Models for Southeast Load

Mônica Barros

March, 22nd, 2002

Objectives

Our primary objective in this report is to present simple models that will allow us to forecast the DAILY load in the Southeast subsystem in Brazil.

The database consists of observations collected since January 1^{st}, 2001, and all models were fitted using as the final point March, 18th, 2002.

We take a step-by-step approach, and we present a tentative model. We believe we will not have our "final" model until we obtain temperature data for major metropolitan areas in the Southeast region, since it is a widely known fact that electricity consumption is heavily dependent on this variable. Moreover, it is important to note that the temperature effect will probably be relevant when we consider several important load centers simultaneously.

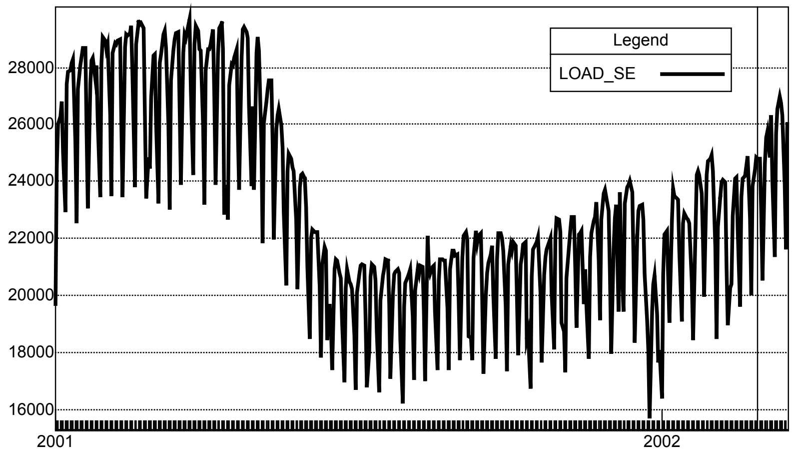
Why is Daily Load Forecasting Important for Trading Purposes?

Short-term (daily and eventually, hourly) load forecasting is essential for trading, since it will give us a good indication on what type of position we should assume, and the volume of this position. For example, if we have a fairly accurate prediction that load is about to increase by x% on the next week (due, for example, to a weather forecast that predicts extremely high temperatures), in a totally free market we would expect prices to go up, and could use this information to our advantage.

Unfortunately, the current situation in Brazil is still very much away from a free market set-up, but under the newly proposed offer and demand system, short term load forecasts will be as important as short term inflow energy forecasts, so we need to attack the problems from both sides!

Below we exhibit a time series graph with the load data for the Southeast subsystem. One should note the dramatic impact of rationing (from June 2001 to February 2002).

Time Series Graph – Southeast Load



Tentative Model 1

Model Structure: Linear Scale, Constant included in the model, Lagged Dependent Variables (up to 7 days), Dummy Variables for National Holidays and Rationing, Day of Week Indicator, Least Squares Estimation

Dependent Variable: LOAD_SOUTHEAST
Method: Least Squares
Date: 03/07/02 Time: 12:30
Sample(adjusted): 8/01/2001 5/03/2002
Included observations: 422 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	**2050.533**	**1331.534**	**1.539979**	**0.1243**
LOAD_SOUTHEAST(-1)	0.332666	0.041826	7.953559	0.0000
LOAD_SOUTHEAST(-2)	-0.114157	0.040054	-2.850093	0.0046
LOAD_SOUTHEAST(-3)	**0.027442**	**0.040547**	**0.676788**	**0.4989**
LOAD_SOUTHEAST(-4)	**0.017247**	**0.041990**	**0.410735**	**0.6815**
LOAD_SOUTHEAST(-5)	**-0.037280**	**0.040406**	**-0.922639**	**0.3567**
LOAD_SOUTHEAST(-6)	**0.100908**	**0.066705**	**1.512752**	**0.1311**

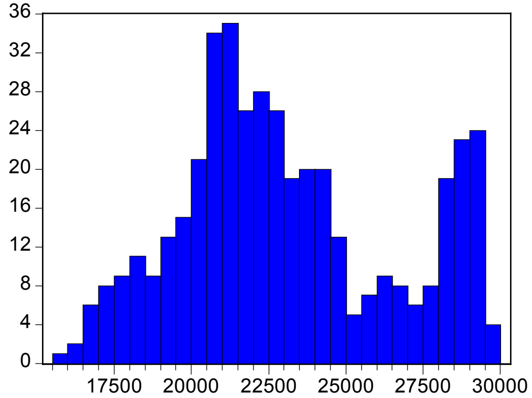
LOAD_SOUTHEAST(-7)	0.595833	0.053326	11.17337	0.0000
NAT_HOLIDAYS	-2928.877	330.2739	-8.868024	0.0000
WEEKDAY	**24.71201**	**68.12823**	**0.362728**	**0.7170**
RATIONING	**-423.4708**	**310.8725**	**-1.362201**	**0.1739**

R-squared	0.887131	Mean dependent var	23052.64
Adjusted R-squared	0.884385	S.D. dependent var	3497.637
S.E. of regression	1189.274	Akaike info criterion	17.02579
Sum squared resid	5.81E+08	Schwarz criterion	17.13123
Log likelihood	-3581.442	F-statistic	323.0397
Durbin-Watson stat	1.259962	Prob(F-statistic)	0.000000

Note: Non-significant variables shown above in boldface.

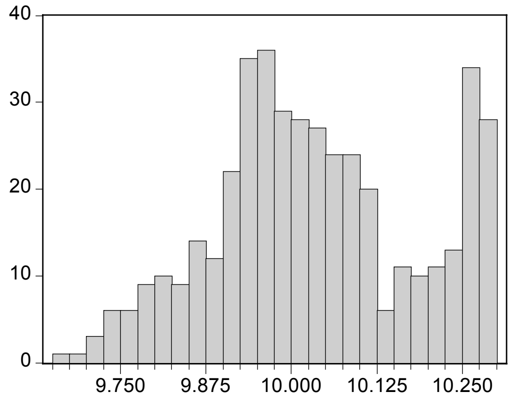
Obviously, there is a lot of room for improvement in this model. We can start by dropping out the insignificant variables and also by creating dummies for each weekday, instead of putting them all together in a single variable. Also, Normality of the dependent variable might be an issue, as shown in the following histogram (note the high value of the Jarque-Bera statistic, a clear indication of non-normality). Moreover, it wouldn't surprise me at all if the load distribution were bimodal, representing, to an extent, the effect of rationing (June 1st, 2001 to February 28th, 2002).

Histograms – dependent variable on original and log scale



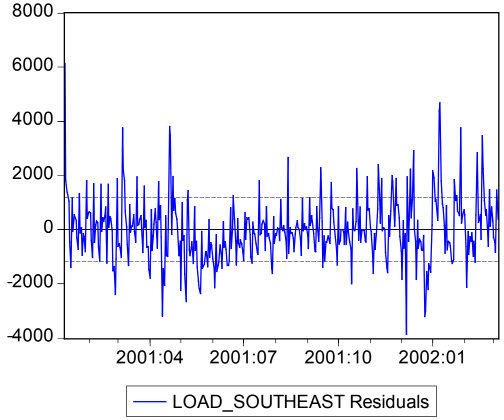
Series: LOAD_SOUTHEAST	
Sample 1/01/2001 5/03/2002	
Observations 429	
Mean	23075.80
Median	22292.00
Maximum	29704.00
Minimum	15708.00
Std. Dev.	3485.951
Skewness	0.331409
Kurtosis	2.187413
Jarque-Bera	19.65580
Probability	0.000054

If we log transform the data, there is no clear improvement in terms of Normality, as shown in the next histogram. I guess the problem is really the bi-modality I previously mentioned.



Series: LOG(LOAD_SOUTHEAST)	
Sample 1/01/2001 5/03/2002	
Observations 429	
Mean	10.03526
Median	10.01198
Maximum	10.29904
Minimum	9.661925
Std. Dev.	0.150244
Skewness	0.066686
Kurtosis	2.262865
Jarque-Bera	10.03065
Probability	0.006635

Needless to say, the residuals of the model are still quite bad, as the next graph shows.



Tentative Model 2

Model Structure: Linear Scale, Constant included in the model, Lagged Dependent Variables (lags 1, 2, 3, 5, 6 and 7), Dummy Variables for National Holidays, Rationing, Weekends, Mondays and Wednesdays, Least Squares Estimation.

Dependent Variable: LOAD_SOUTHEAST
Method: Least Squares
Date: 03/07/02 Time: 16:41
Sample(adjusted): 8/01/2001 5/03/2002
Included observations: 422 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3072.134	866.2896	3.546313	0.0004
LOAD_SOUTHEAST(-1)	0.665403	0.040540	16.41337	0.0000
LOAD_SOUTHEAST(-2)	-0.082546	0.039324	-2.099112	0.0364
LOAD_SOUTHEAST(-3)	0.124398	0.043531	2.857663	0.0045
LOAD_SOUTHEAST(-5)	0.149711	0.036362	4.117247	0.0000
LOAD_SOUTHEAST(-6)	-0.137910	0.032591	-4.231564	0.0000
LOAD_SOUTHEAST(-7)	0.185818	0.030101	6.173177	0.0000
NAT_HOLIDAYS	-2558.808	220.2461	-11.61795	0.0000
RATIONING	-529.0360	200.7525	-2.635265	0.0087
WEEKEND	-2635.212	204.4063	-12.89203	0.0000
MONDAY	1634.653	231.9577	7.047202	0.0000
WEDNESDAY	**434.2261**	**228.5706**	**1.899746**	**0.0582**

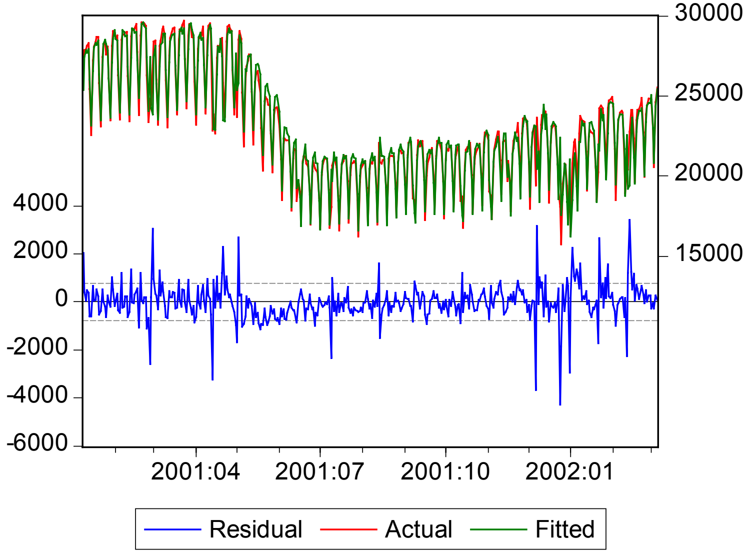
R-squared	0.952112	Mean dependent var	23052.64
Adjusted R-squared	0.950827	S.D. dependent var	3497.637
S.E. of regression	775.5981	Akaike info criterion	16.17317
Sum squared resid	2.47E+08	Schwarz criterion	16.28819
Log likelihood	-3400.539	F-statistic	741.0601
Durbin-Watson stat	1.792038	Prob(F-statistic)	0.000000

There´s a significant improvement in model fit when compared to tentative model 1. Some points need to be made:
The effect of national holidays (keeping all other variables constant) is a reduction of about 2559 MW in the daily load.
Similarly, rationing represented a decrease of roughly 529 MW on daily load (keeping all other effects fixed).
The effect of weekends is even more dramatic, and represents a load decrease of approximately 2635 MW.

One should keep in mind that current load (as of March 5th, 2002) is 25,542.

M. Barros Consultoria Ltda.
e-mail: barrosm@alumni.utexas.net
mailto:info@mbarros.com

Observed and Fitted Values, Residuals



Some residuals are still quite large, and we expect to improve model fit with the inclusion of temperature in some of the major load centers in the region, and we intend to incorporate this information in our models as soon as we obtain the data.

Short-term forecasting capacity for tentative model 2

We next use the structure proposed in tentative model 2 to forecast load up to one week ahead, thus we take as the end of the in-sample period February 26th, 2002. Note that rationing did not officially end until February 28th, 2002, even though its end was announced some 10 days earlier, so our forecasts actually include a period of structural break.

	Actual SE Load	Forecasted SE Load	Error	% Error
27/02/02	24,839	21,903	2,936	11.8%
28/02/02	24,816	21,804	3,012	12.1%
01/03/02	24,792	22,336	2,457	9.9%
02/03/02	22,958	20,575	2,383	10.4%
03/03/02	20,542	18,569	1,973	9.6%
04/03/02	24,625	22,325	2,300	9.3%
05/03/02	25,542	23,211	2,331	9.1%

Forecast errors are still quite large, and definitely improvements will be made by including other explanatory variables in the model. However, the current structure is simple enough to make it still useful. One point that should be kept in mind is that all forecasts underestimate the actual load, which can be due to two factors:

- Load started "picking up" as the end of rationing was announced in mid-February or,
- Temperature effect – the last couple of days have been warmer than on the previous weeks, and we had a fairly mild summer so far, which in part explained the low levels of load previously observed.

Quite probably, the reason for such low forecasts is a combination of both factors already mentioned.

Tentative Model 3

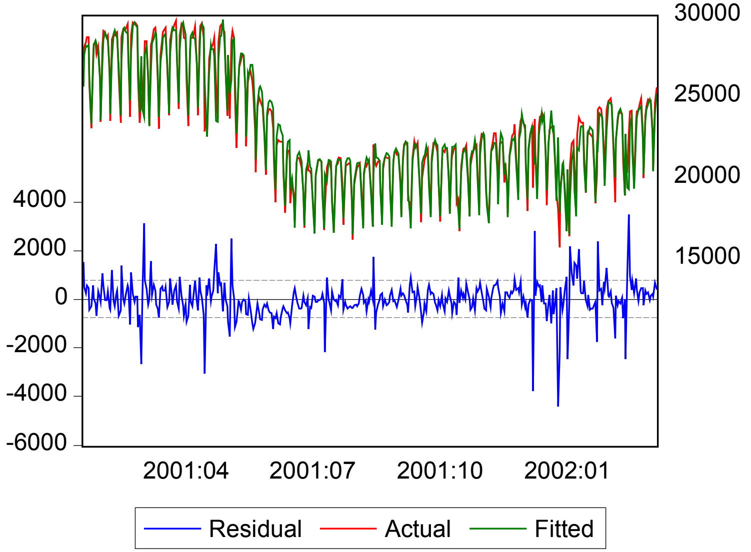
We start with a longer set of lagged covariates (originally up to 14 days as possible explanatory variables) and tentatively include some indicator variables in the model, namely: dummies for each day of the week (except Friday, to avoid collinearity) and a dummy for weekends, rationing and national holidays dummy variables.

The final model structure is:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	987.64	304.72	3.24	0.1%
LOAD_SOUTHEAST(-1)	0.60	0.03	18.13	0.0%
LOAD_SOUTHEAST(-3)	0.07	0.02	2.80	0.5%
LOAD_SOUTHEAST(-5)	0.12	0.03	3.85	0.0%
LOAD_SOUTHEAST(-7)	0.12	0.03	3.36	0.1%
LOAD_SOUTHEAST(-12)	0.10	0.04	2.68	0.8%
LOAD_SOUTHEAST(-13)	-0.16	0.03	-5.12	0.0%

LOAD_SOUTHEAST(-14)	0.14	0.03	4.54	0.0%
NAT_HOLIDAYS	-2778.93	220.68	-12.59	0.0%
MONDAY	1334.63	214.66	6.22	0.0%
WEEKEND	-2672.70	196.81	-13.58	0.0%
R-squared	**0.9528**	Mean dependent var		22,989.46
Adjusted R-squared	**0.9516**	S.D. dependent var		3,483.37
S.E. of regression	766.36	**Akaike info criterion**		**16.15**
Sum squared resid	237000000	**Schwarz criterion**		**16.25**
Log likelihood	**-3339.57**	F-statistic		814.93
Durbin-Watson stat	**1.67**	Prob(F-statistic)		-

Observed and Fitted Values, Residuals



Short-term forecasting capacity for tentative model 3

We repeat the procedure used for tentative model 2 and produce one week ahead forecasts starting on February 27th, 2002. The remarkable fact is: even though there is a structural break in this period, **the forecasting capacity of this model is quite superior to the previous one**, as the next table shows.

It remains to be seen if this excellent forecasting capacity can be maintained at other periods, since I personally feel the model might be slightly "too large" (however all lags included in the model are significant!).

Tentative Model 3

	Actual SE Load	Forecasted SE Load	Error	% Error
27/02/02	24,839	24,546	292.90	1.2%
28/02/02	24,816	24,868	-52.30	-0.2%
01/03/02	24,792	24,647	144.80	0.6%
02/03/02	22,958	22,735	222.70	1.0%
03/03/02	20,542	20,439	103.30	0.5%
04/03/02	24,625	23,946	679.10	2.8%
05/03/02	25,542	25,119	422.60	1.7%

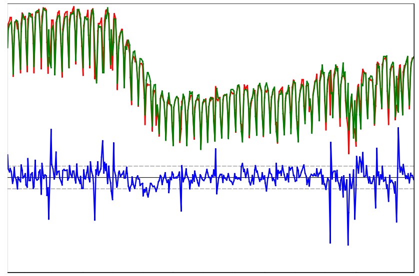
Tentative Model 4

We next attempt to incorporate the temperature effect into the tentative daily load models for the Southeast subsystem as a means of improving their predictive ability.

We obtained two new data series (average and maximum daily temperatures at the Botafogo measuring station in the Southern part of the city of Rio de Janeiro). This series probably does not represent extreme changes in temperatures in the city, as there are areas where the maximum observed temperatures can be 15 to 20 degrees Fahrenheit higher than in the measuring station for which we have available data at this time.

It seems remarkable to us that the maximum temperature series at the above mentioned station has a very poor explanatory ability, and "drops" out quickly of almost any model we attempt to fit. Surprisingly, the average temperature series remains in the "final" model, together with several lags of the dependent variable.

Model Structure: Linear Scale, Constant **not** included in the model, Lagged Dependent Variables (lags 1, 3, 5, 7, 12, 13, 14), Dummy Variables for National Holidays, Weekends, Mondays, Average Daily Temperature at Botafogo Measuring Station, Least Squares Estimation.

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Observed and Fitted Values, Residuals
					
LOAD_SOUTHEAST(-1)	0.53	0.03	16.18	0.0%	
LOAD_SOUTHEAST(-3)	0.07	0.02	2.92	0.4%	
LOAD_SOUTHEAST(-5)	0.10	0.03	3.47	0.1%	
LOAD_SOUTHEAST(-7)	0.13	0.03	4.13	0.0%	
LOAD_SOUTHEAST(-12)	0.11	0.03	3.28	0.1%	
LOAD_SOUTHEAST(-13)	-0.17	0.03	-5.79	0.0%	
LOAD_SOUTHEAST(-14)	0.14	0.03	4.61	0.0%	
NAT_HOLIDAYS	-2835.96	209.54	-13.53	0.0%	
MONDAY	1102.99	205.00	5.38	0.0%	
WEEKEND	-2692.80	181.64	-14.83	0.0%	
TEMP_AVER_RJ	103.74	13.42	7.73	0.0%	
R-squared	0.96	Mean dependent var		2298089.0%	
Adjusted R-squared	0.96	S.D. dependent var		349815.2%	
S.E. of regression	727.39	Akaike info criterion		1604.3%	
Sum squared resid	211,000,000	Schwarz criterion		1615.1%	
Log likelihood	-3277.87	Durbin-Watson stat		166.5%	

We haven't tested the short-term forecasting ability for tentative model 4 since we haven't generated forecasts for the temperature series.

M. Barros Consultoria Ltda.
e-mail: barrosm@alumni.utexas.net
mailto:info@mbarros.com