedure ARIMA and new procedure Model 6

Months	ICMS	Arima(1, 0, 2)[1]	Model 6
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

Notes: [1] Values obtained from the software *eViews*

[2] Calculations done in *MS-Excel*

To verify the efficiency of equation 6 and its eventual applicability for the elaboration of tax collection forecasts, estimates of the tax collections for a period of 4 months beyond the researched sample period, January to April of the year 2002, were done. For the months of February, March and April in which the model demanded lags in the year 2002 (outside of the research interval) for the variables ECM_{t-1} and ECM_{t-2}, the model ARIMA was employed for the obtainment of the respective values. In the case of the variable $Dlicms_{t-2}$ for the months of March and April the proper forecasts generated in this estimate were used.

Table 7 presents the values obtained and forecasted by equation 6 for the first four months of 2002 (MAPE = 2.55%).

From the results obtained in the first 4 months of 2002, the quality of the forecasting model elaborated in this paper can be verified. It should be emphasized

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

that the forecast for the month of January 2002, when the values of all the exogenous variables for the model were known, was only 0.17%. Furthermore the model only depends upon three exogenous variables (consumption of electric energy, industry revenues and consultations with SPC) which make it very easy to use in practice.

Table 7 – ICMS observed and forecast values for the year of 2002

Months	ICMS	Model 6
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
MAPE[1]		2.555

Note: Calculations done in MS-Excel

Final Considerations

Following the methodology developed by the LSE, also known as the general to specific approach, in this work a forecasting model for the long-term was constructed and afterwards a model was estimated for the short-term with the incorporation of an error correction term, to be employed in the forecast of value added tax collections.

The forecasts for the value added tax of Santa Catarina formulated in this work presented for 2001 results significantly better than those obtained with the model presently employed (ARIMA), according to criteria established by percentage error.

For the forecasts in the short-term, up to one year, the creation of an application in a spreadsheet is suggested with the historic data employed in this paper, as well as the values calculated for the error correction mechanism and an area for the insertion of up-to-date values of explanatory variables. It is also recommended that the dynamic model be revised periodically to incorporate new effects and changes in the relevant variables occurring in the State.

References

Armstrong, J. S. (1985), Long-Range Forecasting: from Crystal Ball to Computer, New York: John Wiley.

Charemza, W. W. and Deadman, D. F. (1997), New Directions in Econometric Practice, Cheltenham: Edward Elgar.

Corvalão, E. D. (1999), Estudo comparativo de modelos de previsão aplicados à arrecadação do ICMS no Estado de Santa Catarina, Trabalho de Conclusão de Estágio, Departamento de Ciências da Administração, Universidade Federal de

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

Santa Catarina. (in Portuguese)

Cuthbertson, K.; Hall, S. G. and Taylor, M. P. (1992), Applied econometric techniques, Ann Arbor: University of Michigan Press.

Doornik, J. and Hansen, H. (1994), An omnibus test for univariate and multivariate normality. Nuffield College working paper. Available at: http://www.nuff.ox.ac.uk/Economics/papers/index1995.htm. Access: 19th November, 2009.

Doornik, J. A. and Hendry, D. F. (1994), PcGive 9.0 Interactive Econometric Modelling of Dynamic Systems, London: International Thomson Publishing.

Engle, R. F, and Granger, C. W. J. (1987), "Co-integration and Error Correction: Representation, Estimation and Testing", Econometrica, Vol. 55, pp. 251-276.

Engle, R. F.; Granger, C. W. J. and Hallman, J. J. (1989), "Merging short-run and long-run forecasts: an application of seasonal cointegration to montly electricity sales forecasting". Journal of Econometrics, Vol. 40, No. 1, pp. 45-62.

Ericsson, N. R.; Campos, J. and Tran, H-A. (1990), "PCGIVE and David Hendry's Econometric Methodology". Revista de Econometria, Vol. 10, No. 1, pp. 7-117.

Fildes, R. (1985), "Quantitative Forecasting – the State of the Art: Econometric Models", Journal of the Operational Research Society, Vol. 36, No. 7, pp. 549-580.

Gilbert, C. L. (1986), "Professor Hendry's econometric methodology", Oxford Bulletin of Economics and Statistics, Vol. 48, pp. 283-307.

Hendry, D. F. (1980), "Econometrics: Alchemy or Science?", Economica, Vol. 47, No. 188, pp. 387-406

Hendry, D. F.; Pagan, A. R. and Sargan, J. D. (1984), Dynamic Specifications, In: Z., Griliches and M. D. Intrilligator (eds.), The Handbook of Econometrics, Amsterdam: North Holland.

Pagan, A. (1990), "Three Econometric Methodologies: A Critical Appraisal", In: Granger, C. W. J. (ed.), Modeling economic series, London: Oxford University Press.

Samohyl, R. W. (2009), Controle Estatístico de Qualidade, Rio de Janeiro: Elsevier-Campus.

Biography

Eder Daniel Corvalão is a System Analyst from Centro de Informática e Automação de Santa Catarina (CIASC). He received the degree of Doctor in Industrial Engineer at the Federal University of Santa Catarina – Brazil, in 2009. His areas of interest are Time series analysis, Forecasting and Data Mining.
Contact: eder@ciasc.gov.br

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 105-121

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.
Contact: samohyl@deps.ufsc.br

Gutemberg Hespanha Brasil is a full professor of the Department of Statistics at the Federal University of Espírito Santo – Brazil. He received the degree of Doctor of Electrical Engineer, in 1989 at the Pontífica Universidade Católica do Rio de Janeiro. Professor Gutemberg is a member of the Brazilian Operations Research Society. His areas of interest are Time series analysis, Bayesian models, Econometrics and Forecasting.

Contact: ghbrasil@terra.com.br

Article Info

Received: November, 2009

Accepted: May, 2010