The Philippine Statistician
1990, Vol. 39, Nos. 1-2 & 3-4
pages 39-51

THE USE OF SEMIPARAMETRIC MODEL IN TIME SERIES ANALYSIS

by

Erniel B. Barrios
Assistant Professor, Statistical Center
University of the Philippines

Tirso Eduardo C. Diaz
Instructor, Statistical Center
University of the Philippines

Abstract

The Box-Jenkins technique is a popular non-structural approach of analysing time series data. Box and Tiao(1975) adopted the technique in analysing the structural change in the series motivated by some intervening factors. In this paper, we introduce the semiparametric model and discuss how this can be used in the analysis of time series. The data on peso-dollar exchange rate is used to illustrate the proposed model.

Keywords: Semiparametric Model, Penalized Least Squares, Time Series, Intervention, Structural Change

1. INTRODUCTION

The term structural change brings to mind two notions in economics. It can be interpreted in either the temporal or nontemporal context. In the temporal context, the "change" is brought about by altering the environment where the relationship holds. In such, time is usually considered as one of the independent variables. An example is the peso-dollar exchange rate. Various political and economic phenomena surely affect its rise-and-fall.

In the non-temporal context, the "change" is brought about when an independent variable reaches a specified level. An example for this is the relationship between yield of palay and the amount of rainfall. The occurence of drought as indicated by a very low rainfall volume results in a low yield of palay. As rainfall increases, the yield would also increase until a time when a very high volume of rainfall

associated with natural disasters such as flood will damage existing crops including palay. Consequently, this would reverse the pattern of the relationship between palay yield and rainfall.

In this paper, we consider only the structural change in the temporal context. The "alteration in the environment" (which induces change) is called intervention. The intervention is expected to produce an instantaneous and continuing effect to the system until a new intervention occurs.

Box and Tiao(1975) used the Autoregressive Moving Average(ARMA) model(s) together with some dummy variables in the analysis of the effects of some intervention in the time series. Poirier(1973) used cubic splines to test for the significance of a structural change. In both approaches, time is the only independent variable incorporated in the model.

The semiparametric model introduced in the next section is proposed to solve the same problem. One of the advantages of this is that it can accomodate other independent variables aside from time.

Thus the technique is formulated in a broader perpective than the previously mentioned techniques.

2. THE SEMIPARAMETRIC MODEL

Given n observations on the dependent variable y and on the (p + 1) independent variables x_1, x_2,..., x_p, t. A routinary problem in model-building is to determine and estimate the structure which relates the dependent variable to the independent variables. The simplest solution is to assume that y linearly depends on x_1, x_2, ...,x_p, t. However, in some cases, this assumption would not give good fit. Thus, an alternative structure is desired.

Suppose that we are willing to keep the linear functional relationship between y and x_1, x_2, ...,x_p. On the otherhand, we do not want to take the risk of commiting on a functional form of the dependence of y on t. Consequently, we should resolve for a nonparametric form for the dependence of y on t.

The model can then be written as:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + f(t_i) + \epsilon_i$$

$$= \mathbf{x}_i'\boldsymbol{\beta} + f(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n \qquad (1)$$

$$E(\epsilon_i) = 0, \quad V(\epsilon_i) = \sigma^2, \quad i = 1, 2, \ldots, n$$

$$E(\epsilon_i \epsilon_j) = 0 \qquad i = j = 1, 2, \ldots n$$

$\mathbf{x}_i$ is the vector of independent variables

$\boldsymbol{\beta}$ is the vector of unknown parameters

f is an unknown function that belongs to the class

$W_2^{(2)}$ where $W^{(2)} = \{f: f, f^{(1)}$ are absolutely continuous and f is square integrable.$\}$

t is a variable in which y depends through an unknown function.

Model (1) in a more compact form becomes

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}(t) + \boldsymbol{\epsilon} \qquad (2)$$

where $\mathbf{Y} = \left[y_1, y_2, \ldots, y_n \right]'$

$$\mathbf{X} = \left[\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n \right]'$$

$$\mathbf{F}(t) = \left[f(t_1), f(t_2), \ldots, f(t_n) \right]'$$

Remark: *In model (2), if $\boldsymbol{\beta} = 0$, it reduces to the nonparametric model while if f is equal to a constant, then it reduces to a parametric model. Thus the term semiparametric model is applied to model (2).*

The problem is now to estimate the vector of parameters β and the unknown function f Nevertheless, situation arise when not all values of f in the entire range of t are desired. In this case, the estimation procedure can be simplified if the function f is discretized and thus reduce the nonparametric f into a parametric function specified as follows:

Suppose further that the points $t_0', t_1', \ldots, t_s'$ $(s < n - p)$ are known.

Define $\delta_j = f\left[\dfrac{t_{j-1}' + t_j'}{2}\right]$, $j = 1,2,\ldots,s$.

Let

$$I(t) = \begin{bmatrix} I_1(t_1) & I_2(t_1) & \ldots & I_s(t_1) \\ I_1(t_2) & I_2(t_2) & \ldots & I_s(t_2) \\ \vdots & \vdots & & \vdots \\ I_1(t_n) & I_2(t_n) & \ldots & I_s(t_n) \end{bmatrix}$$

$$I_j(t_i) = \begin{cases} 1 & \text{if } t_{j-1}' \le t_i < t_j' \\ 0 & \text{otherwise} \end{cases}$$

$$\delta = (\delta_1, \delta_2, \ldots, \delta_s)$$

Substitution of δ in place of F in model (2) results in

$$Y = X\beta + I(t)\,\delta + \epsilon \qquad (3)$$

The problem of estimating β and the nonparametric function f thus reduces to estimating β and δ .

The objective function used is $\dfrac{1}{n} \| Y - X\beta - I(t)\,\delta \|$

$+ K\| V\delta \|$, where $\|\cdot\|$ is the Euclidean norm. (4)

Remark: *The quantity (4) is called the penalized least squares.criterion.*

The matrix $V_{(s-2\ xs)}$ in the second term of (4) is the second differencing operator with the following elements

$$v_{ij} = \begin{cases} \dfrac{1}{(q_{i+1} - q_i)(r_{i+1} - r_i)} & i = j \\ -\dfrac{1}{(q_{i+1}-q_i)(r_{i+1}-r_i)} - \dfrac{1}{(q_{i+2}-q_{i+1})(r_{i+1}-r_i)} & j=i+1 \\ \dfrac{1}{(q_{i+2} - q_{i+1})(r_{i+1} - r_i)} & j = i + 2 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where} \quad q_i = \frac{t_{i-1}' + t_i'}{2} \qquad i = 1,2,\ldots,S$$

$$r_i = \frac{q_i + q_{i+1}}{2} \qquad i = 1,2,\ldots,S-1$$

The procedure and some results given in this section were based on Engle, Granger, Rice and Weiss (1986).

THEOREM 1 Given the model $\mathbf{Y} = \mathbf{H}\theta+\epsilon$

where $\mathbf{H} = (\mathbf{X},\mathbf{I}(t))$, $\theta' = (\beta', \delta')$

$$\mathbf{U}_{(s-2)x(p+s)} = (\mathbf{0} , \mathbf{V}).$$

Then $$\hat{\theta} = \left[\frac{\mathbf{H}'\mathbf{H}}{n} + \mathbf{K}\,\mathbf{U}'\mathbf{U}\right]^{-1} \frac{\mathbf{H}'\mathbf{Y}}{n} \qquad (5)$$

minimizes the penalized least squares criterion given in (4).

Remark: *Since we have linear functions as argument in the two norms that appeared in (4), theorem 1 can be proven by applying simple calculus as in ordinary least squares. It can also be shown (Barrios, 1990) that* $\hat{\theta}$ *is a consistent estimator of* θ *under mild conditions.*

THEOREM 3 (Distribution of θ)

Suppose $\epsilon \sim N(0, \sigma_\epsilon^2 I)$. Then

$$\hat{\theta} \sim N(\mu, \sigma_\epsilon^2 \Sigma)$$

where $\mu = \left[\frac{H'H}{n} + K U'U \right]^{-1} \frac{H'H}{n} \theta$

$$\Sigma = \left[\frac{H'H}{n} + K U'U \right]^{-1} \frac{H'H}{n^2} \left[\frac{H'H}{n} + K U'U \right]^{-1} \qquad (6)$$

The proof is straightforward from observing that θ is a linear combination of Y which is normally distributed with mean $H\theta$ and variance $\sigma_\epsilon I$.

The optimal value of the smoothing parameter K can be obtained both graphically and analytically. However, for our purpose, we will assume that a reasonable value of K is available.

The simultaneous estimation procedure we have just discussed requires the definition of matrices of large dimensions which obviously complicates computation of the estimates. The cummulative error in inverting matrices is proportional to the size of the matrix. Thus, inversion of larger matrices would incur larger error than small matrices. Furthermore, the method requires another objective function in the determination of the smoothing parameter. In this section, we discuss the method which minimizes the objective function (4) in three stages. This method provides an outright estimate for K. Though the method we will discuss next would require matrices with smaller dimension, it can not be used if all the independent variables are categorical.

Denote the objective function (4) by $Q(\beta, \delta, K)$, i.e.

$$Q(\beta,\delta,K) = \frac{1}{n} \| Y - X\beta - I(t)\,\delta \|^2 + K \| V\delta \|^2$$

The steps in obtaining estimates of β, δ and K are outlined as follows:

i) Differentiate $Q(\beta,\delta,K)$ with respect to β and equate to 0; i.e. $\dfrac{Q(\beta,\delta,K)}{\beta} = 0.$ (10)

Denote by $\hat{\beta}(\delta, \ )$ the solution of (10). Note that this is a function of δ and K.

ii) Substitute $\beta(\delta,K)$ in $Q(\beta,\delta,K)$, denoted by $\hat{Q}(\beta,\delta,K)$ which is a function of δ and K. Then differentiate Q with respect for δ and equate to 0 , we have

$$\frac{\hat{Q}(\hat{\beta},\delta,K)}{\underset{\sim}{\delta}} = 0 \qquad (11)$$

Denote by $\hat{\delta}(K)$ the solution of (11).

iii) Substitute $\hat{\delta}(\ )$ in $\hat{Q}(\hat{\beta},\delta,K)$ denoted by $\hat{Q}(\hat{\beta},\hat{\delta},\ )$ a function of K alone. Then differentiate Q with respect to K and solve for K . Denote the solution by $\hat{K}$.

THEOREM 4.

Define $P = I - X(X'X)^{-1}X'$, $A = I(t)PI(t)(V'V)^{-1}$

$$C_1 = \frac{1}{n^3} Y' PI(t) (V'V)^{-1} I'(t) PI(t) (V'V)^{-1} I'(t) PY$$

$$C_2 = \frac{1}{n^5} Y' PI(t) (V'V)^{-1} A' I'(t) PI(t) A(V'V)^{-1} I(t) PY$$

$$C_3 = \frac{1}{n^2} Y' PI(t) (V'V)^{-1} I'(t) PY$$

$$C_4 = \frac{1}{n^3} Y' PI(t) A(V'V)^{-1} I'(t) P^{V}$$

$$C_5 = \frac{1}{n^4} Y' PI(t)(V'V)^{-1} I'(t) PI(t) A(V'V)^{-1} I'(t) PY$$

$$C_6 = \frac{1}{n^4} Y' PI(t)(V'V)^{-1} A'(V'V) A(V'V)^{-1} I'(t) PY$$

$$C_7 = \frac{1}{n^3} Y' PI(t)' (V'V)^{-1} (V'V) A(V'V)^{-1} I'(t) PY$$

Assume that a positive root of the equation

$$(2C_3) K^3 + (-2C_1 - 4C_4 + 4C_7) K^2 + (6C_5 - 3C_6) K = 4C_2$$

exists and denote it by $\hat{K}^*$.

Let $$\delta = \left[I'(t)PI(t) + n \hat{K}^* V'V \right]^{-1} I'(t) PY$$

$$\hat{\beta} = (X'X)^{-1} \left[X'Y - X'I(t) \hat{\delta} \right]$$

then $\hat{K}^*$, $\hat{\delta}$, and $\hat{\beta}$ minimizes $Q(\beta,\delta,K)$

The proof is accomplished by following the steps given before Theorem 4. Mild conditions can be imposed to show consistency.

3. SEMIPARAMETRIC MODEL IN TIME SERIES ANALYSIS

Suppose Y is a variable which we have observed over time say, Y_t, t =1,2,...,T. We refer to $\{Y_t\}$ t= 1, 2, ...,T as the time series. It can be represented by the general structural model given by

$$Y_t = \mu_t + S_t + \epsilon_t \ , \ t = 1, 2, \ldots T.$$

Y_t is the t^{th} observation, μ_t is the trend component, S_t the seasonal component and ϵ_t the irregular component. The dependence of the mean and variance of Y_t on time is indicated

by the trend component μ_t. Thus, for a stationary process, $\mu_t = 0$.

The Autoregressive Moving Average (ARMA) model assumes a known parametric structure of the trend component. One identifies the parametric structure by looking at the correlogram of the series. In practice, we compute the empirical correlogram and compare this to the theoretical correlogram of known models to identify the structure of the given data set. In doing so, some confusion may ensue and result in the incorrect identification of the underlying parametric structure.

Thus, to avoid the risk of incorrect identification of the parametric structure of the trend component (assuming the seasonal component is 0), we propose the nonparametric specification of the trend component say

$$Y_t = f(t) + \epsilon_t \qquad (2)$$

where f is a smooth function of time.

To account for the seasonal effect of this type, we adopt some dummy variables. Suppose we have a quarterly seasonal period, we use three dummy variables defined as follows:

$$X_{1t} = \begin{cases} 1 \text{ if } t \text{ is a first quarter} \\ 0 \text{ otherwise} \end{cases}$$

$$X_{2t} = \begin{cases} 1 \text{ if } t \text{ is a second quarter} \\ 0 \text{ otherwise} \end{cases}$$

$$X_{3t} = \begin{cases} 1 \text{ if } t \text{ is a third quarter} \\ 0 \text{ otherwise} \end{cases}$$

Model (2) now becomes

$$Y_t = f(t) + \beta_1 X_{1t} + \beta_2 X_2 + \beta_3 X_{3t} + \epsilon_t \qquad (3)$$

Furthermore, if other variables denoted by Z_t are known to influence Y, model (3) can be written as (4)

$$Y_t = f(t) + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + Z_t \alpha + \epsilon_t \qquad (4)$$

which is clearly a semiparametric model. β_1, β_2 and β_3 are the corresponding seasonal indices (relative to the fourth quarter). We can test the significance of the β's to test for the significance of the seasonal variations.

To generalize the above formulation to any length of seasonality, simply use dummy variables numbering to one less than the number of seasons.

The simplest form of intervention is the one where the effect (structural change) in the series is a constant and instantaneous over time (or until the next intervention). For instance, suppose there are 3 interventions which are expected to induce structural change in the series. Then we define 3 intervention variables as:

$$W_{1t} = \begin{cases} 1 & \text{if } t \leq T_1 \\ 0 & \text{otherwise} \end{cases}$$

$$W_{2t} = \begin{cases} 1 & \text{if } T_1 \leq t \leq T_2 \\ 0 & \text{otherwise} \end{cases}$$

$$W_{3t} = \begin{cases} 1 & \text{if } T_2 < t \\ 0 & \text{otherwise} \end{cases}$$

Then model (4) can be modified as

$$Y_t = f(t) + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + Z_k'\underline{\alpha} + \beta_1^* W_{1t} + \beta_2^* W_{2t} + \beta_3^* W_{3t} + \epsilon_t \qquad (5)$$

β_1^* is the structural increase in the mean of the series due to the first intervention, β_2^* is the structural change in the mean due to the second intervention given that the first intervention has already occured. This may also include the cumulative effect of the first intervention. β_3^* can be interpreted similarly. However, if one suspects that the structural change in the slopes of the covariates and in the seasonal indices may exist, the cross-product of W's and X's and W's and Z's can be included in model (5).

4. APPLICATION TO PESO-DOLLAR EXCHANGE RATE

The exchange rate from Oct. 1, 1985 - Apr. 17, 1986 was analyzed using the Box & Tiao approach by Ridao(1987). The following events were assumed to affect structural change

1. The murder of Evelio Javier on Feb. 11, 1986.
2. The announcement of civil disobedience by Mrs. Aquino on Feb. 17, 1986 and
3. The restoration of the writ of habeas corpus on March 2, 1986 by President Aquino.

The following are the intervention variables used:

$$W_{1t} = \begin{cases} 1 \text{ if Feb. 11, 1986 } < t \leq \text{ Feb. 17, 1986} \\ 0 \text{ otherwise} \end{cases}$$

$$W_{2t} = \begin{cases} 1 \text{ if Feb. 17, 1986 } < t \leq \text{ March 22, 1986} \\ 0 \text{ otherwise} \end{cases}$$

$$W_{3t} = \begin{cases} 1 \text{ if } t > \text{ March 22, 1986} \\ 0 \text{ otherwise} \end{cases}$$

A 5-day interval or equivalent to 1-week of trading appears to be the logical way of discretizing f. Thus, each week is considered as one interval. A program was written in turbo pascal implementing the two approaches discussed in section 2. However, only the first one was used because no covariate was included in the model. Only the intervention variables were used in the parametric component. With = 1 the estimates of the parameters are given in table 1.

Consider $\beta_1^* = 0.8223112$. This means that the murder of Mr. Javier influenced the decline of the peso value by an average of about 82 cents relative to the US dollar. The announcement of civil dis-obedience by Mrs. Aquino further triggered the decline of the peso value to an average of about 2.88 pesos relative to the US dollar. Finally, the joint effect of the three interventions cut the fall in the peso value induced by the first two interventions by about P 1.

To check the fit of the model to the data, the mean squared prediction error was computed to be 0.003898703. The standard error is 0.0622439 or the average error in estimating the exchange rate from the model is about 6 cents.

If one aims to use the model for forecasting purposes, other economic variables should also be included and the second approach in estimating the parameters be used.

References:

1. Barrios, E. B. (1989). *Splines in Nonparametric and Semiparametric Regression.* Tech. Report # 89-03, Statistical Center, University of the Philippines.

2. Barrios, E. B. (1990). Semiparametric Regression Models, Unpublished, Ph.D. dissertation.

3. Box and Tiao (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *JASA*, 70.

4. Engle, R., Granger, C., Rice, J. and Weiss, A. (1986). Semiparametric Estimates of the Relation .Between Weather and Electricity Sale. *JASA*, **81**, 310-320.

5. Heckman, N. E. (1986). Spline Smoothing in a Partly Linear Model. *JRSS* ser.B, **48**, 244-248.

6. Poirier (1973). Piecewise Regression Using Cubic Splines. *JASA*, **68**, 515-525.

7. Ridao (1987). *Modelling the Daily Peso-Dollar Exchange Rate from Oct. 1, 1985 to Apr. 17, 1986 by Intervention Analysis.* Unpublished Master's Thesis, UP Statistical Center.