STATISTICAL ANALYSIS OF THE BRAZILIAN ELECTRICITY SECTOR: A TOP-DOWN LONG RANGE ENERGY CONSUMPTION AND SUPPLY FORECAST MODEL

Paula Medina Maçaira
Pontifical Catholic University of Rio de Janeiro, PUC-Rio
Electrical Engeneering Department
paulamacaira@gmail.com
Felipe Leite Coelho Silva
Federal Rural University of Rio de Janeiro, UFRRJ
Mathematics
felipeleite@ufrrj.br
Fernando Luiz Cyrino Oliveira
Pontifical Catholic University of Rio de Janeiro, PUC-Rio
Industrial Engineering Department
cyrino@puc-rio.br
Rodrigo Flora Calili
Pontifical Catholic University of Rio de Janeiro, PUC-Rio
Metrology Department
rcalili@esp.puc-rio.br
Plutarcho Maravilha Lourenço
Pontifical Catholic University of Rio de Janeiro, PUC-Rio
Institute of Energy, IEPUC
plutarchoml@gmail.com

Abstract

To produce accurate prediction of time series, it is first necessary to know and understanding all the historical information available. This way, the series of electrical energy consumption and production in Brazil were collected and analysed by descriptive statistics in order to find the presence of Stylized Facts before submitting them to standard forecasting methods. Two well known forecasting approaches were used in the modelling stage of the series. The first one, the Two Parameters Holt method as the univariate method and Dynamic Regression method as a causal model, using the Brazilian GDP as the unique explanatory variable. The results obtained by the fitted models show that the gap between the energy consumption and production series tends to increase with the time horizon (up to 2050 in this paper). However, if the energy generation do not materialize, there will be problems to supply the projected consumption.

KEYWORDS: Brazilian Electricity Sector. Dynamic Regression. Holt.

Main area: EST - Estatstica. EN - PO na rea de Energia. MP - Modelos Probabilsticos.

1 Introduction

The behavior of energy consumption in Brazil has been widely investigated over the past years. This interest, in general is due to the great financial and social importance of the sector; its failure or shortage can cause all sort of damages to the country. In this context, for planning purposes, the study of long range forecasts of the electricity market behavior are rather important.

Figure 1 illustrates the system of generation and transmission of electricity in Brazil. Currently, the national System Operator (ONS from now on) is responsible for the coordination of the generation and transmission of electricity, while the Electrical Energy Commercialization (CCEE from now on) administers the contracts, the liquidation of the spot market and the energy auction mechanisms in Brazil. The Brazilian grid is composed of four interconnected subsystems named: Southeast & Midwest, South, North and Northeast.

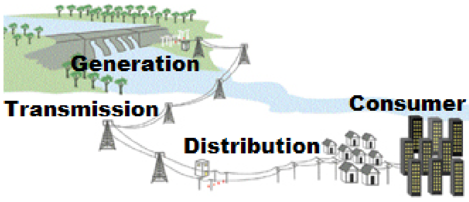


Figure 1: Generation and transmission of electricity
Source: Diário de Pernambuco

The data used in this work were the annual observations of consumption and electricity generation (measured in GWh), ranging from 1991 till 2013, available at the Energy Research Company (EPE from now on). It was also used the monthly observations on energy generation available at ONS, covering the period from 2000 till 2013. The GDP observations used in this work were collected at the Organization for Economic Co-operation and Development (OCDE from now on).

The objective of this paper is to carry out first a statistical analysis of the dataset collected followed by the fitting of the two long term type of top down forecast covering the horizon ranging from 2014 till 2050, considering the disaggregation of the consumption series in the following sectors: commercial, industrial, residential and public + transportation + livestock. This was followed by the analysis of the annual consumption growth ratio and its relationship with the corresponding annual growth GDP ratio. The last analysis was dedicated to the production series disaggregated by in the four subsystems stated above. In Bossmann et al. (2013), Elsland et al. (2013), Fleiter et al. (2010) and Jakob et al. (2012) is possible to see works with top down and bottom up approaches.

In this paper is applied two parametric models, first the Two-Parameter Holts method with damped trend and second the Dynamic Regression Model, both explained hereafter. More works in electricity forecasting can be found in Ghods and Kalantar (2011), Santana et al. (2012), Xue and Geng (2012), Anguita et al. (2012) and Bonanno et al. (2009).

This paper is organized as follows: In section 2 it is briefly described the theoretical aspects of the two models used in the modeling exercise. This is followed by section 3 where a statistical analysis of the Brazilian production and consumption energy series is presented. In section 4 are shown the results obtained by the application of Holts method with damped trend and dynamic regression with GDP (and its lags) as explanatory variable. Finally, in section 5 the main conclusions are drawn based on the results obtained from the models and directions for continuing this research are presented.

2 Methodology

In this paper is applied two parametric models, first the Two-Parameter Holts method with damped trend and second the Dynamic Regression Model, both explained hereafter.

Let y_t be a stochastic process in discrete time t, representing the energy series (either one, consumption and/or production) at time t, $t = 1, \cdots, T$. Also, let $\widehat{Y}_t(h)$ be the forecast of Y_{t+h} made at time t. The Two-Parameter Holts method with damped trend of the forecasting equation is given by:

$$\widehat{y}_t(h) = l_t + \sum_{i=1}^{h} \phi^i b_t, \tag{1}$$

where l_t and b_t denote estimates of the level and of the trend (slope), respectively, of the series at time t and, ϕ represents the damping parameter, with $0 < \phi < 1$. The estimate of the level of the series is described by

$$l_t = \alpha y_t + (1-\alpha)(y_{t-1} + \phi b_{t-1}), \tag{2}$$

where α the level smoothing parameter, with $0 < \alpha < 1$. The estimate of the trend component is given by:

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}, \tag{3}$$

where β represents the trend smoothing parameter, with $0 < \beta < 1$. For more details see: Montgomery and Johnson (1990).

Moving now to the second approach, i.e., the Dynamic Regression Model, it is described by the following general equation:

$$y_t = \sum_{i=1}^{k} \beta_i(L) x_{it} + \frac{1}{a(L)} \epsilon_t \tag{4}$$

Where y_t is the dependent (or endogenous) variable, x_{it} are the explanatory (or exogenous) variables (GDP in the present modelling exercise); $\beta_i(L) = \frac{b_i(L)}{a(L)}$ and $a(L)$, $b_1(L)$, ..., $b_k(L)$ are finite order lag polynomials of degrees r, s_1, ..., s_k, respectively, and ϵ_t is assumed to be white noise.

This equation decompose the value of y_t in two terms: the first term captures the systematic dynamics due to the influence of the explanatory variables on the actual value of y_t while the second term reflects residual dynamics, which is the dynamics in y_t not explained by the k explanatory variables in the model. This representation is a special case of a Transfer Function model. For details, see (Cochrane and Orcutt, 1949).

The accuracy measure used to compare the models is the Mean Absolute Percentage Error (or MAPE), as used in Fu and Nguyen (2003) and Turkay and Demren (2011). The equation to calculate MAPE is described by

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{5}$$

where A_t is the actual value and F_t is the forecast value.

3 Statistical Analysis

In this section, a descriptive statistical analysis of the energy series (consumption and production) is carried out. The two series (total Brazil) are plotted in Figure 2 below. Looking at Figure 2, it is possible to see the clear cut difference between the production and the consumption series that is growing year by year. It is also possible to visualize an approximate linear growth trend in both series. Also in Figure 2 one can clearly see the effect of the energy rationing that happened in 2001, when the reservoir of the hydro plants reached the lowest level of water of the history, forcing the government to decree the compulsory consumption reduction of 20% of all classes of consumers. The consumers that fail to attain this target reduction were heavily penalized. A similar break in the growth (not as deep as the 2001 one) happened later on, in 2009, caused by the world financial crisis started in 2008.

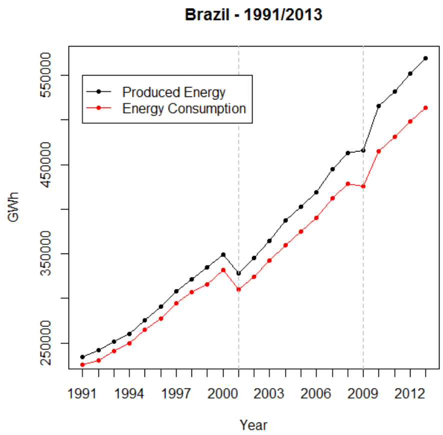
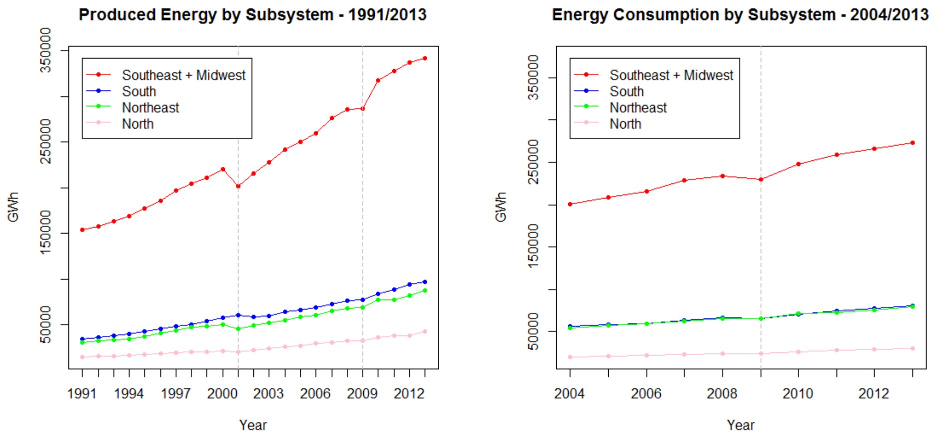


Figure 2: Brazilian produced energy and energy consumption data (year)
Source: The authors

Moving now to Figure 3(a), it is displayed the production series for each one of the four subsystems, while in Figure 3(b) is displayed the corresponding consumption series for each subsystem; the former corresponding to the period from 1991 till 2013 and the latter from 2004 till 2013. Important to mention the unequal share of consumption among the subsystems. The Southeast+Midwest alone is responsible for more than a half of the nations generation and consumption, followed by the South, Northeast and the North subsystems. To give an idea of the discrepancy of the distribution, in 1991, the Southeast+Midwest subsystem produced 65.7% of the energy consumed in Brazil, followed by by the South with 14.7%, the Northeast with 13.3% and the North with 6.3%. These percentages, in 2013 became, respectively: 60.0%; 17.0%; 15.5% and 7.5%, indicating an increase in the production of 15.6% in South subsystem, 16.6% in the Northeast and 18.7% in the North. On the other hand in the most important subsystem (Southeast+Midwest), there was a decrease of 8.6% in the energy production. Also important to mention is the fact that the 2001 rationing did not reach the South region and this is reflected in the almost insensitive movement on the consumption series of this subsystem around the years 2001 and 2002 (differently from the sharp decay in consumption in the subsystems Southeast+Midwest and Northeast that were most heavily affected by the rationing.

In order to provide a view of the production x consumption series as shown in Figure 2 for the total Brazil, it is displayed in Figure 4 the same plots of these two series separately for the four subsystems. Here, as in the Brazilian total of Figure 2, one can also observe the



(a) Brazilian produced energy (year) (b) Brazilian energy consumption (year)

Figure 3: Produced and Consumed Series
Source: The authors

increase in the difference between the two series, specially in the North subsystem where this difference is more pronounced.

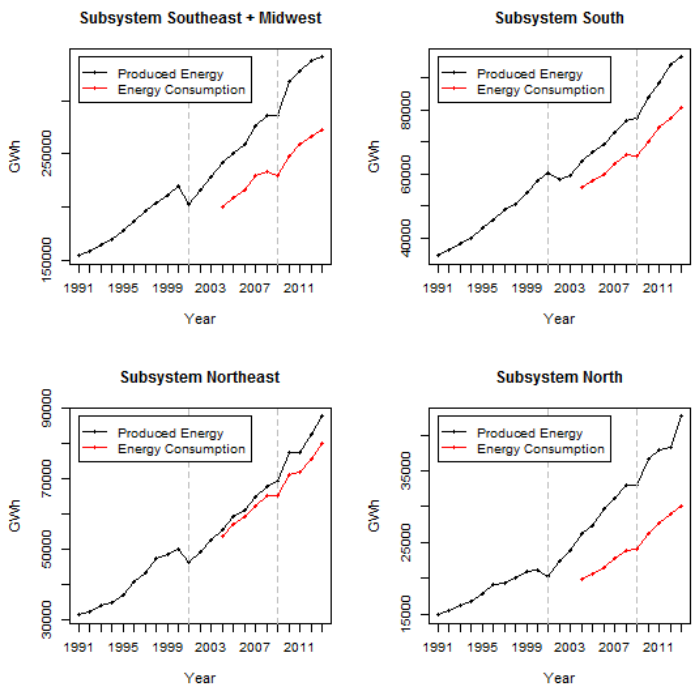


Figure 4: Brazilian produced energy and energy consumption data by subsystem (year)
Source: The authors

Looking now at Figure 5, a further disaggregation of the consumption series is displayed, now by activity sectors (Residential, Commercial, Industrial and Public + Transportation + Livestock). One interesting fact can be observed in this graph. While in the 2001 rationing all sectors were affected, in the financial crises of 2008/2009 only the industrial sector was affected. In fact, for this important sector, the ratio of decrease in consumption in 2001 in relation to 2000 was 4.99% while the same ratio in 2009 with respect to 2008 was 5.3%;

indeed very similar percentages.

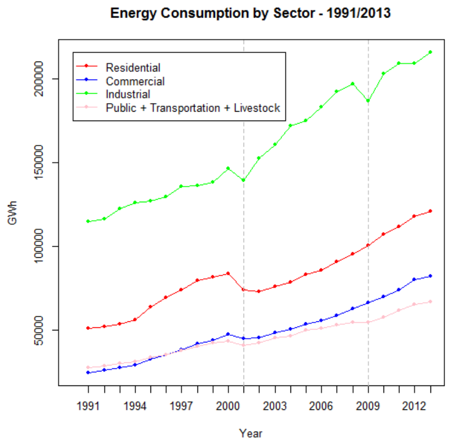


Figure 5: Brazilian energy consumption data by sector (year)
Source: The authors

Continuing the descriptive analysis, Figure 6 displays the monthly energy production and consumption, covering the period from January 2000 until December 2013 (for the production series, source EPE) and the period from January 2004 until December 2013 (for the consumption series; source ONS). This display allows similar conclusions to what was detected in Figure 2, however with the behaviour of the series on a month by month basis. As a matter of fact, as the graph shows, the 2001 rationing started in March, 2001 for the two subsystems; Southeast+Midwest, Northeast and in September, 2001 in the North subsystem. The South subsystem did not enter the rationing as the reservoirs of the hydro plants of that region were not affected by the dry period of the other 3 regions.

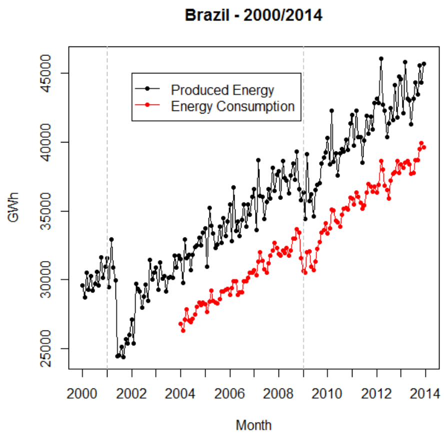
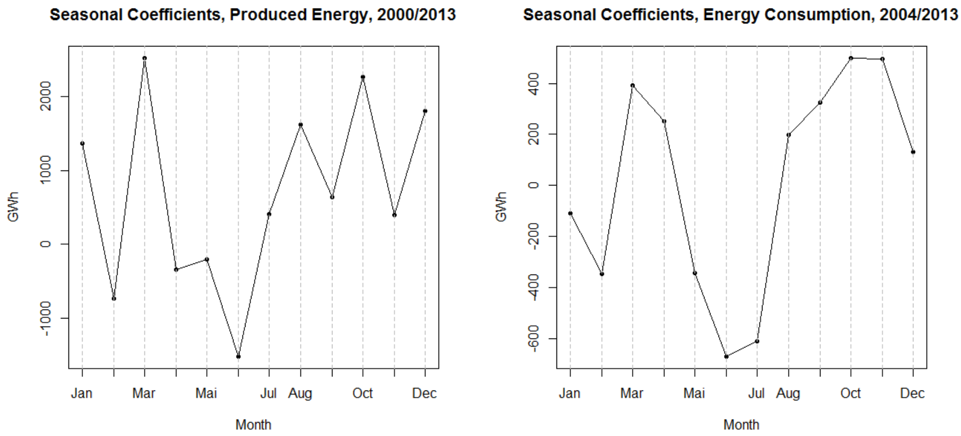


Figure 6: Brazilian produced energy and energy consumption data (month)
Source: The authors

It is also possible to extract from the monthly series (Figure 6) the seasonal factors that indicate how the series behave over the months within a year. Figure 7(a) displays the seasonal coefficients of the production series and in Figure 7(b) the so responding coefficients for the consumption series. Analysing these factors, one can see that in May, June and July both series have a somehow similar behave, i.e., the coefficients are smaller, while on the months of March, August, September and October these coefficients are bigger than those in the other months for both series.



(a) Produced energy (b) Energy consumption

Figure 7: Seasonal coefficients
Source: The authors

Finally, in Figure 8 the annual GDP growth rate and the annual energy consumption growth rate are compared. A quick look at this plot reveals that there is a strong positive correlation between these two series, a clear cut indication that the GDP is, indeed, a strong explanatory variable to be considered in a transfer function kind of model to explain the energy consumption.

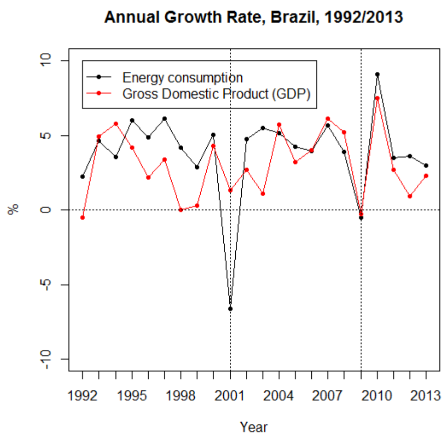
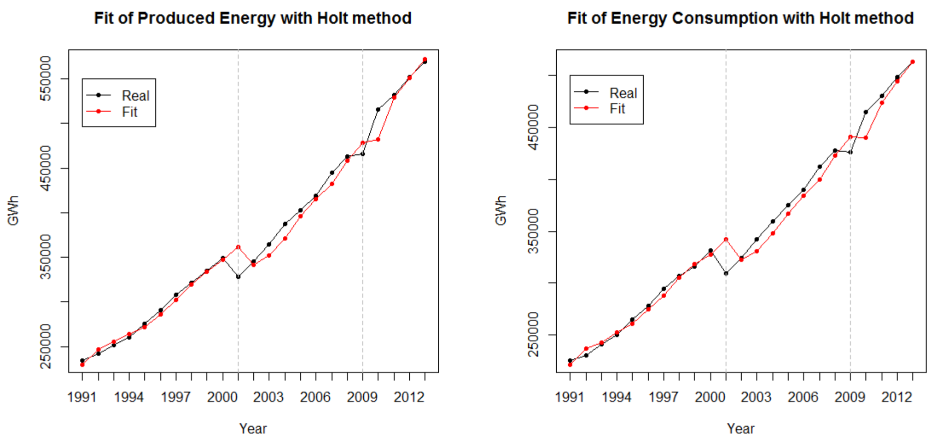


Figure 8: Annual growth rate of GDP and energy consumption data
Source: The authors

4 Results

In this section the relevant results of the fitting of the two models (Holt and Dynamic Regression) to the electricity series being studied in this paper are presented. All the analysis was carried out using the R Software (R Core Team (2012)) and the packages: tseries, forecasts ggplot2 and dynlm.

Firstly, the two annual production and consumption series were submitted to the Holt method. The adjustment for both; Production and consumption series are shown in Figures 9(a) and 9(b), respectively. The quality of this adjustment, measured by the MAPE (Mean Absolute Percentage Error) was, respectively 2.16% (Production) and 2.21% (Consumption); indeed a very good fit to the data for both series.



(a) Produced energy data (b) Energy consumption data

Figure 9: Fit with Holt method
Source: The authors

Moving now to the extrapolation of both series till 2050, it was first estimated the damping coefficient to be used in the forecasting equations. It was found that the optimal value for this coefficient was 0.98 for both series. In Figure 10 are displayed the point forecasts for the two extrapolations. It can be seen from the projections that the gap between the historical data tends to increase with the time horizon as one moves into the future. Unless the consumption in the country increases in a higher rate than that projected, the produced results point to a possible obsolete supply capacity in the future.

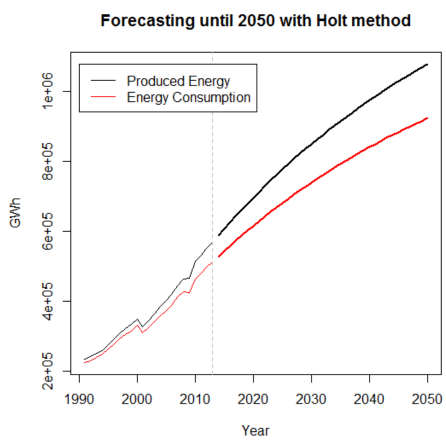


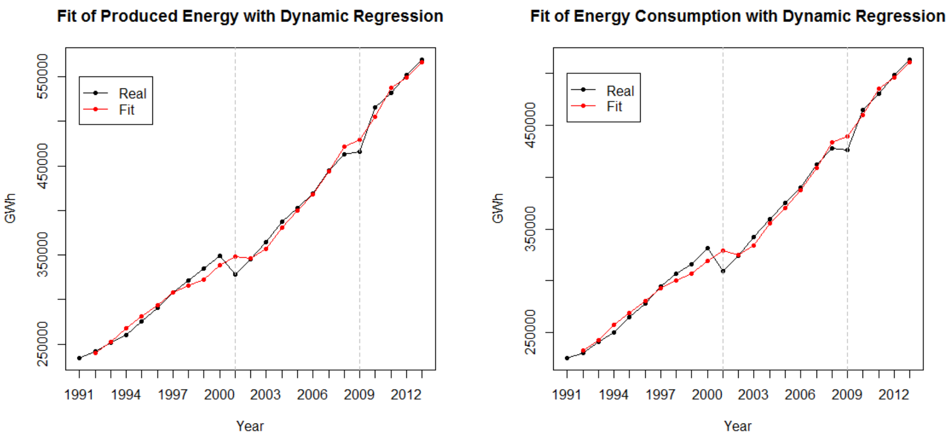
Figure 10: Comparison of forecast until 2050 with Holt method
Source: The authors

The model found for produced energy series with Holt methods is in Equation 6 and for energy consumption series is in Equation 7.

$$\widehat{y}_t(h) \quad = \quad 0,7905 + \sum_{i=1}^{h} 0,98^i 0,1943 \qquad (6)$$

$$\widehat{y}_t(h) \quad = \quad 0,7905 + \sum_{i=1}^{h} 0,98^i 0,1559 \qquad (7)$$

To conclude the model fitting exercise of this paper, the Dynamic Regression approach was used in two situations; in the first, the endogenous variable was the production series while in the second the endogenous variable was the consumption series. In both the explanatory variable was the country GDP (and its lagged values). The MAPE estimated for both models within sample was 1.55% and 1.6% for both, production and consumption series, respectively. As expected, the MAPE value reduced quite substantially from the univariate to this transfer function application, and also confirms the importance of the GDP series to explain the evolution of these electricity series. The two adjusted models can also be seen in Figures 11(a) and 11(b).



(a) Produced energy data (b) Energy consumption data

Figure 11: Fit with Dynamic Regression
Source: The authors

In order to compare the two projections up to time 2050 of the two series, in Figure 12 it is shown the point projections obtained by the two series. For the GDP forecasting was used the projected values in the 2012 OECD Economic Outlook, see more in OECD (2012). As one can see, the gap between the two series still exists, even though in a lesser discrepancy than that obtained with the univariate extrapolation.

Finally, in Figures 13(a) and 13(b) the difference between both methods forecasts are displayed for the production and consumption series. From Figure 13(a) one can extract that the Dynamic Regression model points to a faster increase in the projection than that projected by the Holts method. The same happens in the comparison for the consumption series shown in Figure 13(b).

In Table 1 is possible to see the estimate, standard error and t value for dynamic model fitted to produced energy series, the most important result is that all the variables are significant at 1% level. And in Table 2 is showed the same results to energy consumption

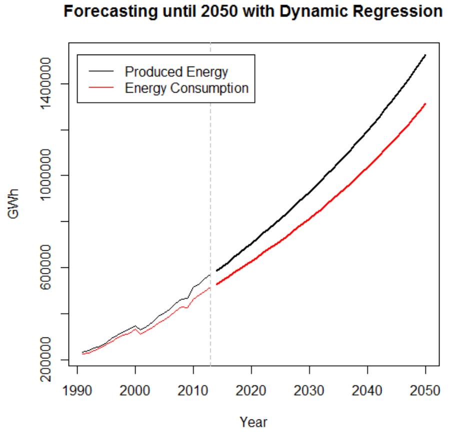
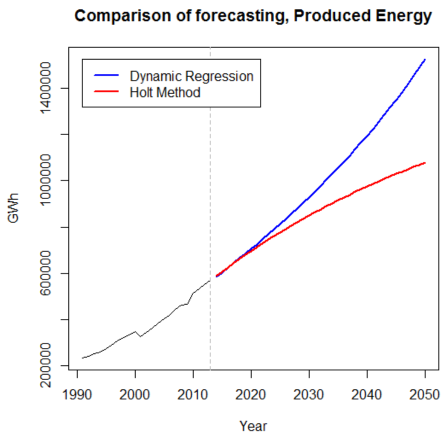
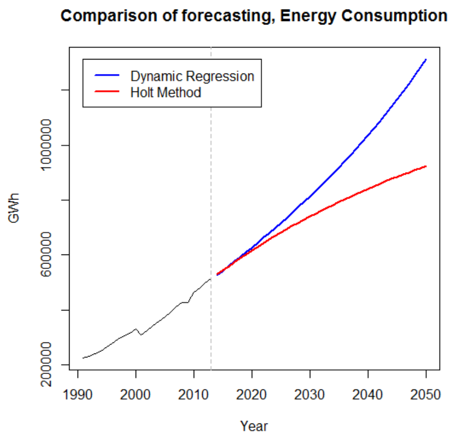


Figure 12: Comparison of forecast until 2050 with Dynamic Regression
Source: The authors



(a) Produced energy series



(b) Energy consumption series

Figure 13: Comparison of forecasting
Source: The authors

series, and also all the variables are significant at 1% level.

Table 1: Statistics for dynamic model fitted to produced energy series

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
Intercept	-6,38E+04	1,62E+04	-3,944508	0,00087
GDP	7,88E-02	1,75E-02	4,492257	0,00025

Table 2: Statistics for dynamic model fitted to energy consumption series

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
Intercept	-3,64E+04	1,15E+04	-3,170316	0,005038
GDP	6,64E-02	1,52E-02	4,356786	0,00034

5 Conclusions

At the introductory section of this paper, it was briefly described how the Brazilian electricity sector is organized, its size, its complexity and an overview of its historical development. It was also presented the peculiarities of the energy transmission and distribution grid which makes it a real challenge without any similar system in the world. The historical series of both, production and consumption at the aggregate and disaggregated level (subsectors and sectors) were presented in section 2 and submitted to a thorough descriptive analysis in the following section, where the important Stylized Facts of each could be extracted to be used in the forecasting model fitted in the next section. Two approaches were used to produced the forecasts, a straight forward univariate Two-Parameter Holt method with Damped Trend in the forecasting equation and the Dynamic Regression method to model the relationship between these series and the nations GDP series. The results obtained by both approaches show a very good fit to the data, both method producing in sample MAPE's very low indeed, with a substantial gain of the Dynamic Regression model over the Holt counterpart (a reduction of the MAPEs from 2.2% to 1.5%), a clear indication of the improvement on the fitting accuracy by the introduction of the GDP (and its lagged values) in the model formulation. The forecasts up to year 2050 obtained from both approaches for the two series indicates that the gap between production and consumption already observed on the historical data tend to increase even more as the time horizon. On the other hand, although the projected demand highly probably wont be affected by environment constraints, the same can not be said for the supply side, that may not be attained due to serious environmental restrictions.

The results obtained in this modelling exercise with one univariate model and a transfer function kind of model (with only one explanatory variable) can be viewed as a simple benchmark long range forecasts. Obviously, more sophisticated top down econometric approaches including other explanatory variables, or the bottom up kind of methods where the projections are made for disaggregated sector and sub-sectors of the various consumer classes for each sub-systems separately, using technological and macro-economic variables are, by far, more accurate. However, these top down kind of methods should not be neglected, on the contrary, they can be considered good starting point to tackle this important problem of long range energy forecasts.

Acknowledgement

The authors would like to thank the R&D program of the Brazilian Electricity Regulatory Agency (ANEEL) for the financial support.

References

Anguita, D., L. Ghelardoni, and A. Ghio (2012). Long-term energy load forecasting using auto-regressive and approximating support vector regression. In *Energy Conference and Exhibition.*

Bonanno, F., G. Capizzi, and G. Tina (2009). Long-term energy performance forecasting of integrated generation systems by recurrent neural networks. In *International Conference on Clean Electrical Power.*

Bossmann, T., B. Pfluger, and M. Wietschel (2013). The shape matters! how structural changes in the electricity load curve affect optimal investments in generation capacity. Fraunhofer Institute for Systems and Innovation Research ISI.

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto- correlated error terms. *Journal of the American Statistical Association 44*, 32–61.

Elsland, R., W. Eichhammer, and B. Schlomann (2013). Is enough electricity being saved? impact of energy efficiency policies addressing electrical household appliances in germany until 2030. Fraunhofer Institute for Systems and Innovation Research ISI.

Fleiter, T., D. Fehrenbach, E. Worrell, and W. Eichhammer (2012). Energy efficiency in the german pulp and paper industry e a model-based assessment of saving potentials. *Energy*, 1–16.

Fleiter, T., S. Hirzel, M. Jakob, J. Barth, L. Quandt, F. Reitze, F. Toro, and M. Wietschel (2010). Electricity demand in the european service sector: A detailed bottom-up estimate by sector and by end-use. In *Improving Energy Efficiency in Commercial Buildings*.

Fu, C. and T. Nguyen (2003). Models for long-term energy forecasting. In *Power Engineering Society General Meeting, 2003, IEEE*.

Ghods, L. and M. Kalantar (2011). Different methods of long-term electric load demand forecasting: A comprehensive review. *Iranian Journal of Electrical & Electronic Engineering 7*, 249–259.

Jakob, M., G. Catenazzi, and T. Fleiter (2013). Ex-ante estimation of the eu ecodesign directives impact on the long-term electricity demand of the tertiary sector. TEP Energy GmbH and Fraunhofer Institute for Systems and Innovation Research (ISI).

Jakob, M., T. Fleiter, G. Catenazzi, S. Hirzel, F. Reitze, and F. Toro (2012). The impact of policy measures on the electricity demand of the tertiary sector of the european countries: An analysis with the bottom-up model forecast. In *Improving Energy Efficiency in Commercial Building Conference*.

Montgomery, D. and L. Johnson (1990). *Forecasting and Time Series Analysis*. McGraw-Hill Book.

OECD (2012, June). *OECD Economic Outlook*, Volume 2012. OECD.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Santana, A. L., G. B. Conde, L. P. Rego, C. A. Rocha, D. L. Cardoso, J. C. Costa, U. H. Bezerra, and C. R. Francs (2012). Predict decision support system for load forecasting and inference: A new undertaking for brazilian power suppliers. *Electrical Power and Energy Systems*, 33–45.

Turkay, B. E. and D. Demren (2011). Electrical load forecasting using support vector machines. In *Electrical and Electronics Engineering (ELECO)*.

Xue, B. and J. Geng (2012). Dynamic transverse correction method of middle and long term energy forecasting based on statistic of forecasting errors. In *IPEC, 2012 Conference on Power & Energy*.