Working Paper 91-21
July 1991

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

FORECASTING DAILY DEMAND FOR ELECTRICITY WITH MULTIPLE-INPUT NONLINEAR TRANSFER FUNCTION MODELS: A CASE STUDY*

J. R. Cancelo**
Antoni Espasa***

Abstract

A model for forecasting the daily demand for electricity energy in Spain is presented. The trend and the weekly seasonality are modelled by using past values of the series; a complex intervention analysis is carried out to capture the effects of changes in the working conditions; and meteorological variables enter in the model with transfer functions, which allow nonlinear, dynamic, season-and type-of-day- dependent responses, as well as exhaustion effects and an implicit assessment of the increase on the stock of appliances.

Key words: Nonlinear ARMAX Models, Model Selection, Intervention Analysis.

* Invited paper at the Eleventh International Symposium on Forecasting, New York. This research has been commissioned by Predicción y Coyuntura, S.L., and sponsored by Red Eléctrica de España, S.A.

** Departamento de Economía Aplicada, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain.

*** Departamento de Economía, Universidad Carlos III, Getafe, 28903 Madrid, Spain.

Preliminary Version: May 1991

FORECASTING DAILY DEMAND FOR ELECTRICITY WITH MULTIPLE-INPUT NONLINEAR TRANSFER FUNCTION MODELS: A CASE STUDY (*)

J. R. Cancelo (**)

A. Espasa (***)

(*) Invited paper at the Eleventh International Symposium on Forecasting, New York. This research has been commissioned by Predicción y Coyuntura, S.L., and sponsored by Red Eléctrica de España, S.A.

(**) Applied Economics Department, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain.

(***) Economics Department, Universidad Carlos III, Getafe, 28903 Madrid, Spain.

ABSTRACT

A model for forecasting the daily demand for electric energy in Spain is presented. The trend and the weekly seasonality are modelled by using past values of the series; a complex intervention analysis is carried out to capture the effects of changes in the working conditions; and meteorological variables enter in the model with transfer functions, which allow nonlinear, dynamic, season- and type-of-day- dependent responses, as well as exhaustion effects and an implicit assessment of the increase on the stock of appliances.

KEYWORDS: Nonlinear ARMAX Models, Model Selection, Intervention Analysis

1. TIME SERIES MODELS FOR THE DAILY FORECASTING OF THE ELECTRICITY DEMAND

The purpose of this paper is to present a model for obtaining daily forecasts of the demand for electricity in Spain. In addition to its use as an instrument for solving a specific problem on forecasting, the whole process of model building is of general interest in the sense that it may be applied to a wide set of problems related to daily forecasting of economic activity variables.

Electric energy can not be stored, so producers must anticipate accurately future demand to keep it satisfied while avoiding over-production. Speaking broadly, two different horizons must be considered when forecasting:

1) the long run, that refers to the forecasts made for several years into the future, and whose aim is to determine the number and capacity of the new power stations to be built, and

2) the short run, aimed to optimize present resources by producing at a given moment of time just the energy that will be demanded. This kind of forecasts usually are made with hourly, daily or weekly horizons.

In this paper only the daily forecasting problem will be considered. In Spain the distribution of the electric energy is centralized at a state-owned enterprise, Red Eléctrica de España, S. A. (REE), and one of its tasks is to set the production of the several producers, given their capacity and the expected demand. As a consequence, it is of the greatest importance for REE to have good short term forecasts of the aggregated demand.

Up to 1989 those forecasts were made by a qualified staff on a judgmental basis, as this personnel had acquired a big experience that allowed them to carry out accurate subjective forecasts. These forecasts, being subjective, consider, in principle, all the information that is known at a given moment. But these experiences are difficult to transmit to other people, and often the resulting forecasts are strongly dependent on the specific person that has made them; this last characteristic must be avoided, as it may condition the whole performance of the company on the presence of specific individuals.

Moreover, socioeconomic conditions change and the demand for electricity shows a growing complexity, which individuals can not properly assess, and this obliges subjective forecasters to rely on statistical techniques that synthesize certain aspects of the real world.

This paper is organized as follows: in section 2 the main characteristics of the daily demand for electricity in Spain are reviewed; section 3 deals with modelling the changes in the working conditions. Sections 4 and 5 are devoted to the weather conditions: while section 4 presents the methodology for estimating their influence, section 5 shows the final estimates. Section 6 analyzes the dynamics of the disturbance term and the criteria used for model validation; the last section, section 7, summarizes the main conclusions. Five appendixes are added with the technical details.

2. DAILY DEMAND CHARACTERISTICS AND THEIR MODELLING

The series to be modelled is the peninsular Spanish daily demand for electric energy. The initial sample included data from January 1, 1983 to December 31, 1988, but since then the performance of the model has been periodically studied and the model has been reestimated each time a new semester of observations became available. No structural change has been detected in these new estimations. All estimations and analysis presented in the paper were made with the 1983-1988 sample, unless otherwise stated.

Figures 1, 2 and 3 point out two of the main characteristics of this series: a growing trend and several seasonal oscillations. The trend and the seasonal oscillations are reflected in figure 1, and a more detailed picture of these oscillations is presented in figure 2 (annual) and in figure 3 (weekly).

It can also be seen (figure 4) that the series is very sensitive to the presence of a holiday, a vacation period or some other event that affects the normal daily activity: in the example shown in this figure the weekly seasonal pattern is distorted by the presence of a holiday on Thursday, October 12. Throughout the paper this kind of events will be referred to as changes in the working conditions.

Even if at first sight it can not be detected, weather conditions exert a big influence on the demand; temperature differences from a winter to the next can induce large variations on the demand for the same month of two consecutive years.

So the main characteristics of the series to be analyzed are:

- a growing trend
- seasonal oscillations of weekly and annual periodicity
- to be highly sensitive to the changes of the working conditions

FIGURE 1.– SPANISH DAILY DEMAND

(Demand for electricity in 1983–1989)

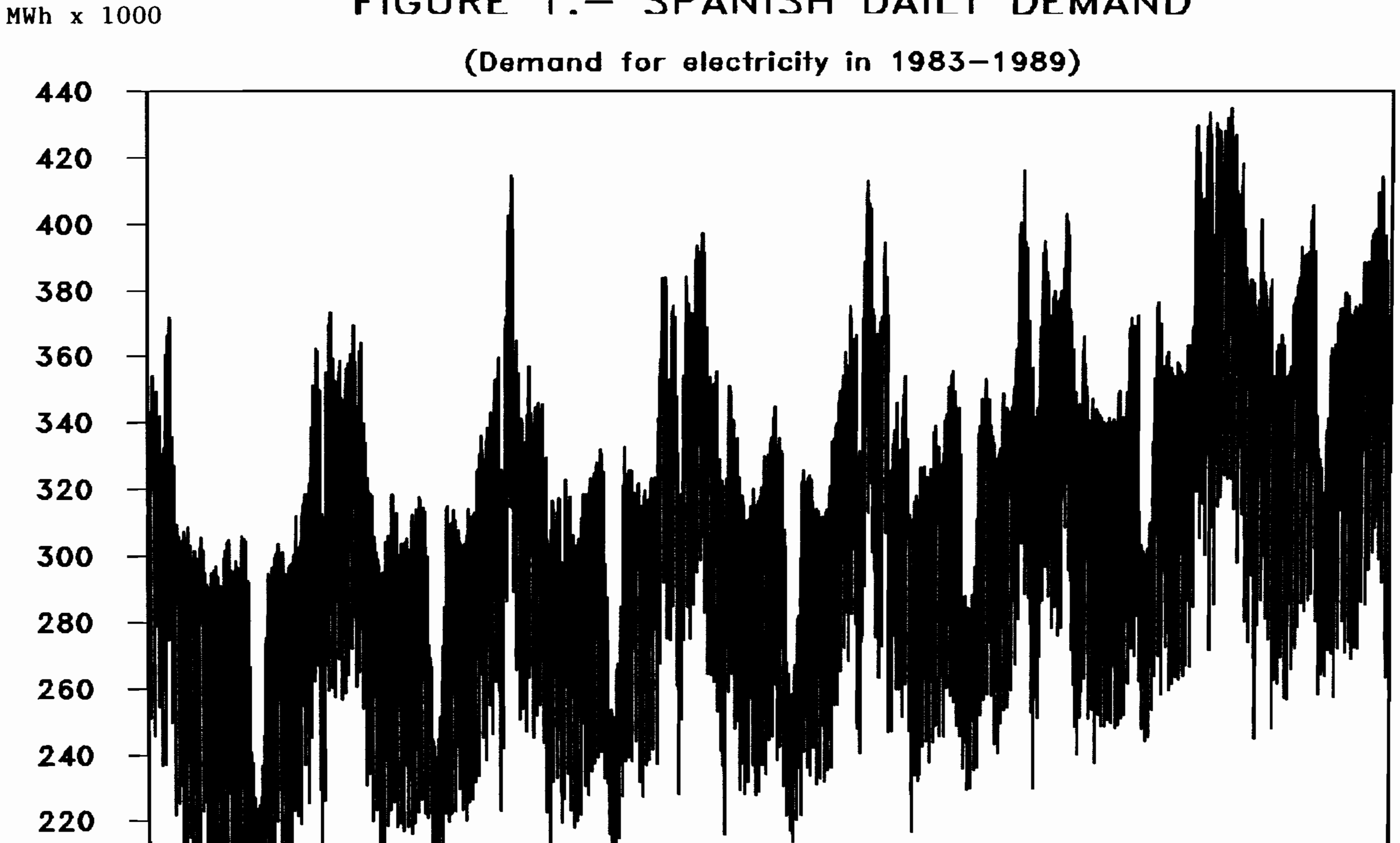


FIGURE 2.— DAILY DEMAND FOR 1989

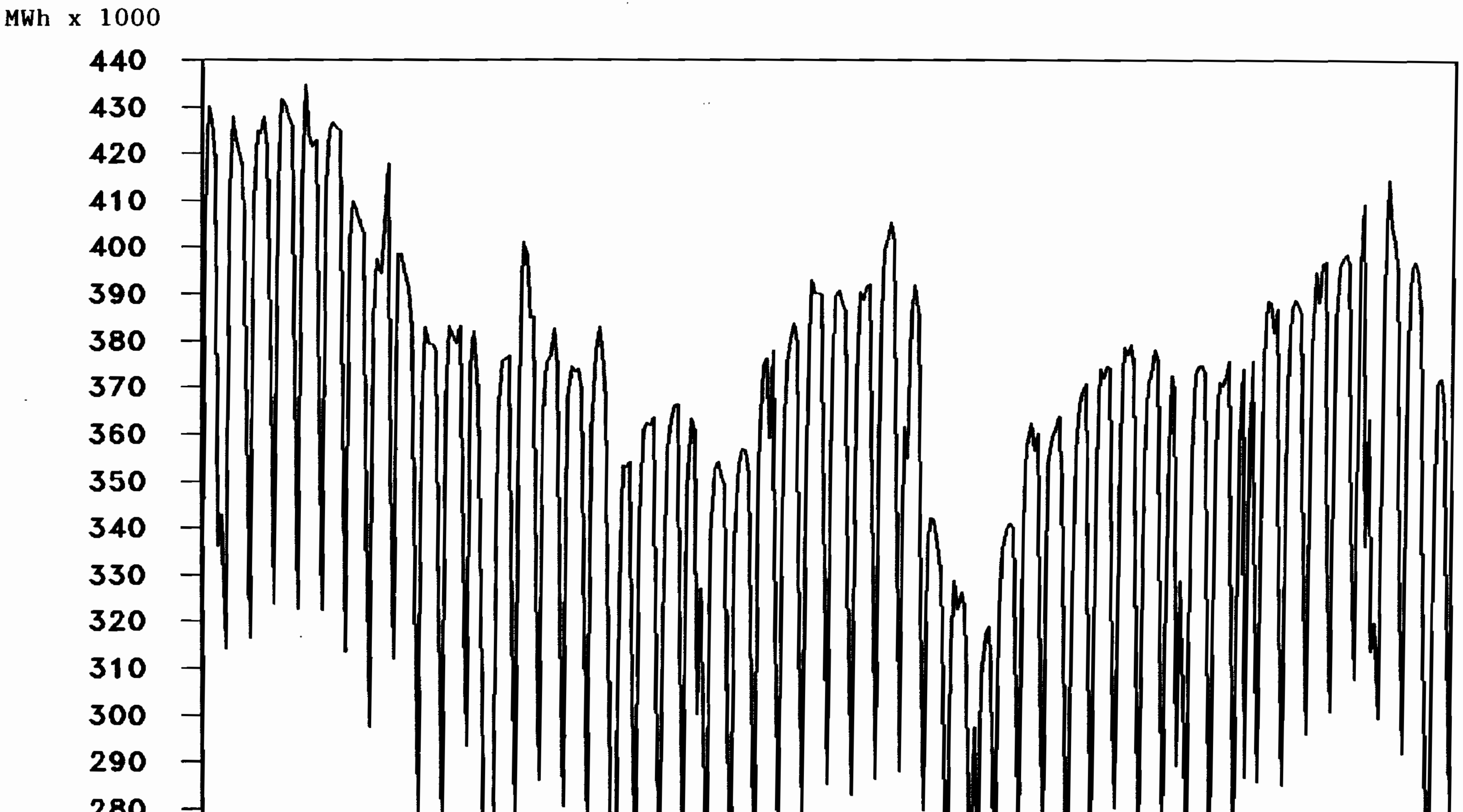


FIGURE 3.- FOUR-WEEK PERIOD DEMAND

(Monday Sep 4 to Sunday Oct 1, 1989)

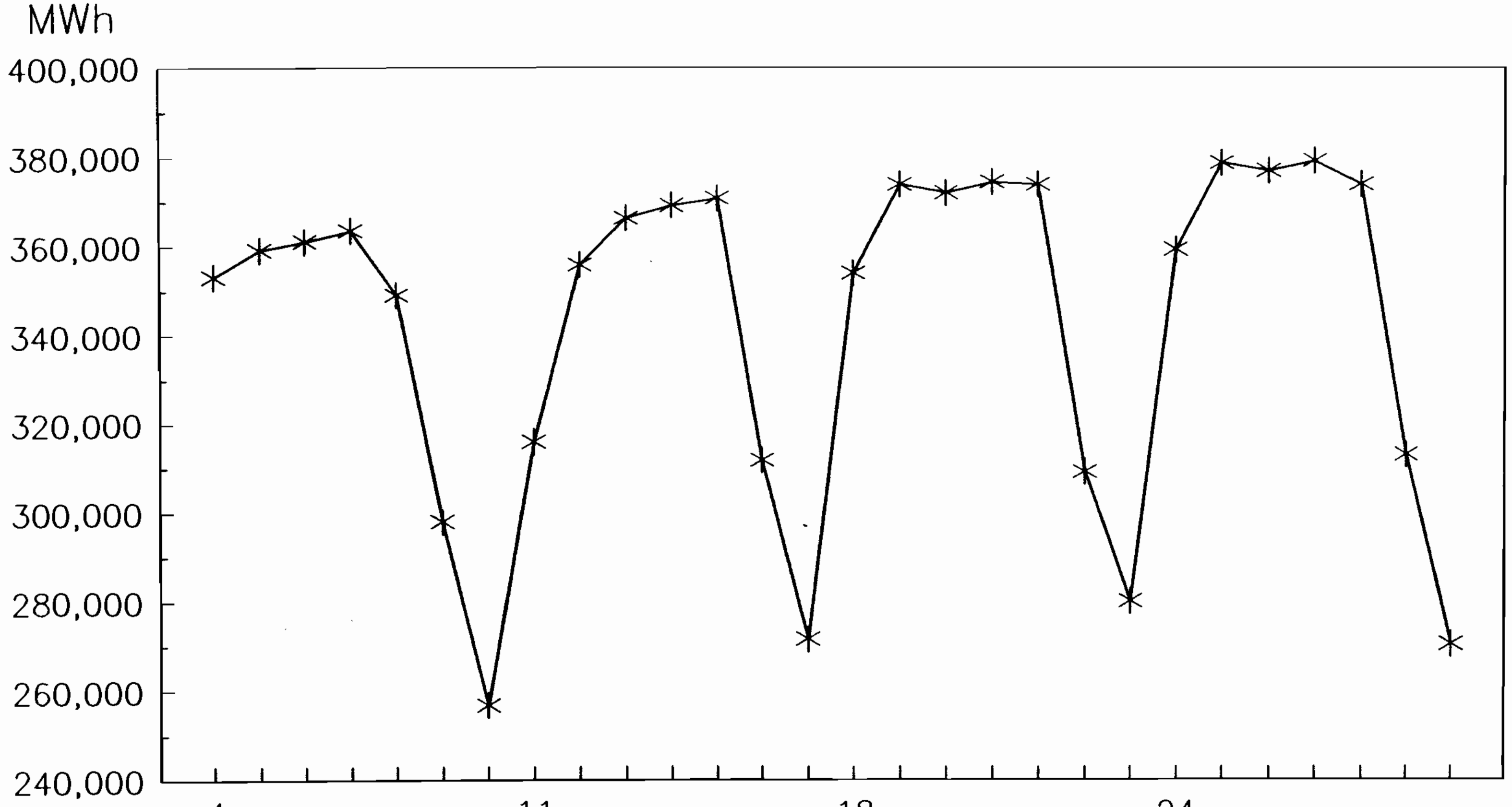
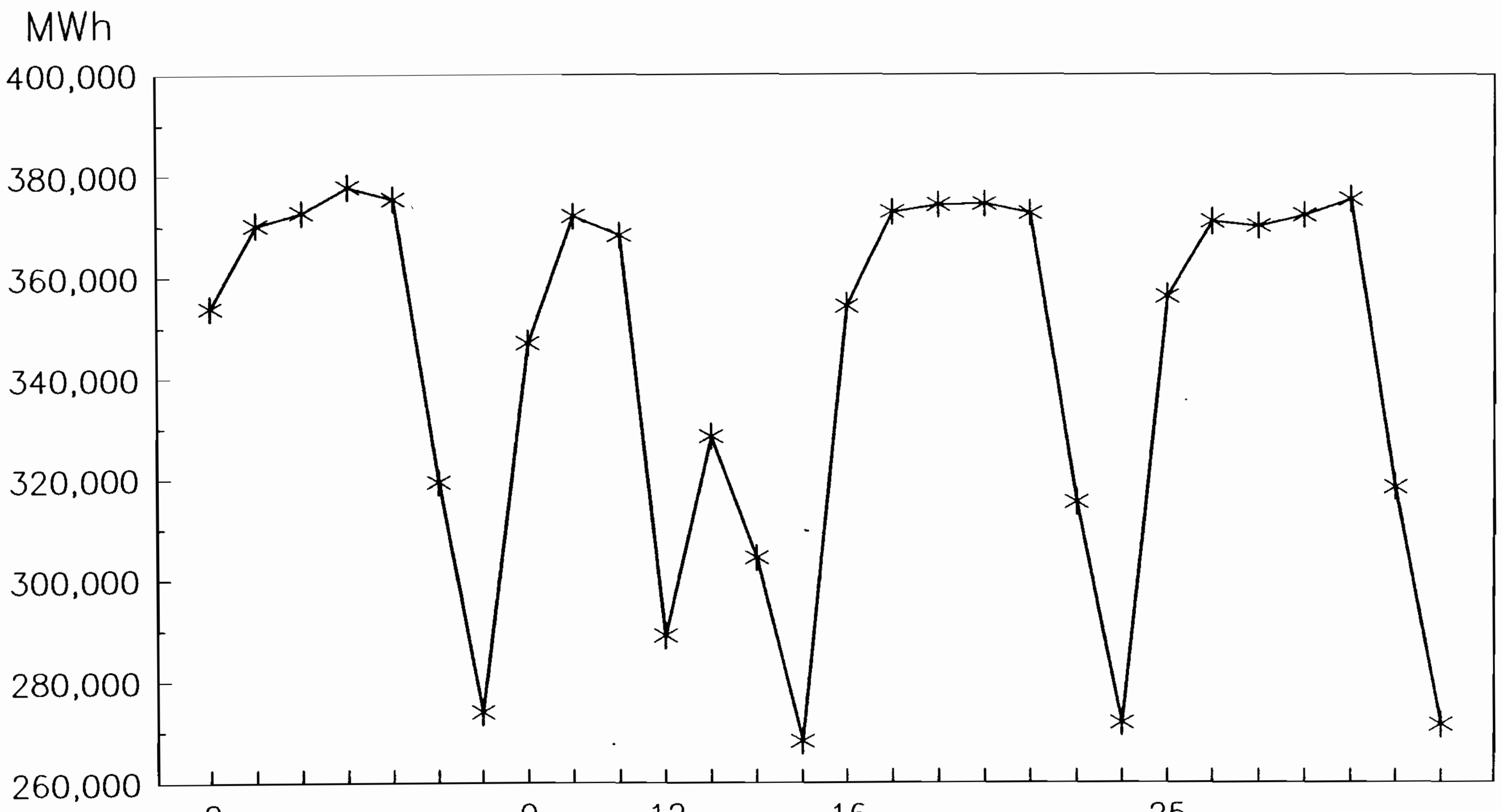


FIGURE 4.– DISTORTION DUE TO A HOLIDAY

(Monday Oct 2 to Sunday Oct 29, 1989)



- to depend on the weather conditions.

So a daily forecasting model must incorporate all these characteristics if good forecasts are to be obtained.

The trend growth is related to a general growth of the economic activity, and the weekly seasonal oscillations are quite systematic. No explanatory variables have been included in the model to deal explicitly with these characteristics[1], therefore they must be incorporated through the dynamics of the system by using the Box and Jenkins (1970) approach.

Assuming temporarily that the other characteristics are negligible, the trend and the weekly cycle can be captured by considering

$$\Delta\Delta_7 \ln D_t = w_t \qquad (1)$$

or equivalently

$$\ln D_t = \ln D_{t-1} + (\ln D_{t-7} - \ln D_{t-8}) + w_t \qquad (1')$$

where $\ln D_t$ is the natural logarithm of the demand in an specific day and w_t is a stationary disturbance[2].

Under this data generation process, both the trend and the weekly seasonality are stochastic, as they depend on past values of the demand. This approach generalizes deterministic models, in which the trend is assumed to follow a straight line and the seasonal coefficients for each day of the week are fixed. It has been checked that in this case these stochastic schemes provide much better forecasts than deterministic ones.

The formulation (1)-(1') has been obtained under the assumption that the other factors that influence the series are negligible. The annual seasonality turns to be mainly dependent on the working and on weather conditions; so (1)-(1') would be applicable to a demand that had been previously corrected of working and weather effects, call it D_t^c:

$$\ln D_t^c = \ln D_{t-1}^c + (\ln D_{t-7}^c - \ln D_{t-8}^c) + w_t \, . \qquad (2)$$

This corrected demand can be obtained from

$$\ln D_t^c = \ln D_t - f_1(L)\ WOR_t - f_2(L)\ WEA_t \, , \qquad (3)$$

where WOR represents the set of explanatory variables that refer to the changes of the working conditions, WEA the weather variables and $f_1(L)$ and $f_2(L)$ are polynomials on the lag operator L.

By combining (2) and (3) the general expression for the model is obtained

$$\ln D_t = \ln D_{t-1}^c + (\ln D_{t-7}^c - \ln D_{t-8}^c) + f_1(L)\ WOR_t + f_2(L)\ WEA_t + w_t \, , \qquad (4)$$

where the trend and the weekly seasonality are determined by past values of the demand once the working and weather effects have been removed from them, so that they can reflect more precisely the secular and seasonal evolution of the series. Moreover, the value on a specific day t depends on the working and weather conditions of t and of the recent past, as well as on short term dynamics that may be embodied on the disturbance w_t.

Using equation (3) it is possible to express (4) as

$$\ln D_t = \ln D_{t-1} + (\ln D_{t-7} - \ln D_{t-8}) + f_1^*(L)\ WOR_t + f_2^*(L)\ WEA_t + w_t \, , \qquad (4')$$

which is the specification finally estimated[3].

It must be noted that with a model like (4) or (4') the user is left with a single equation for the whole year, with no need to distinguish between winter model and summer model. Even if this single model is harder to build, it seems more adequate for the daily tracking of the demand. Moreover, the only restriction it imposes a priori with respect to season-specific models refers to the stochastic process related to w_t, which must be the same for all the seasons: it seems a rather soft restriction. All the other effects can be season-dependent if necessary.

In effect, in equation (4') we have a model which is assumed that is valid for every season within a year. As we shall see in subsequent sections, this restriction is not very strong since finally it is allowed that $f_2^*(L)$ will differ along a calendar year.

3. <u>CHANGES IN THE WORKING CONDITIONS AND THEIR EFFECT ON THE DEMAND</u>

The modelling of this kind of effects has been achieved by carrying a sophisticated intervention analysis (Box and Tiao, 1975), both for the type of dummy variables and the dynamic filters involved.

Figures 5 and 6 show that the distortions caused by the working conditions on the stationary transformation of the series are very important: figure 5 represents the stationary transformation of the original series and figure 6 shows this transformation applied to the series once it has been corrected for the effects of the working conditions, i.e., once the final estimates of them are obtained. The standard deviation of the series graphed on figure 5 is 0.0740, while for the corrected series is only 0.0212.

The working conditions are classified into three groups, which are discussed separately.

3.1 BANK HOLIDAYS

The main problem to be treated when modelling the effect of bank holidays is the reduced sample information, which entails estimating with few degrees of freedom. In order to increase the precision of the estimators, it was assumed that the effect of a bank holiday on the demand only depends on the day of the week where the holiday occurs, i.e., it does not depend on the specific holiday being celebrated: by taking advantage of this feature, the number of degrees of freedom could be augmented. However it was detected that this uniform treatment entailed bad forecasts for certain specific holidays, as it rendered important prediction

FIGURE 5.— DIFFERENCED ORIGINAL SERIES

(1983–1988)

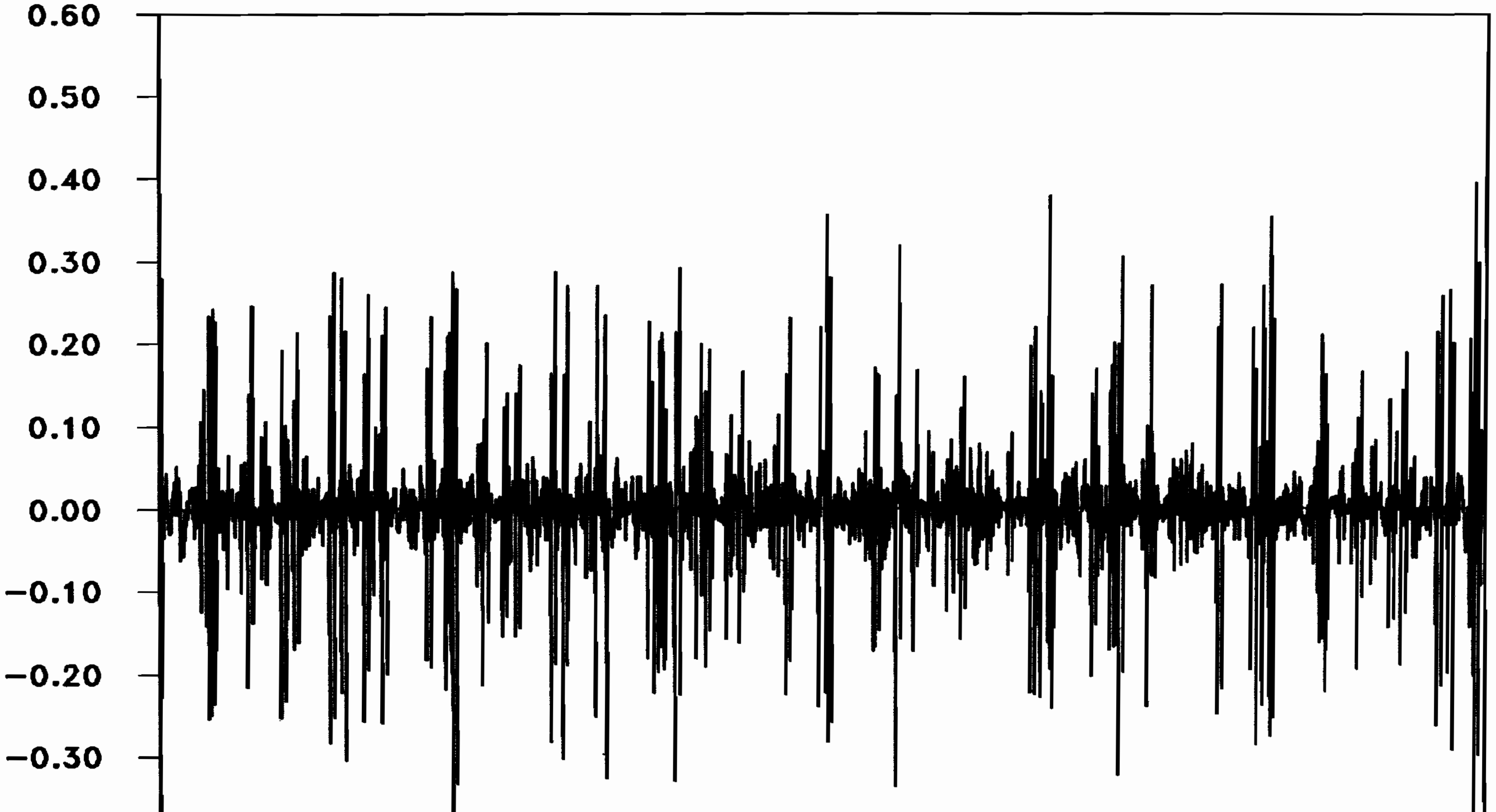
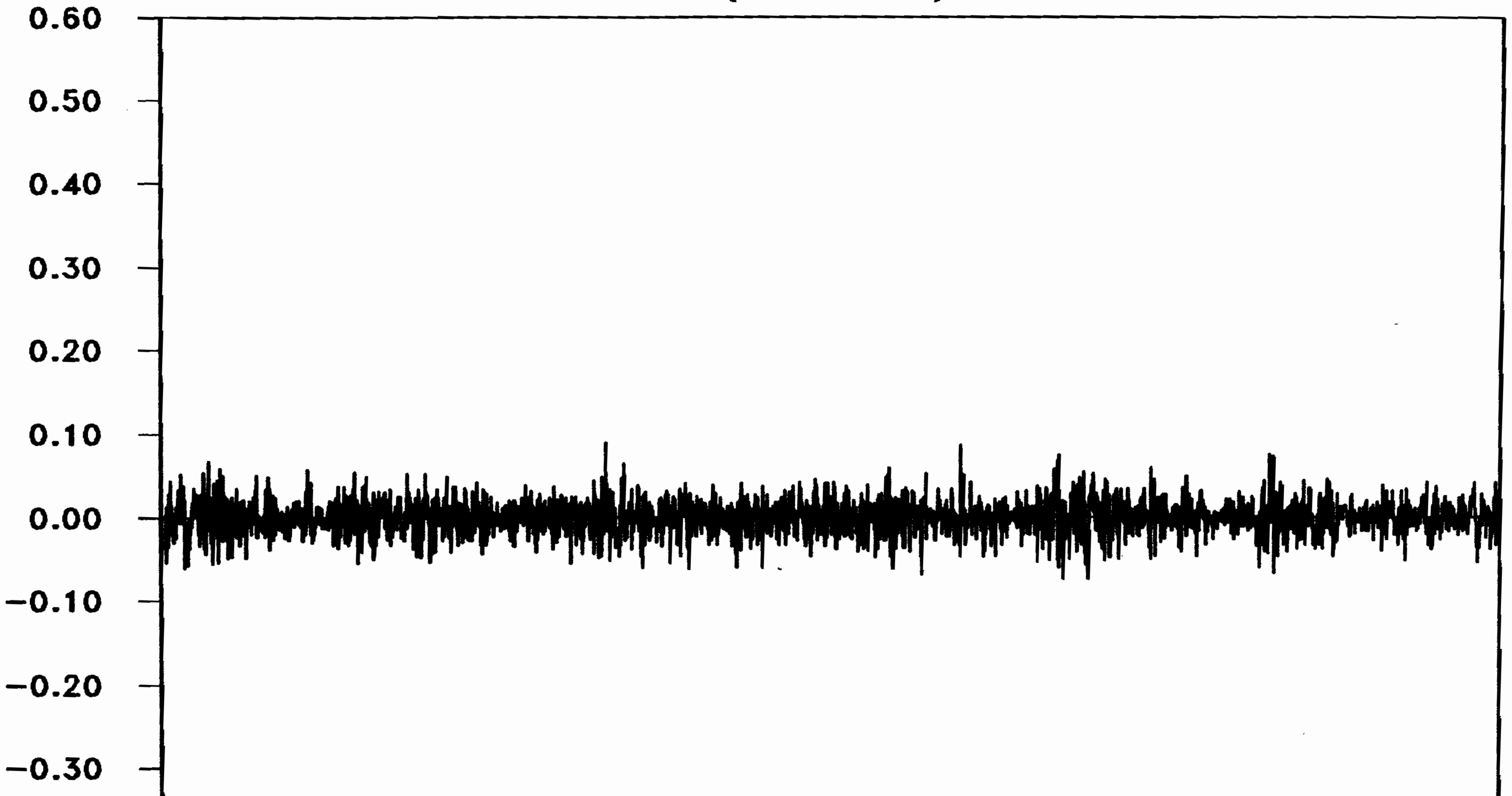


FIGURE 6.— DIFFERENCED CORRECTED SERIES

(1983–1988)



errors for all the years in the sample. So it was decided to distinguish two subgroups: general bank holidays and special ones.

3.1.1 General Bank Holidays

For these holidays an uniform treatment proves to be adequate; the main part of the holidays belong to this subgroup. Six dummy variables were used to model this type of holidays: MON (for those holidays that occur on Mondays), TUE (Tuesday), WED (Wednesday), THU (Thursday), FRI (Friday) and SAT (Saturday).

If these variables are to permit to include the effect of holidays in the model, it follows that they can not be of the 0 / 1 type, as there exists nationwide, regional and local holidays. As a consequence, and taking variable MON as an example, its value for day t is equal to:

- one, if t is a Monday and a holiday for the whole Spanish territory.[(4)]
- x, if t is a Monday and a holiday in part of the Spanish territory; the value x is obtained by dividing an average demand of the area affected by the holiday by the Spanish peninsular average demand, and rounding the result to the first decimal.[(5)]
- zero, in any other case.

Once the dummy variables have been defined, the next step is to consider their transfer function. The following issues have been taken into account:

a) the extra day effect: for a holiday occuring on a Tuesday (Thrusday), the demand for the Monday before (the Friday after) is smaller than what it would be if there were no holiday this week.

b) the post-holiday effect: the day after a holiday the whole demand is smaller than a normal day because the working activity tends to begin somewhat later. The only exception are the holidays occuring on Saturday.

Summarizing the above discussion, the dynamic effects that have been detected for each dummy are:

- MON: it affects Monday and Tuesday,
- TUE: Monday, Tuesday and Wednesday,
- WED: Wednesday and Thursday,
- THU: Thuersday, Friday and Saturday,
- FRI: Friday and Saturday,
- SAT: Saturday,

and the estimated coefficients are given in table 1.

3.1.2 Special Holidays

There are special holidays that do not follow the above scheme. This subgroup includes the following days:

- January 6: it will be discussed in section 3.2 that the demand is affected by the existence of a vacation period between Christmas Eve and New Year's Day. As January 6 is very near of this vacation period, the effect on the demand is different than it would be for a general holiday.
- May 1: It often is the first holiday of the year with a fine weather, so many people decide to take a small vacation out of their homes. Its effect on the demand is the greatest of all the Spanish holidays.
- August 15: as it will be argued in section 3.2, the monthly demand shows a minimum on August, as this is the month in which most Spanish people enjoy their summer vacation. So the reduction in the demand is the smallest.

Table 1: Estimated Coefficients for General Holidays

Day of the Holiday	Effect on MON	TUE	WED	THU	FRI	SAT
MON	-.2566	-.0409				
TUE	-.1017	-.2962	-.0333			
WED			-.2604	-.0398		
THU				-.2520	-.1083	-.0255
FRI					-.2544	-.0839
SAT						-.0795

These three holidays have been separatedly considered by including the variables:

- JAN6MON, JAN6TUE, ..., JAN6SAT
- MAY1MON, MAYITUE, ..., MAY1SAT, MAY1SUN
- AUG15MON, AUG15TUE, ..., AUG15SAT

The first four letters indicate the specific holiday they refer to, and the other three the day of the week where the holiday occurs.

To stress their specifity, note that for instance May 1 alters the demand even when it occurs on Sunday, an awkward result which was nevertheless very clear in the data. As in the initial sample at most only one non-zero observation of each variable of these dummy variables was available, it was tested if some cross-restrictions could be imposed: for example, a simmetry in the responses for a Tuesday and a Thursday. The restrictions that finally have been included in the model are given in the appendix B.

It was carefully checked that the special holidays, as well as the Christmas holidays treated in section 3.2.3, need this type of modelling. To base the estimates upon just one observation is not a good practice, but it seems to be better than to impose the same coefficient for holidays that really have very different effects on the demand. When making day-to-day real forecasts, these one-observation estimates have proved to be quite satisfactory.

3.2 VACATION PERIODS

Three vacation periods are considered: Easter, August and Christmas.

3.2.1 Easter

In Spain Friday, Saturday and Easter Sunday are holidays in the whole territory; and depending on the regions and varying from one year to another, either the Thursday before or the Monday after Easter Sunday is also a holiday. However there is a growing number of persons that take off work the whole week, so the demand for electricity falls since the Monday of the Holy Week until the Tuesday after Easter Sunday. So the following variables have been included: EASMON1, EASTUE1, EASWED, EASTHU, EASFRI, EASSAT, EASSUN, EASMON2 and EASTUE2: the first three letters refer to Easter and the other three / four to the day of the week.

All variables are binary ones excepting EASTHU and EASMON2; the value of EASTHU for day t is given by

- x if t is the Holy Thursday, where x represents the weight of the geographic area in which it is a holiday, and it is calculated in the way employed for the general holidays.
- zero, if t is not Holy Thursday.

EASMON2 is defined similarly.

3.2.2 August

In August the demand for electricity experiences two different types of modifications, as it can be seen in figure 2:

a) There is a general decrease on the level of the series, as the demand for any day of the week is smaller in August than in any other month.

b) There is a change on the weekly seasonal pattern: the within-week differences are reduced. If the deviations of each day of the week with respect to the weekly average are calculated, it is seen that they are very similar for Tuesday, Wednesday, Thursday and Friday, and much smaller than if the same deviations are calculated for any other month. Moreover, the deviations are less negative (smaller in absolute value) for Monday, Saturday and Sunday. These changes are not constant for the four / five weeks of August, as they are more pronounced for the central weeks of the month.

One added difficulty stems from the fact that August 1, the official date of the beginning of the vacation, does not always occur on a Monday. If August 1 happens to be, for example, a Wednesday, it is very likely that many people will advance their vacation to the previous weekend, so the distortions will affect the last days of July. August 31 poses a similar problem.

The way of dealing with this kind of effects can be summarized as follows:

a) It is considered that the vacation period of August stands for four or five weeks, depending on the days of the week in which August 1 and August 31 happen to occur.

b) Several truncated step variables are used to capture both the level and the weekly pattern changes. In order to clarify their definition, assume that August 1 turns to be on Saturday (so August 30 occurs on Sunday). All the following step variables will end on August 30 for this year:

- first week: one truncated step variable starting on Monday (it equals one from August 3); another one starting on Tuesday (from August 4); another one in Saturday (from August 8) and another one in Sunday (fom August 9).
- for each of the other weeks a similar set of four dummies have been employed starting in the corresponding day of the week in question.

c) These truncated step variables are denoted by AUGyxxx, where y = 1, 2, 3, 4 stands for the week and x = MON, TUE, SAT, SUN stands for the day of the week.

d) The main advantage of this parameterisation is to avoid high correlations among the estimators of their coefficients.

e) In those cases where five weeks are needed (for instance, when August 1 occurs on Thursday), the so-called first week in b) begins on the first Monday of August (in the above example, on August 5), and a previous "zero week" is added. The truncated step variables end on the Sunday of this week (August 4 in the example); these variables are denoted as AUG0xxx, where xxx is defined as in c).

Figure 7 shows the estimated effect on the demand for electricity when August 1 occurs on Saturday.

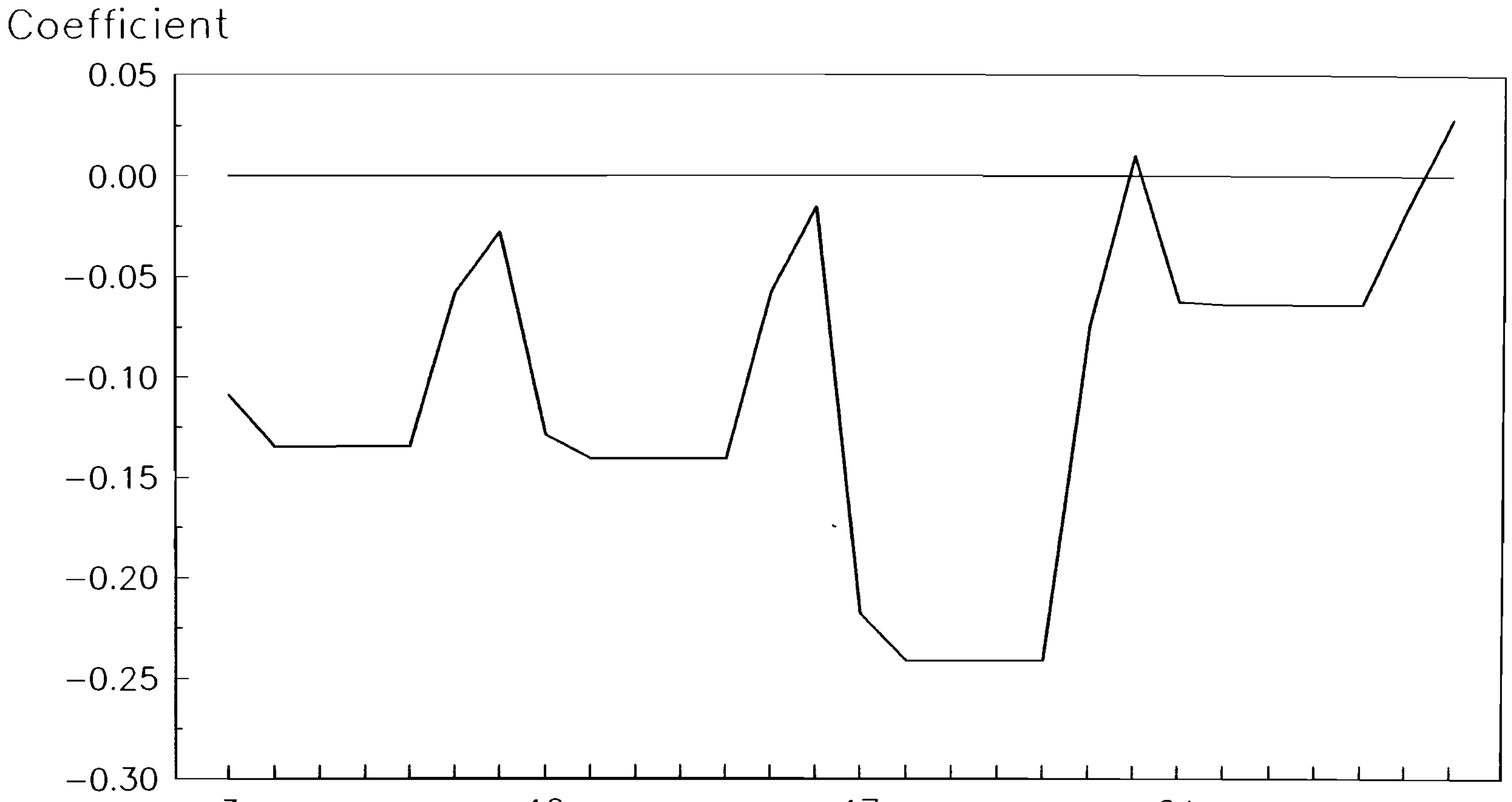
3.2.3 Christmas

Two types of effects can also be distinguished on Christmas: a) a general level shift, and b) the specific effect of the holidays of this period. The general decrease of the level is modelled by including a truncated step variable that equals one from December 24 to January 2 and zero elsewhere; it is denoted as XMAS. The specific effect of the holidays are modelled by taking into account the day of the week they happen to occur. The following variables are added to the model:

- DEC24MON, DEC24TUE, ..., DEC24SUN
- DEC31MON, DEC31TUE, ..., DEC31SUN

and in all cases they are given a dynamic filter of the form (w_0 + w_1.L). Moreover, extra days off work and post-holiday effects are considered when necessary by modyfing the filter. Whenever it was

FIGURE 7.– ESTIMATED EFFECT FOR AUGUST
(example)



possible cross restrictions have been imposed to increase the degrees of freedom.

3.3 OTHER ANOMALIES

Along the sample period several non-systematic changes in the working conditions have been detected, such that they have distorted the normal economic activity and so the demand for energy. These are multiple-causes changes: elections, general strikes, etc.; we shall not detail them, and their estimated effects are not shown in this paper.

Some other anomalies are related to the official hour adjustments of March and September: the last Sunday of March has officially 23 hours (variable DAY23), and the last Sunday of September 25 hours (DAY25). The effects on the daily demand of removing and adding one hour are estimated at -3.23% and 2.92% respectively.

4. EFFECTS OF THE WEATHER CONDITIONS ON THE DEMAND FOR ELECTRICITY: METHODOLOGY FOR ESTIMATION

The demand for electric energy is affected by the weather conditions, as: a) extreme temperatures start electric heating or cooling systems, and b) luminosity changes are compensated by using electric light.

So the following variables were selected as representative of the weather conditions: maximum and minimum daily temperature, solar light hours within a day and degree of cloudiness. Their specific definition is discussed in the data appendix (appendix A).

In this first section devoted to the weather conditions, the methodology for estimating their nonlinear dynamic effects will be considered, leaving the presentation of the final estimates for section five. We shall concentrate on the relationship between temperature and demand, as it happens to be both the most complex and the most useful for forecasting.

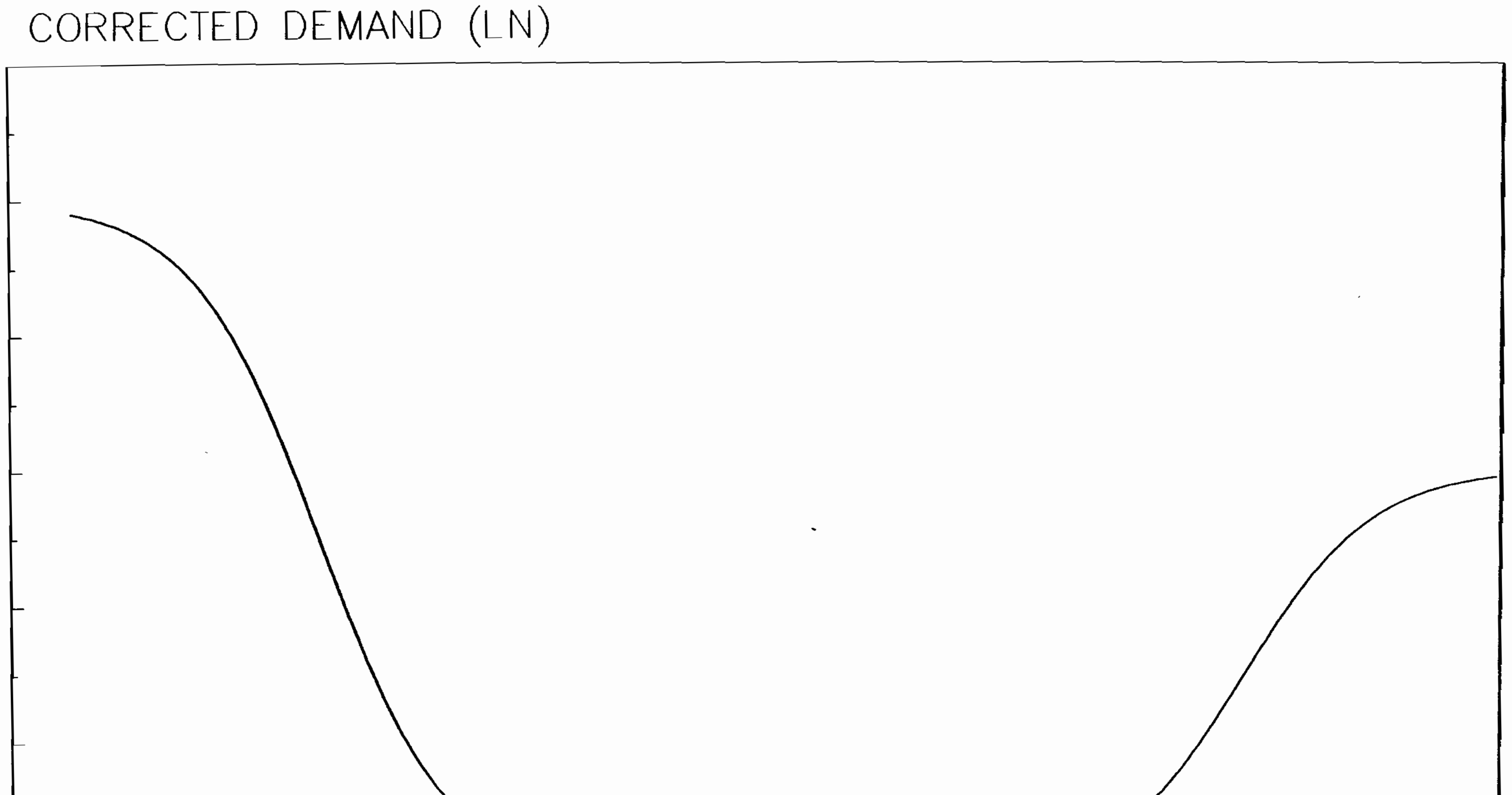
4.1 A PRIORI INFORMATION CONCERNING THE RELATIONSHIP BETWEEN TEMPERATURE AND DEMAND FOR ENERGY

From a theoretical viewpoint, the relationship between demnad for energy and temperature is represented on figure 8. The main features of this relationship are as follows:

a) There exist two temperatures, T^* and T^{**}, such that they define a neutral range, in which the temperature is nice and there is no reason to alter it artificially.

b) Below T^* a cold sensation is perceived: heating systems are started up, and the demand for energy increases, being the demand higher as the cold sensation grows (i.e., the lesser the temperature).

FIGURE 8.– SHAPE OF THE THEORETICAL RELATION
BETWEEN TEMPERATURE AND DEMAND FOR ELECTRICITY



c) Above T^{**} a hot sensation is perceived; the starting of cooling systems imply an additional demand for energy, and as the temperature rises the demand also rises.

d) The response function may be a non-linear one on both sides, specially on the cold zone.

This set of basic characteristics of the relationship will be referred as the Static Basic Effect. However with daily data this relationship is not strictly contemporaneous; on the contrary it can be decomposed as:

- a contemporaneous effect, that captures the influence on the demand for day t of the temperature of this same day t, and
- a previous history effect, that captures the influence of temperatures observed previously to t on the demand for t.

When this dynamic decomposition is considered, the so-called Dynamic Basic Effect is obtained.

There may exist some other additional effects that complement the Basic Effect:

a) Exhaustion Effect: there may exist a temperature low enough as to oblige that every heating system operate at their maximum capacity. As a consequence, if the temperature decreases further, it will have no additional effect on the demand. A similar effect may be found with high temperatures.

b) Season Effects: the influence of a given temperature may be different according to the season of the year in which it is registered, as it may be considered as a normal and permanent temperature or as an anomalous and transitory one.

c) Type-of-Day Effect: the response of the demand for a given temperature may change according to if it is registered on a working day or on a non-working day.

d) Increasing-Stock-of-Appliance Effect: for a small period of time this stock can be taken as a constant, but when several years are considered it may happen that the same temperature has different effects on the demand according to if it is registered at the beginning or at the end of the sample.

4.2 A METHODOLOGY FOR ESTIMATING THE RELATIONSHIP

It is no easy task to estimate the relationship between temperature and the demand, due to its complexity. As a consequence the following sequential strategy was considered:

1) Firstly only the Dynamic Basic Effect was considered, as once this effect is estimated then it is easy to enlarge the model. On the contrary, the specification of the Dynamic Basic Effect entails to combine a search of functional specification with a search of dynamic specification; moreover, both searches interact, in the sense that not necessarily the functional form is the same for all lags.

To estimate the Dynamic Basic Effect the dynamic threshold procedure described in Cancelo and Espasa (1990) was used. So a first estimation of the relationship was obtained, call it model A.

2) Model A entails a priori constraints on the other four effects enumerated above; so the next stages are to test if these constraints are or not rejected by the data. So, in a second step the issue considered was testing the exhaustion effect: to do so several candidates for exhaustion thresholds were considered, and specifications imposing this exhaustion were estimated. These models were then compared with model A.

Exhaustion effects, both for the cold and hot zones, seem to exist(6). Let this resulting model be referred to as Model B.

3) In the third stage the seasons of the year were defined, as it does not seem adequate to use the standard classification[7], and the corresponding temperature variables were generated.

Afterwards, a general version of model B that allowed for season effects was estimated, non-significant coefficients were deleted, and likelohood ratio tests were used to determine which constraints could be imposed on the inter-seasons response functions. The final conclusion is the existence of certain specific effects of the temperature depending on the season of the year.

4) The type-of-day effect was next considered. To do so the whole sample was divided into two groups, working and non-working days - see appendix A -, and the corresponding temperature variables were generated.

The final model of stage 3 imposes the effect of a given temperature to be the same for a working and a non-working day. So a more general model that does not impose this constraint was estimated, non-significant coefficients were eliminated and the restrictions of equal coefficients were tested. The final results point to response functions that vary according to the type of day.

It must be stressed that this type-of-day effect is designed in such a way that the effect of a given temperature depends on the type of the day we are considering, not on the type of the day in which this temperature was registered. As an example, let t be a working day and t+1 a non-working one; so the effect of a given temperature registered in t on the demand for t+1 will be determined by the lag one coefficient of the response function for non-working days.

5) Lastly, it was analyzed the imposing of a increase of the stock of appliances effect on the above model. This study focused mainly on the hot zone, as it was known that the sales of cooling systems in Spain experienced a boom during the second half of the eighties.

This was confirmed by the data: the final gain of the response function for the hot zone is greater for the final years of the sample than it was for the first years.

As a summary, a sequential methodology has been used; its starting point is a model which provides a consistent estimation of the dynamics and non-linearity, and additional effects are then added.

Once the final model has been obtained, a more general version, where the dynamics and the non-linearity were enlarged, was estimated in order to check that no significant threshold was drop out throughout the process. This sensibility analysis supported the proposed specification.

It may be argued that the particular order followed in this sequential testing procedure can influence the final result. Even if this is an open possibility, in this case the resulting model seems to be robust to changes in the order of stages 2-5.

It must be stressed that the main temperature effect has already been captured with the Dynamic Basic Effect; additional effects improve the model performance at very specific dates, but their general contribution is small. The model obtained at step 1 (model A), with only one meteorological variable, has a residual standard error of about 0.0137, while for the final specification this paper is devoted to - with four different meteorological variables - the same statistic equals 0.0130. This reduction can be better evaluated if it is taken into account that an univariate model with intervention analysis, which does not include any meteorological variable, has a residual standard error of 0.0161.

5. EFFECTS OF THE WEATHER CONDITIONS ON THE DEMAND FOR ELECTRICITY: FINAL ESTIMATES

5.1 MAXIMUM TEMPERATURE

As a proxy of the temperature for day t the maximum temperature corresponding to this day has been used. So the general discussion of the previous section was applied to estimate the relationship between maximum temperature and demand for electricity.

First of all, a neutral zone is detected between 20 C (68 F) and 24 C (75.2 F); temperatures within this range do not alter the demand for electricity. Values below 20 C are said to be in the cold zone, and above 24 C in the hot zone. In both cases their effects on the demand depends on the season of the year.

5.1.1. Cold Zone

A.1 Winter

a) There is an exhaustion temperature equal to 9 C (48.2 F), so all values below 9 C alter the demand in the same quantity.

b) The function is nonlinear between 9 C and 20 C; this nonlinearity is approximated by a piecewise linear function that can be decomposed in:

- a linear function between 14 C (57.2 F) and 20 C.
- a linear function between 11 C (51.8 F) and 14 C with a greater (in absolute value) slope than in the preceeding case.
- a linear funtion between 9 C and 11 C, with a lesser (in absolute value) slope than the other two functions.

c) Demand for day t is affected for temperatures ranging from t to t-7; the global lagged response is more important than the strictly contemporaneous one.

d) There is an interrelationship between dynamics and nonlinearity, as most of the nonlinearity of the response is concentrated on the first lag.

e) A distinction must be made between working and non-working days, as the demand is greater for the latter for a given temperature; Both contemporaneous and previous history effects differ according to the type of the day.

f) No appliances-stock-effect is detected.

A.2 Spring

For this season the response function is identical to that for winter in the range 11 C to 20 C; below 11 C no temperatures have been observed.

A.3 Summer

There are no observations below 14 C. In the range 14 - 20 C the response is linear and contemporaneous, with a much smaller effect on the demand than in the previous seasons. No type-of-day effect was detected.

A.4 Autumn

The same response function as for summer.

5.1.2. Hot Zone

B.1 Winter

Temperatures above 24 C do not alter electric demand.

B.2 Spring

The same result as for winter is found.

B.3 Summer

a) An exhaustion temperature is detected at 33 C (91.4 F), so higher temperatures do not provoke a further effect on the demand.

b) The function in the range 24 C - 33 C is linear.

c) There are contemporaneous, first-order-lagged and second-order-lagged effects.

d) The response function is the same for working and non-working days.

e) The increase in the stock of cooling systems shifts the response function over time, as the effect of a given temperature is greater from 1988 on than in previous years.

B.4 Autumn

The data point out that the same response function as for summer can be applied.

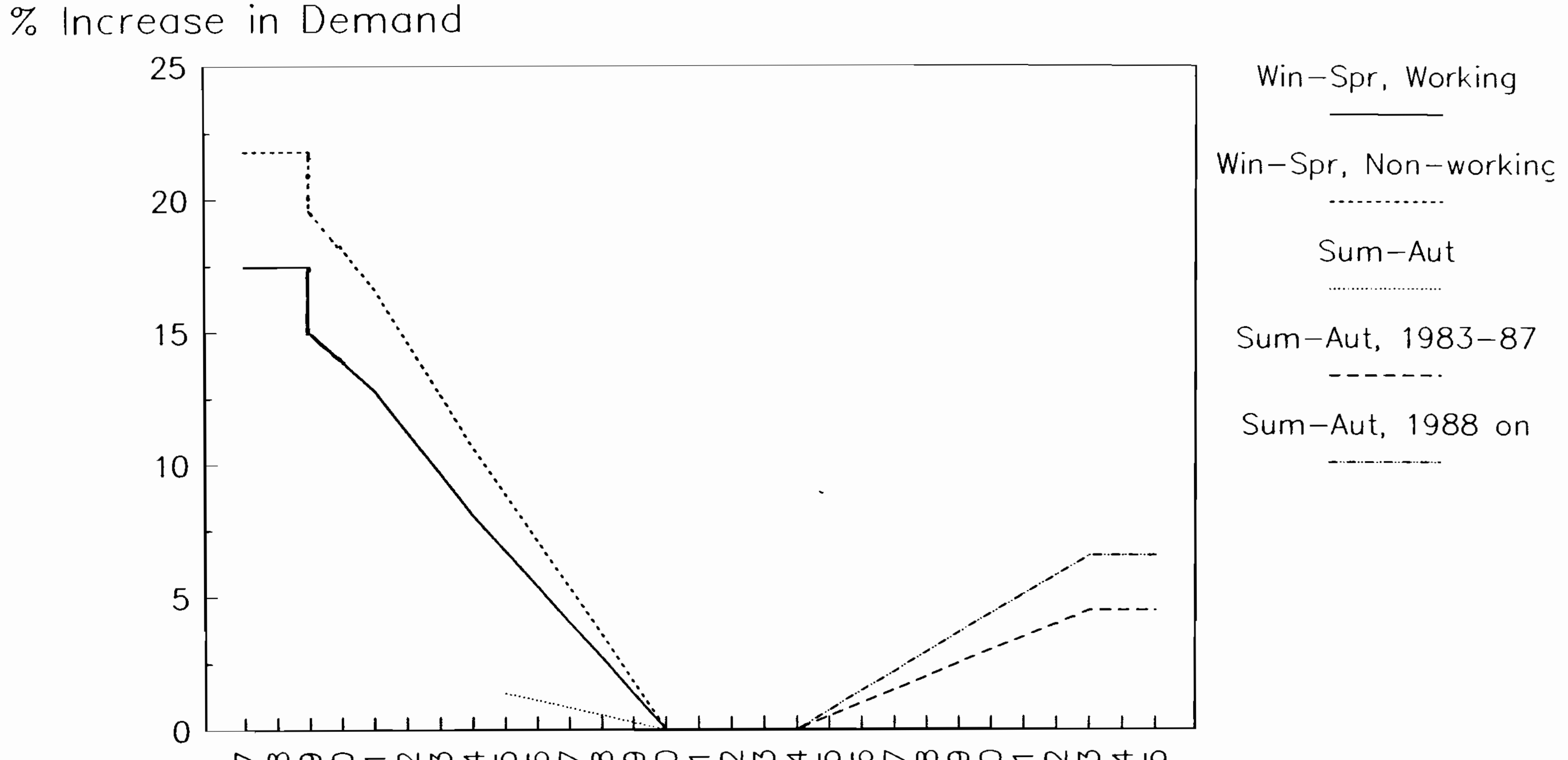
Figure 9 summarizes the final gain of each temperature on the demand. A more detailed information, concerning estimated coefficients and some figures that illustrate the shape and dynamics of the whole response function, is presented in appendix C.

5.2 TEMPERATURES DISPERSION

The relationship presented in subsection 5.1 is based upon taking the daily maximum temperature as a proxy for the temperature of t, and it implicitly assumes that this temperature has been constant for the whole day. However, for a given day the temperature is not constant, rather it oscillates. So a new variable is added to consider this oscillation.

The temperature dispersion for t, denoted by TEMDIS, is defined as the difference between maximum and minimum temperatures for t (8). The expected effect of this variable is conditioned upon the value of the maximum temperature:

FIGURE 9.- TEMPERATURE EFFECT
FINAL GAIN



- with a maximum temperature within the cold zone, the higher the dispersion the more marked the cold sensation and thus the higher the demand.
- with a maximum temperature within the hot zone, the higher the dispersion the less marked the hot sensation and thus the lesser the demand.

The main characteristics of the response function are:

1) TEMDIS has no influence when the maximum temperature belongs to the hot zone.

2) There are two different response functions according to the value taken by the maximum temperature:
- when the maximum temperature ranges between 14 C and 20 C, and
- when it is below 14 C. In this case the same value of the dispersion has a greater effect on the demand.

3) The response functions are linear in each range of variation of TEMDIS.

4) The dynamics are such that dispersion values corresponding to t, t-1 and t-2 influence the demand for t.

Appendix C also refers to this variable.

5.3 LUMINOSITY

To take into account the effect of the luminosity on the electric energy demand two indicators were available. The first proxy is a general one, that evolves in a cyclical deterministic way through the year: solar light hours within a day; the other one is a indicator related to the specific weather conditions for day t, the degree of cloudiness.

5.3.1 Solar Light Hours

This variable quantifies the number of solar light hours during normal working time; it displays a cyclical deterministic evolution oscillating from a maximum of 12 hours to a minimum of 9.2; it is denoted as SOLHOU.

Its effect is strictly contemporaneous and linear: the demand of a 9.2-solar-light-hours day with respect to a 12-hours day is estimated to be around 3.6% higher.

5.3.2 Degree of Cloudiness

The response function of this variable, which will be referred to as DEGCLO and whose values range from 0 to 8 - see appendix A -, it can be summarized as follows:

a) No dynamics are detected, being its effect strictly contemporaneous.

b) There is a season effect.

c) As for winter and spring, there is a nonlinear relationship, approximated by a piecewise linear function decomposed in:

- a linear function for values between 0 and 6
- a linear function, with a greater slope, for values between 6 and 8.

d) For summer and autumn the response function is linear for the whole range of variation of DEGCLO.

Appendix C also deals with these points.

6. STOCHASTIC PROPERTIES OF THE DISTURBANCE TERM AND MODEL EVALUATION

Once explanatory variables have been included, the disturbance term still shows a systematic evolution, whose modelling would contribute to improve the forecasts of the model. Simple and partial correlograms indicate that the short run residual dynamics can be modelled by an MA model; this moving average process turns to be very similar to that resulting for the univariate model with intervention analysis built in a previous stage of this research.

So the residual generating process is given by a MA (1,2) (7,14) (357,364,365,728,731,735) process; consequently the forecasts include a correction term depending on some past prediction errors. These are

- the errors for the two previous days, in order to introduce slightly adjustments on the trend term.
- the errors for the same day of the week of the two previous weeks, along with the corresponding satellites, to tune up the weekly seasonal, and
- the errors for the same day and neighbours of the two previous years, along with their satellites and leap years adjustments, to adapt the annual seasonal.

The estimated coefficients appear in appendix D. It must be stressed that these estimates are very similar to those obtained with the univariate model with intervention analysis, with the minor exceptions of the estimates of the coefficients Θ_1 and Θ_2 of the MA (1,2) process.

The model residuals have succeeded all the usual misspecification tests: zero mean, no autocorrelation, simmetry, kurtosis and normality. The innovations' standard error is equal to

0.0130, so the 95% confidence interval for the next observation is given by the corresponding one period forecast plus / minus 2.55% of this latter value.

To validate the model a residual analysis by subsamples was also carried out. The split of the sample was made according to the criteria: days of the week, months and years, taking only one of them at each time. In all cases (7+12+6 = 25) except one (the mean for the residuals corresponding to July) the null hypothesis of non-significant mean is not rejected. On the contrary, it seems to remain a slight heteroskedasticity: residual standard errors for different days of the week vary from 0.0120 to 0.0140, for months from 0.0098 to 0.0160 and for years from 0.0104 to 0.0152, as it can be seen in the appendix D.

As a consequence, it seems that further improvements could be achieved by modelling monthly heteroskedasticity. This heteroskedasticity may be due to the presence of nonlinear meteorological effects still not modelled, or to the restriction implicit on using a single model for the different days of the week, which can alter the estimated responses for the weather variables[9]. In any case, its influence on the final specification and estimates seems to be a minor one.

Several prediction tests were also performed to analyze post-sample stability, with no rejection of the stability hypothesis.

7. CONCLUSIONS

A complex model for forecasting daily electric demand has been specified and estimated. The main issues of this kind of forecasting problem seem to be the modelling of the changes in the working conditions and the influence of the weather conditions.

As for the changes in the working conditions, all extra-sample information is worth: a public holidays calendar, reflecting local, regional and general holidays, is a basic tool; another one is a daily record of the main events, which can provide hints to get the cause of an anomalous observation. Nor subjective memories are negligible, as experienced personnel achieve fine forecasts by relying almost solely on them.

This kind of analysis must be complemented with a careful study of the data. Data can help to define a dummy variable, to establish its dynamic filter and to impose restrictions that increase the degrees of freedom. As quoted in section 3, the stationarity transformation of the original series has a standard error equal to 0.0740, and the same statistic for this transformation of the series corrected by intervention analysis equals 0.0212.

The modelling of the influence of the weather conditions turns to be a hard exercise on functional and dynamic search for spcification. There are several effects to be considered, and a general methodology to handle them must be developped: starting from the approach presented in Cancelo and Espasa (1990), a generalisation is developped to include further effects. This methodology goes beyond this specific problem and is applicable to a larger set of daily data modelling.

Extra-sample information also deserves a central role. This kind of information must focus:

1) on the type of theoretical effects that are to be expected - see the discussion of section 4 on temperature - as these are to guide the empirical search, and

2) on the measurable magnitudes that can serve as proxy variables for the underlying meteorological phenomena - see the treatment of luminosity -.

Bearing in mind the types of responses we are searching for, and by applying the sequential strategy outlined in 4.2, the information contained in the data is efficiently used and good approximations of the response functions are obtained.

The adequacy of the model is assured by misspecification tests and by a sensibility analysis. Subsample analysis checks structural stability according to different classifications of the data; prediction tests center on post-sample stability, and sensibility analysis confirms the nonlinearity approximations and the lag responses. Both prediction tests and sensibility analysis are central to avoid the problem of data mining, which is likely to occur when using data crunching techniques.

The final model is being used since May 1989 to generate official forecasts of the demand for electricity. It has surpassed alternative forecasts, mainly based on subjective approaches. Moreover, it has been implemented on a computer in such a way that day-to-day forecasts and data loading can be managed by non-specialized personnel, allowing the experts to concentrate on other issues. When it is needed to modify the forecasts of the model due to the appearence of an anomalous fact, then:

1) the model can provide information about the effects previous anomalous facts have had on the demand, which in turn may help to quantify the expected influence of this distortion, and

2) the model can decompose its own forecast, detailing the effect of each factor it considers (maximum temperature, degree of cloudiness, etc); this information is valuable as it permits the experts to focus on quantifying the effect of the anomaly and then adding the usual effects.

FOOTNOTES

(1) A daily indicator of industrial activity was used to capture the trend but it was not successful.

(2) The logged variable is use to stabilize the variance. It is checked that the series has not been overdifferenced.

(3) It is to be expected that adding the explanatory variables contained in WOR and WEA will not alter the stationarity transformation, as these variables explain neither the trend nor the weekly seasonality. This belief was fully confirmed by the data.

(4) Even if the series being analyzed refers only to the peninsular Spanish demand, we shall refer to it as the Spanish demand to clarify the exposition.

(5) In order to establish these weights, annual data of the Spanish "provincias" corresponding to 1985 were used. Very similar weights would result if any other year were used.

(6) The adding of exhaustion effects does not provide a clear improvement with the respect to model A - which is very similar to the model reported in Cancelo and Espasa (1990) -; moreover, the shape of the response function for the cold zone had to be slightly adjusted to accommodate the exhaustion. As the forecasting performances are almost the same, and given that the model with exhaustion effects has more parameters, model A was preferred from

a statistical viewpoint. However, the model with exhaustion effects suited more the beliefs of the qualified personnel of REE, who advised its selection.

(7) In some areas of the inland Spain (like Madrid, that represents 10% of the total demand), heating systems are started at the beginning of november. See the operative definitions of the seasons in appendix A.

(8) This is equivalent to add to the model the daily minimum temperature, as a model which includes the maximum temperature and the temperature dispersion can be reparameterized so that the explanatory variables are both extreme temperatures. The advantage of the parameterisation based on the temperature dispersion is to obtain more ortogonal estimators of the temperature effects.

(9) Related research on daily peak models indicates that the daily demand at the peak hours needs to be modelled with one model for each day of the week.

REFERENCES

- Bogard, C., G. George, G.M. Jenkins and G. McLeod (1982): Analysing a Large Number of Energy Time Series for a Utility Company, in G.M. Jenkins and G. McLeod (eds), Case Studies in Time Series Analysis, volume 1, 113-149, Gwilym Jenkins and Partners Publications.

- Box, G.E.P. and G.M. Jenkins (1970): Time Series Analysis: Forecasting and Control, Holden Day.

- Box, G.E.P. and G.C. Tiao (1975): Intervention Analysis with Applications to Economic and Environmental Problems, Journal of the American Statistical Association, vol. 70, 70-79.

- Cancelo, J.R. and E. Espasa (1990): Dynamic Threshold Modelling: an Application to Daily Models, paper presented at the Sixth World Congress of the Econometric Society.

- Hesterberg, T.C. and A. Papalexopoulos (1988): Short Term Electric Load Forecasting Using Linear Regression, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 608-612.

- Hillmer, S. (1984): Forecasting Electric Loads, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 413-418.

- Jenkins, G.M. (1982): Some Developments of Transfer Function Analysis, in G.M. Jenkins and G. McLeod (eds), Case Studies in Time Series Analysis, volume 1, 170-194, Gwilym Jenkins and Partners Publications.

- Rhys, J.M.W. (1984): Techniques for Forecasting Electricity Demand, The Statistician, vol. 33, 23-33.

APPENDIX A: DEFINITIONS

DATA DEFINITION:

- Demand for Electric Energy: net Spanish demand corresponding to the peninsular territory; obtained by taking total production from all sources plus international interchanges balance less intermediate autoconsumption and pumping consumption.

- Maximum Temperature: weighted average of the maximum temperatures registered in ten selected weather stations throughout the Spanish part of the peninsula (almost half a million of squared kilometers), and supposed to represent their area's climate. The weights are equal to the quota of the demand for each area with respect to the total demand.

- Minimum Temperature: weighted average of minimum temperatures built in the same way as the maximum temperature.

- Solar Light Hours: number of solar light hours during normal working time, taken to be 8.00 a.m. to 8.00 p.m.; it is a purely cyclical deterministic variable as it evolves according to astronomical rules.

- Degree of Cloudiness: the observable magnitude it comes from is the local sky cloudiness, a meteorological variable which takes integer values from 0 (no clouds) to 8 (fully cloudy); the values of the local sky cloudiness are combined into an index obtained as discussed when presenting the maximum temperature.

PERIODS-OF-THE-YEAR DEFINITION:

- Classification by Seasons: the seasonal calendar used in this paper is:

* Winter: from November 1 to March 31.
* Spring: from April 1 to May 31.
* Summer: from June 1 to September 30.
* Autumn: from October 1 to 31.

- Classification by type-of-day: all days are working days except for the following, which are non-working days:

* Saturdays,
* Sundays,
* nationwide holidays,
* regional holidays if the demand for the affected area exceeds by 50% the total demand, and
* generalized days-off-work: if a certain Tuesday (Thursday) is a non-working day, then the correponding Monday (Friday) is also a non-working day.

APPENDIX B: INTERVENTION ANALYSIS

All the coefficients have been estimated with the sample 1983-1988, except otherwise stated.

Table B.1: Estimated Coefficients for January 6

Day of the Holiday	Effect on					
	MON	TUE	WED	THU	FRI	SAT
MON	-.2782	-.0254				
TUE	-.1003	-.2847	-.0379			
WED			-.2782	-.0254		
THU				-.2847	-.1003	
FRI					-.2847	-.1003
SAT(1)						-.1001

(1) Estimated with 1989 data.

Table B.2: Estimated Coefficients for May 1

Day of the Holiday	Effect on MON	TUE	WED	THU	FRI	SAT
MON(1)	-.3395	-.0667				
TUE	-.1142	-.3537	-.0539			
WED		-.0538	-.3537	-.0538		
THU				-.3080	-.1025	
FRI					-.3080	-.1025
SAT(2)						
SUN						-.0503

(1) Estimated with 1989 data.

(2) No sample information for this case in the period 1983-1990.

Table B.3: Estimated Coefficients for August 15

Day of the Holiday	Effect on					
	MON	TUE	WED	THU	FRI	SAT
MON	-.1809	-.0412				
TUE(1)	-.0588	-.2002	-.0405			
WED			-.1809	-.0412		
THU				-.1548	-.0416	
FRI					-.1548	-.0416
SAT						-.0693

(1) Estimated with 1989 data.

Table B.4: Estimated Coefficients for Easter

Variable	Coefficient
EASMON1	-.0065
EASTUE1	-.0198
EASWED	-.0301
EASTHU	-.2641
EASFRI	-.3432
EASSAT	-.1451
EASSUN	-.0646
EASMON2	-.3927
EASTUE2	-.0671

Table B.5: Estimated Coefficients for August

Variable Beginning on	Week				
	"Zero"	First	Second	Third	Fourth
MON	-.0505	-.1088	-.1013	-.2026	-.0727
TUE	-.0344	-.0258	-.0118	-.0236	-.0014
SAT	.0538	.0765	.0834	.1668	.0480
SUN	.0129	.0303	.0424	.0848	.0444

Table B.6: Accumulated Coefficients since Monday, First Week

Variable Beginning on	Week			
	First	Second	Third	Fourth
MON	-.1088	-.1291	-.2177	-.0624
TUE	-.1346	-.1409	-.2413	-.0638
SAT	-.0581	-.0575	-.0745	-.0158
SUN	-.0278	-.0151	.0103	.0286

Table B.7: Estimated Coefficients for the Christmas Period (1)

Day of the Holiday	Effect on						
	MON	TUE	WED	THU	FRI	SAT	SUN
MON	-.2296	-.4202	-.0713				
	-.2082	-.3546	.0046				
TUE	-.0845	-.1345	-.3853	-.0968			
	-.0318	-.1277	-.3253	-.0539			
WED			-.1345	-.3853	-.0968		
			-.1277	-.3253	-.0539		
THU				-.1345	-.3853	-.0968	
				-.1277	-.3253	-.0539	
FRI(2)							
SAT	-.0460					-.0498	-.0947
	.0294					-.0381	-.0423
SUN(3)	-.3126	-.0661					
	-.3391						
XMAS	-.0646						

(1) For each cell the first figure refers to the Christmas Day effect and the second one to the New Year Day effect. The day of the holiday is the day of the week in which December 24 / December 31 happen to occur.

(2) No sample information for this case in the period 1983-1990.

(3) Estimated with 1989 data.

APPENDIX C: RESULTS FOR THE WEATHER VARIABLES

Table C.1: Additional Demand (%) due to the Temperature:

Cold Zone, Winter / Spring, Working Days.

Temp. Value	Lag 0	1	2	3	4	5	6	7	Final Gain
19	0.49	0.25	0.22	0.13	0.12	0.07	0.04	0.02	1.34
18	0.97	0.50	0.45	0.25	0.23	0.15	0.09	0.05	2.69
17	1.46	0.74	0.67	0.38	0.35	0.22	0.13	0.07	4.02
16	1.95	0.99	0.89	0.51	0.46	0.30	0.18	0.09	5.37
15	2.43	1.24	1.11	0.63	0.58	0.37	0.23	0.12	6.71
14	2.92	1.49	1.34	0.76	0.69	0.45	0.27	0.14	8.05
13	3.41	1.97	1.56	0.89	0.81	0.52	0.32	0.17	9.65
12	3.90	2.45	1.78	1.02	0.93	0.59	0.36	0.19	11.22
11	4.38	2.94	2.00	1.14	1.04	0.67	0.41	0.22	12.80
10	4.87	2.94	2.23	1.27	1.16	0.74	0.45	0.24	13.90
9 and below	5.36	4.03	3.15	2.10	1.27	0.82	0.50	0.26	17.49

Table C.2: Additional Demand (%) due to the Temperature:

Cold Zone, Winter / Spring, Non-Working Days.

Temp. Value	Lag 0	1	2	3	4	5	6	7	Final Gain
19	0.53	0.32	0.29	0.20	0.16	0.11	0.07	0.08	1.76
18	1.06	0.64	0.58	0.40	0.32	0.23	0.15	0.17	3.55
17	1.59	0.96	0.86	0.60	0.48	0.35	0.23	0.25	5.32
16	2.12	1.28	1.15	0.80	0.63	0.46	0.30	0.33	7.07
15	2.65	1.60	1.44	1.00	0.79	0.58	0.38	0.41	8.85
14	3.18	1.92	1.73	1.20	0.95	0.69	0.45	0.50	10.62
13	3.71	2.47	2.02	1.40	1.11	0.81	0.53	0.58	12.63
12	4.24	3.03	2.30	1.60	1.27	0.92	0.60	0.67	14.63
11	4.77	3.59	2.59	1.80	1.43	1.04	0.68	0.75	16.65
10	5.30	3.66	2.88	2.00	1.59	1.16	0.75	0.83	18.17
9 and below	5.83	4.82	3.87	2.52	1.75	1.27	0.83	0.91	21.80

Table C.3: Additional Demand (%) due to the Temperature:

Cold Zone, Summer / Autumn.

Temp. Value	Increment
19	0.27
18	0.54
17	0.82
16	1.09
15	1.36

Table C.4: Additional Demand (%) due to the Temperature:

Hot Zone, Summer / Autumn.

Temp. Value	1983 to 1987 Lag 0	1983 to 1987 Lag 1	1983 to 1987 Lag 2	1983 to 1987 Final Gain	FROM 1988 ON Lag 0	FROM 1988 ON Lag 1	FROM 1988 ON Lag 2	FROM 1988 ON Final Gain
25	0.14	0.29	0.07	0.50	0.28	0.29	0.15	0.73
26	0.28	0.57	0.14	0.99	0.57	0.59	0.30	1.45
27	0.42	0.86	0.21	1.49	0.85	0.88	0.45	2.18
28	0.56	1.14	0.29	1.99	1.14	1.17	0.60	2.91
29	0.70	1.43	0.36	2.48	1.43	1.47	0.74	3.64
30	0.84	1.72	0.43	2.98	1.71	1.76	0.89	4.37
31	0.98	2.00	0.50	3.48	2.00	2.06	1.04	5.10
32	1.12	2.29	0.57	3.97	2.28	2.35	1.19	5.82
33 and above	1.26	2.57	0.65	4.47	2.57	2.65	1.34	6.55

Table C.5: Additional Demand (%) per Degree of Dispersion

Values of the Maximum Temperature	Lag 0	Lag 1	Lag 2	Final Gain
Between 14 C - 20 C	0.035	0.056	0.039	0.130
Below 14 C	0.035	0.089	0.039	0.163

Table C.6: Additional Demand (%) due to the Degree of Cloudiness

DEGCLO value	Winter / Spring	Summer / Autumn
1	0.171	0.070
2	0.342	0.140
3	0.513	0.210
4	0.684	0.280
5	0.855	0.350
6	1.026	0.420
7	1.497	0.490
8	1.970	0.560

FIGURE C.1.– TEMPERATURE EFFECT: LAG 0

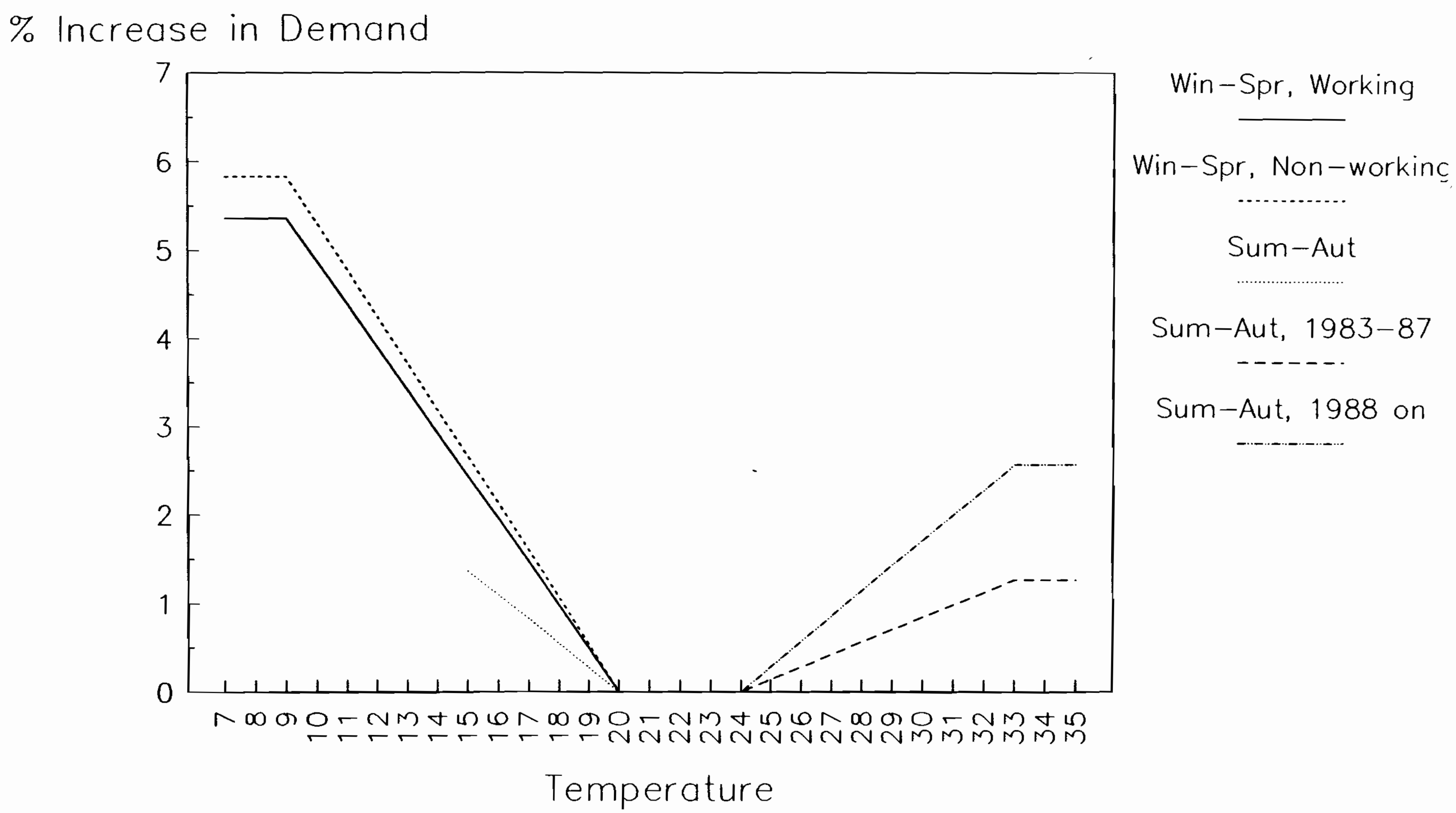


FIGURE C.3.- TEMPERATURE EFFECT: LAGS 3 TO 7

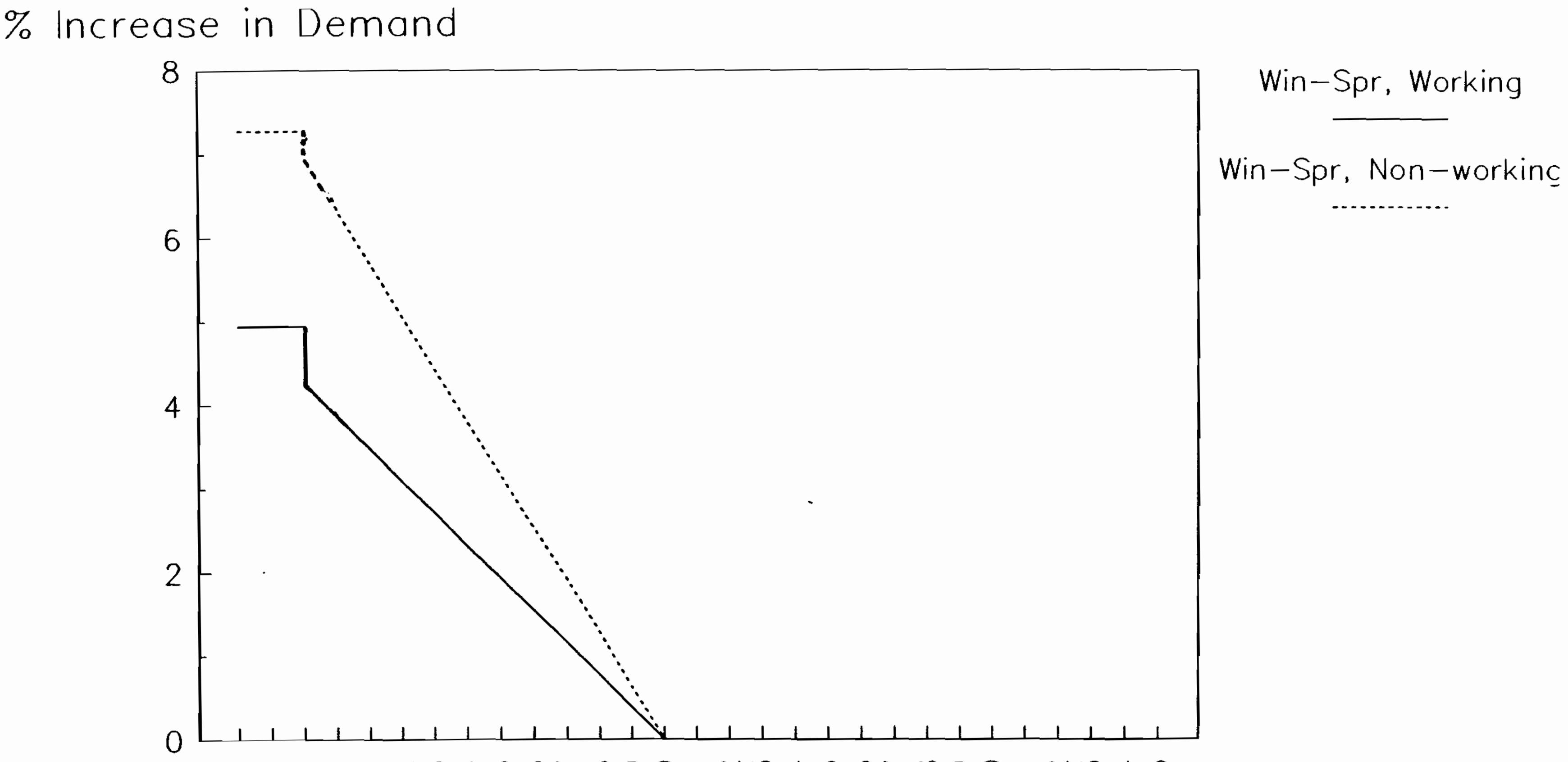
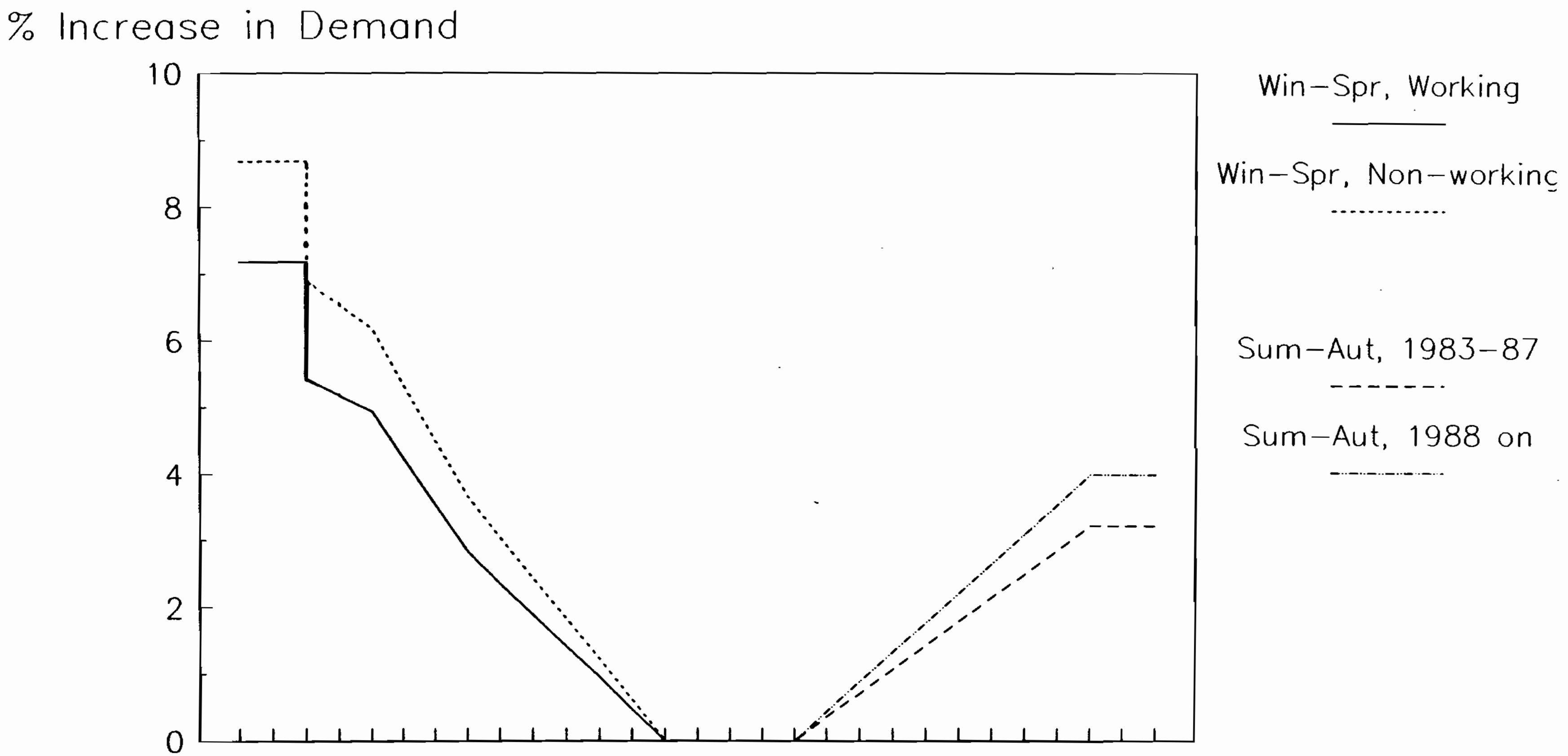


FIGURE C.2.- TEMPERATURE EFFECT: LAGS 1 AND 2



APPENDIX D: SHORT TERM RESIDUAL DYNAMICS

The estimated MA model for the residuals is given by

$$\underset{}{(1 - \underset{(0.02)}{0.20}\, L - \underset{(0.02)}{0.15}\, L^2)} \cdot (1 - \underset{(0.02)}{0.83}\, L^7 - \underset{(0.02)}{0.09}\, L^{14}) \cdot (1 +$$

$$+ \underset{(0.02)}{0.07}\, L^{357} + \underset{(0.02)}{0.15}\, L^{364} + \underset{(0.02)}{0.03}\, L^{365} + \underset{(0.03)}{0.10}\, L^{728} + \underset{(0.03)}{0.08}\, L^{731} +$$

$$+ \underset{(0.03)}{0.06}\, L^{735})$$

Standard deviations between parenthesis. Only one correlation above .75 in absolute value, the correlation between the estimators of Θ_7 and Θ_{14}.

Table D.1: Residual Standard Deviations by Days of the Week

Day	S.D. (%)
MON	1.34
TUE	1.20
WED	1.20
THU	1.24
FRI	1.40
SAT	1.30
DOM	1.40

Table D.2: Residual Standard Deviations by Months

Month	S.D. (%)
JAN	1.30
FEB	1.25
MAR	1.36
APR	1.42
MAY	1.31
JUN	1.29
JUL	1.19
AUG	1.38
SEP	1.27
OCT	1.08
NOV	0.98
DEC	1.60

Table D.3: Residual Standard Deviations by Years

Year	S.D. (%)
1983	1.52
1984	1.30
1985	1.39
1986	1.04
1987	1.29
1988	1.22

APPENDIX E: THE FINAL MODEL

This appendix is devoted to summarize the final model as a whole. We have choosed to present the last estimation, which has been carried out with the sample January 1, 1983 to October 31, 1990; so in general the estimates presented in this appendix will be slightly different from those reported throughout the paper. However there are some significant differences in the dynamic response to the maximum temperature, as the coefficients for lags six and seven turn out to be nonsignificant.

Before presenting the model, the operative definition of the weather conditions' variables must be discussed.

MAXIMUM TEMPERATURE VARIABLES:

Let T_t the maximum temperature for day t, measured in Celsius degrees x 10; then

$$C20_t = \begin{cases} 110 & \text{if } T_t \text{ is less or equal to } 90 \\ 200 - T_t & \text{if } T_t \text{ is between } 90 \text{ and } 200 \\ 0 & \text{if } T_t \text{ is greater or equal to } 200 \end{cases}$$

$$C14_t = \begin{cases} 50 & \text{if } T_t \text{ is less or equal to } 90 \\ 140 - T_t & \text{if } T_t \text{ is between } 90 \text{ and } 140 \\ 0 & \text{if } T_t \text{ is greater or equal to } 140 \end{cases}$$

$$C11_t = \begin{cases} 20 & \text{if } T_t \text{ is less or equal to } 90 \\ 110 - T_t & \text{if } T_t \text{ is between } 90 \text{ and } 110 \\ 0 & \text{if } T_t \text{ is greater or equal to } 110 \end{cases}$$

$$C9_t = \begin{cases} 110 & \text{if } T_t \text{ is less or equal to } 90 \\ 0 & \text{if } T_t \text{ is greater than } 90 \end{cases}$$

$$H24_t = \begin{cases} 0 & \text{if } T_t \text{ is less or equal to } 240 \\ T_t - 240 & \text{if } T_t \text{ is between } 240 \text{ and } 330 \\ 90 & \text{if } T_t \text{ is greater or equal to } 330 \end{cases}$$

Consider now the following dummy variables (for the definition of the different categories see appendix A):

$$DW_t = \begin{cases} 1 & \text{if t is a working day} \\ 0 & \text{otherwise} \end{cases}$$

$$DN_t = 1 - DW_t$$

$$DWS_t = \begin{cases} 1 & \text{if t is winter or spring} \\ 0 & \text{otherwise} \end{cases}$$

$$DSA_t = 1 - DWS_t$$

$$D8387_t = \begin{cases} 1 & \text{if t belongs to the interval (1-1-83, 31-12-87)} \\ 0 & \text{otherwise} \end{cases}$$

$$D88ON_t = 1 - D8387_t$$

Then the temperature variables used in the model are defined as (L denotes the lag operator):

$$C9WS_t = C9_t \cdot DWS_t$$

$$C11WS_t = C11_t \cdot DWS_t$$

$$C14WS_t = C14_t \cdot DWS_t$$

$$C20WSWi_t = (C20_t \cdot L^i) \cdot DWS_t \cdot DW_t \qquad i=0,1,\dots,5$$

$$C20WSNi_t = (C20_t \cdot L^i) \cdot DWS_t \cdot DN_t \qquad i=0,1,\dots,5$$

$$C20SA_t = C20_t \cdot DSA_t$$

$$HSA8387_t = H24_t \cdot DSA_t \cdot D8387_t$$

$$HSA88ON_t = H24_t \cdot DSA_t \cdot D88ON_t$$

TEMPERATURE DISPERSION VARIABLES:

Being $TMIN_t$ the minimum temperature for day t in Celsius degrees x 100, the dispersion variables are

$$TEMDIS20_t = \begin{cases} T_t - TMIN_t & \text{if } T_t \text{ is less or equal to 200} \\ 0 & \text{if } T_t \text{ is greater than 200} \end{cases}$$

$$TEMDIS14_t = \begin{cases} T_t - TMIN_t & \text{if } T_t \text{ is less or equal to 140} \\ 0 & \text{if } T_t \text{ is greater than 140} \end{cases}$$

LUMINOSITY:

Call DEGCLO the degree of cloudiness as in section 5; then if we define

$$DEGCLO6_t = \begin{cases} DEGCLO_t - 6 & \text{if } DEGCLO_t \text{ is greater or equal to 6} \\ 0 & \text{if } DEGCLO_t \text{ is less than 6} \end{cases}$$

the following variables are entered into the model

$$DEGCLOWS_t = DEGCLO_t \cdot DWS_t$$

$$DECLWS6_t = DEGCLO6_t \cdot DWS_t$$

$$DEGCLOSA_t = DEGCLO_t \cdot DSA_t$$

Moreover the solar hours variable that is considered in the model is defined as

$$TRSOLHOU_t = 12 - SOLHOU_t$$

where $SOLHOU_t$ was defined in section 5.

The model, as it is being used for forecasting, is summarized in table E.1; it must be remembered that, as discussed in section 2, all variables have been double differenced by the operators $(1-L)(1-L^7)$.

TABLE E.1.- SUMMARY OF THE FINAL MODEL (1) (2)

VARIABLE	LAG (3)	COEFFICIENT	T-RATIO
C9WS	1	0.7791E-04	2.85
	2	0.3697E-04	2.09
	3	0.3697E-04	2.09
C11WS	1	-0.3860E-03	2.00
C14WS	1	0.2166E-03	3.22
C20WSW0		0.4564E-03	18.89
C20WSW1		0.2328E-03	7.60
C20WSW2		0.1840E-03	7.34
C20WSW3		0.1062E-03	4.32
C20WSW4		0.7221E-04	2.99
C20WSW5		0.3862E-04	1.70
C20WSN0		0.5320E-03	16.82
C20WSN1		0.2464E-03	6.10
C20WSN2		0.2318E-03	6.69
C20WSN3		0.1655E-03	4.58
C20WSN4		0.2831E-04	0.80
C20WSN5		0.1251E-03	3.99
C20SA		0.2767E-03	2.70
HSA8387	0	0.1489E-03	4.25
	1	0.2732E-03	8.22
	2	0.6961E-04	2.13
HSA88ON	0	0.3124E-03	6.70
	1	0.3188E-03	6.87
	2	0.1176E-03	2.56
TEMDIS20	0	0.4226E-04	4.59
	1	0.5209E-04	5.65
	2	0.2362E-04	2.62
TEMDIS14	1	0.3774E-04	3.22
DEGCLOWS		0.1483E-02	7.11
DCLWS6		0.2419E-02	3.29
DEGCLOSA		0.7400E-03	3.28
TRSOLHOU		0.1072E-01	2.80
MON	0	-0.2622	53.95
	1	-0.0424	8.84
TUE	-1	-0.1011	20.01
	0	-0.2907	50.85
	1	-0.0347	6.98
WED	0	-0.2521	53.99
	1	-0.0349	6.89
THU	0	-0.2498	69.93
	1	-0.1109	27.45
	2	-0.0227	6.40
FRI	0	-0.2547	51.41
	1	-0.0772	16.08
SAT	0	-0.0792	21.39
JAN6MON	0	-0.2908	27.92
	1	-0.0281	2.75
JAN6TUE	-1	-0.1306	12.30
	0	-0.2962	24.61
	1	-0.0435	4.11
JAN6WED	0	-0.2535	24.82

	1	-0.0186	1.83
JAN6THU	0	-0.2962	24.61
	1	-0.1306	12.30
JAN6FRI	0	-0.2616	61.69
	1	-0.0445	10.03
JAN6SAT	0	-0.1001	9.82
MAY1MON	0	-0.3395	33.95
	1	-0.0667	6.67
MAY1TUE	-1	-0.1206	16.22
	0	-0.3700	43.90
	1	-0.0663	8.99
MAY1WED	0	-0.3111	26.36
	1	-0.0641	6.15
MAY1THU	0	-0.3124	30.81
	1	-0.1149	11.35
MAY1FRI	0	-0.3004	29.95
	1	-0.0924	9.24
MAY1SUN	0	-0.0464	7.45
AUG15MON	0	-0.1877	25.51
	1	-0.0446	6.24
AUG15TUE	-1	-0.0588	5.55
	0	-0.2002	16.85
	1	-0.0405	3.88
AUG15WED	0	-0.1675	23.46
	1	-0.0331	4.64
AUG15THU	0	-0.1579	15.72
	1	-0.0489	4.83
AUG15FRI	0	-0.1528	15.12
	1	-0.0307	2.99
AUG15SAT	0	-0.0675	7.43
EASMON1		-0.0074	1.84
EASTUE1		-0.0177	3.68
EASWED		-0.0266	5.27
EASTHU		-0.2912	35.88
EASFRI		-0.3489	65.72
EASSAT		-0.1503	28.74
EASSUN		-0.0624	12.49
EASMON2		-0.4150	42.71
EASTUE2		-0.0771	9.59
AUG0MON		-0.0582	10.20
AUG0TUE		-0.0362	6.28
AUG0SAT		0.0562	9.75
AUG0SUN		0.0224	3.89
AUG1MON		-0.1043	22.99
AUG1TUE		-0.0234	5.17
AUG1SAT		0.0757	16.60
AUG1SUN		0.0328	7.16
AUG2MON		-0.0980	26.26
AUG2TUE		-0.0142	3.83
AUG2SAT		0.0830	22.62
AUG2SUN		0.0409	11.13
AUG3MON		-0.1960	26.26
AUG3TUE		-0.0284	3.83
AUG3SAT		0.1660	22.62
AUG3SUN		0.0818	11.13
AUG4MON		-0.0649	14.17

AUG4TUE		-0.0012	0.27
AUG4SAT		0.0438	9.69
AUG4SUN		0.0358	7.97
DEC24MON	0	-0.2278	19.76
	1	-0.4205	33.70
	2	-0.0777	7.26
DEC24TUE	-1	-0.0929	8.61
	0	-0.1556	11.74
	1	-0.3877	29.48
	2	-0.0835	7.63
DEC24WED	0	-0.1080	9.16
	1	-0.3552	27.77
	2	-0.1259	11.32
DEC24THU	0	-0.1366	11.91
	1	-0.4114	32.90
	2	-0.1053	9.76
DEC24SAT	-1	-0.0367	4.76
	0	-0.0426	4.45
	1	-0.0889	9.60
	2	-0.0421	5.43
DEC24SUN	-1	-0.0395	2.97
	0	0.0169	1.22
	1	-0.3126	23.26
	2	-0.0661	5.93
DEC31MON	0	-0.2158	19.85
	1	-0.3678	29.28
	2	-0.0042	0.37
DEC31TUE	-1	-0.0464	4.19
	0	-0.1533	11.50
	1	-0.3402	24.78
	2	-0.0222	1.83
DEC31WED	0	-0.1170	10.82
	1	-0.3050	24.24
	2	-0.1049	9.00
DEC31THU	0	-0.1243	11.47
	1	-0.3263	25.80
	2	-0.0313	2.72
DEC31SAT	0	-0.0228	2.98
	1	-0.0446	4.95
	2	0.0427	4.94
DEC31SUN	1	-0.3391	27.08
XMAS		-0.0667	11.15
DAY23		-0.0386	10.19
DAY25		0.0313	8.24
THE1	1	0.1433	7.42
THE2	2	0.0976	5.05
THE7	7	0.8018	41.20
THE14	14	0.0845	4.23
THE357	357	-0.0407	2.04
THE364	364	-0.1120	5.59
THE365	365	-0.1029	5.20
THE728	728	-0.0534	2.49
THE731	731	-0.0331	1.55
THE735	735	-0.0726	3.37

RESIDUAL STANDARD ERROR = 0.012081
R-SQUARE = 0.995

NOTES:
(1) All variables have been double differenced.
(2) Non-systematic changes in the working conditions are not shown.They add nine more variables and fourteen more coefficients to the model.
(3) Zero except if otherwise stated.