ESTIMATION OF SEMIPARAMETRIC ECONOMETRIC TIME-SERIES MODELS WITH NON-LINEAR OR HETEROSCEDASTIC DISTURBANCES

Francisco Javier Hidalgo Moreno

Submitted for the Degree of Ph. D. in Economics

November 1989

London School of Economics and Political Science

ERRATA

Page 1, line 11 should read "considerable" instead of "considerably".

Page 2, line 6 should read "intrinsic" instead of "intrisic".

Page 11, line 20 should read "stationarity" instead of "stationary".

Page 12, line 3 should read "Judge" instead of "Jugde".

Page 13, line 21 should read "is the use" instead of "is to use".

Page 14, line 16 should read "make use of" instead of "emploit" and in line 19 should read "have been" instead of "has been".

Page 20, line 13 should read "$F_T(x)$ by $F(x)$" instead of "$F(x)$ for $F_T(x)$".

Page 26, the last term of expression (1.6) should be multiplied by

$$\sum_{j=1}^{T} h(y_j)K((x_o-x_j)/a).$$

Page 28, equation (1.7) should be the expression:

$$\sum_{i=1}^{T} h(y_i)W_{T,i}(x).$$

Page 29, the equation of definition 2.1.4 should be

$$\hat{y}_T(x)=\sum_{i=1}^{T} y_i W^{*}_{T,i}(x).$$

Page 30, line 5 should read "depends on" instead of "is conditional to".

Page 37, line 4 should read "principle" instead of "principal".

Page 40, in equation (3.4), by "$Var[P_{o,n}-P_{1,n}]$" we mean "total variation of the difference between $P_{o,n}$ and $P_{1,n}$".

Page 41, the second integral sign in line 7 and 9 should be removed.

Page 60, line 12 should read "behaviour of the estimator of β" instead of "behaviour of β".

Page 61, in the second line of lemma 3.3.1 should read "$\sqrt{T}|\beta_T-\beta|$" instead of "$\sqrt{T}(\beta_T-\beta)$".

Page 62, the integrals in definition 3.3.2 are evaluated in the set A_T.

Pages 65, 66 and 67, θ_T and θ should be removed wherever they appear, and therefore the two paragraphs which follow equation (5.7) in page 67 should be omitted.

Page 68, in definition 3.5.3 should read "$S_{u\gamma}$" instead of "$S_{u\mu}$", and in line 8 should read "extent" instead of "extend".

Page 72, line 6 should read "$a=O(T^{-1/3\lambda})$" instead of "$a=O(T^{-1/3^{\lambda}})$".

Page 84, the second line of (A.11) should read:

"$\le Ma^{(1-\alpha)-r\alpha}\int|K^{(r)}(v)|^{\alpha}dv$" instead of "$\le Ma^{(1-\alpha)-r\alpha}\int|K^{(r)}(v)|\, dv$".

Page 109, line 2 should read "i) plim $^1/_T \sum_t Ev_t^2|\psi_{t-1}=\sigma^2$" where ψ_{t-1} is the information set up to time t-1, and line 4 "ii) lim $^1/_T\Sigma v_t I(|v_t|>\varepsilon\sigma\sqrt{T})=0$".

Page 110, the second inequality of (B.3) should read "$\le {}^1/_T \sup_t E((E_{t-1}\bar{v}_t^2-\bar{v}_t^2))^2$".

Page 116, the L.H.S. of the last equation should read

"$E[\frac{\partial^2}{\partial\beta^*\partial\beta^{*\prime}}\varphi_1(\beta,\beta^*)\Big|_{\beta^*=\beta}]$" instead of "$E[\frac{\partial^2}{\partial\beta^*\partial\beta^{*\prime}}\Big|_{\beta^*=\beta}]$".

Page 143, line 12 should read "that is" instead of "that it".

Page 149, the R.H.S. of equation (2.7) should be multiplied by $\dot{\sigma}_t^2$ and $\dot{\sigma}_t^2$ should be added to the R.H.S. of equation (2.8).

ABSTRACT

This thesis proposes and justifies parameter estimates in two semiparametric models for economic time series. In both models the parametric component consists of a linear regression model. The nonparametric aspect consists of relevant features of the distribution function of the disturbances. In the first model the disturbances follow a possibly non-linear autoregressive model, with autoregression function of unknown form. In the second model the disturbances are both linearly serially correlated and heteroscedastic, the serial correlation and heteroscedasticity being of unknown form. For both models estimates of the regression coefficients of generalized least squares type are proposed, and shown to have the same limiting distribution as estimates based on correct parameterization of the relevant features of the disturbances. Monte-Carlo simulation evidence of the finite sample performance of both estimates is reported.

ACKNOWLEDGEMENTS

I am very grateful to my supervisor Prof. P.M. Robinson who during these years has patiently helped and encouraged me in the realization of this Ph. D. thesis. Also, I am in debt -morally and intellectually- for many fruitful discussions with him. Further, I am very grateful to my wife Beatriz Arias who during these years patiently put up with my ups and downs.

Finally, I am grateful for the financial support of two scholarships from Bank of Spain, and to the research funded by the Economic and Social Research Council (ESRC) reference number: B00232156, that without their financial support, the realization of this Ph. D. thesis would not have been possible.

TABLE OF CONTENTS

TABLES

CHAPTER 4.

CHAPTER 1

INTRODUCTION

This thesis is concerned with precise estimation of time series linear regression models in two semiparametric settings. Throughout it is assumed that we observe a scalar y_t and a k-dimensional column vector x_t at time points $t=1,2,\ldots,T$. These variables are connected by a regression model of the form:

$$(1.1) \qquad y_t=\beta' x_t+u_t \qquad t=1,2,\ldots$$

where u_t is an unobservable disturbance term such that

$$(1.2) \qquad E[u_t|x_s;\ s=1,2,\ldots]=0 \quad \text{a.s.} \qquad \text{for all } t$$

and where β' is the transpose of a kx1 column vector β whose estimation is of interest.

The linear regression model is of considerably theoretical and practical interest in statistics and econometrics. With a view to econometric applications to time series data our specification is of somewhat limited use because it precludes the possibility of lagged dependent variables y_{t-j} in x_t. In view of the interest in non-linear regression models the linearity of the specification (1.1) is also restrictive. Moreover the scalar character of y_t and lack of allowance for any simultaneity equation structures limits the economic relevance of the model. However, we have chosen the model (1.1) not merely because of its intrisic interest, but because the methodology and theory that we develop is complicated and as the simplest interesting case (1.1) is convenient for expository purposes. It seems that our results should extend fairly straightforwardly to non-linear regression, multivariate regressions, and non-dynamic linear simultaneous equation systems. Some extensions to linear or non-linear regression models containing lagged dependent variables should also be possible.

Our description so far of the model (1.1) only partly specifies the data-generating mechanism. On asymptotic efficiency grounds, we seek estimates of β within the class of generalized least squares (G.L.S.) estimates. This requires something to be said about the variances and autocovariances of u_t. We assume that they are not generated by a finite-parameter model, but by a nonparametric one. Two different such models are studied in this thesis, each of which manifests some behaviour not covered by the literature on regression estimation to data, and entails substantially different methodology and theory.

1.1.- NON-LINEAR AUTOREGRESSION MODEL

The first model for u_t is the possibly non-linear first order autoregression (AR(1))

$$u_t=\rho(u_{t-1})+\epsilon_t, \quad t=1,2,\ldots. \tag{1.3}$$

The function ρ is of unknown form and is such that u_t is at least stationary, with independent and identically distributed (i.i.d.) zero mean and finite variance innovations ϵ_t. We also assume that the process $\{u_t\}$ has zero mean and is independent of the sequence $\{x_t\}$, thereby satisfying (1.2). There has been a good deal of interest in non-linear time series models in recent years, such as non-linear moving averages (Robinson, 1977), bilinear models (Granger and Andersen, 1978), and several forms of non-linear autoregression functions as (e.g. Jones, 1978, Haggan and Ozaki, 1980, Tong and Lim, 1980). The model (1.3) is of the latter type, but unlike these authors we specify no functional form for ρ, since this form is likely to be hard to identify in view of the many possibilities once linearity is abandoned.

If u_t is non-Gaussian, but has finite second moments and satisfies suitable regularity conditions, linear modelling is sufficient for the calculation of the G.L.S. estimates, since this only involves second moments of u_t. On the other hand it is not generally the case that the autocovariances of u_t implied by (1.3) even for a given non-linear ρ can be obtained, and they will certainly not have the same structure of the autocovariances generated by a linear AR(1) process. Therefore, there is some interest in employing the non-linear model (1.3) in forming estimates of G.L.S. type. Whereas G.L.S. estimates based on linear time series models tend to have the same asymptotic efficiency as full Gaussian maximum likelihood (M.L.) estimates, this does not seem to be true for our estimates based on a non-linear model, and we have not been able to show that our estimates necessarily improve over linear-in-y estimates. However, we present some favourable Monte-Carlo evidence of finite-sample performance of our estimate. In addition to asymptotic distribution theory the model (1.3) seems a natural starting point for the introduction of nonlinear disturbance structures, in view of the historical importance of the linear AR(1) in regression estimation. It seems relatively straightforward to extend our work to non-linear AR(P) models

$$u_t=\rho(u_{t-1},\ldots,u_{t-P})+\epsilon_t,$$

for some given finite P, when ρ is assumed of unknown form. Extension of our methodology to other non-Markovian or non-linear models seems less feasible because a finite AR transformation cannot be used. Another possibility that we do not employ is that u_t is generated by a linear model with unknown distribution function for the innovations, when G.L.S. estimates can certainly be improved upon.

When the function ρ is known, a G.L.S.-type estimate of β would be

$$(1.4)\qquad \tilde{\beta}=\arg\min_{b}\sum_{t=2}^{T}(y_t-b'x_t-\rho(y_{t-1}-b'x_{t-1}))$$

minimizing over some set of admissible values for β. Because we assume that the functional form of ρ is unknown, it is appropriate to insert smoothed nonparametric estimates of ρ in (1.4), defining another estimate $\bar{\beta}$. Under regularity conditions $\bar{\beta}$ is just as efficient asymptotically as $\tilde{\beta}$ that is based on a known ρ. We can thus say that $\bar{\beta}$ "adapts" for the unknown β learning from the data. In fact the estimate that we study is not the extremum estimate $\bar{\beta}$, but an estimate $\hat{\beta}$ given by one step Gauss-Newton type algorithm towards $\bar{\beta}$ commencing from the ordinary least squares (O.L.S.) estimate of β. However $\hat{\beta}$ has the same asymptotic efficiency as $\bar{\beta}$ and $\tilde{\beta}$.

The combination of the parametric regression model (1.1) and the nonparametric model (1.3) for the disturbances can be called a "semiparametric" model. Correspondingly, an estimate of β that uses nonparametric estimation of the nuisance function ρ is termed a "semiparametric" estimate.

1.2.- LINEAR HETEROSCEDASTICITY MODEL

The second model for u_t in (1.1) is of the form

(2.1) $$u_t = \sigma(x_t)v_t$$

(2.2) $$v_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \qquad \sum_{j=0}^{\infty} |\alpha_j| < \infty.$$

Here ε_t is again a sequence of i.i.d. zero mean and finite variance random variables. They are assumed independent of the sequence $\{x_t\}$, therefore so are the v_t, although the u_t are not independent of the x_t, the condition (1.2) is again satisfied. Neither the α_j in (2.2) nor the function σ in (2.1) are assumed to be known functions of parameters. Thus the model (2.1) and (2.2) for u_t can be termed nonparametric, and the combination of (1.1), (2.1) and (2.2) is again a semiparametric model.

The model (2.1), (2.2), where the u_t are both serially correlated and heteroscedastic conditional on $\{x_t\}$, combines two familiar non-standard features of disturbance behaviour. Even for the case that the α_j and σ are both parametrically modelled or when only one of them is, the literature contains little discussion. Evidence of either or both serial correlation and heteroscedasticity in disturbances is frequently found in econometric data and it seems appropriate to consider allowing for both possibilities in estimating β. In general economic theory seems unlikely to lay down any parametric model for the α_j or σ, and the presence of both features is likely to make the identification of the parametric model especially difficult. Thus, there seems to be grounds for treating the α_j and σ as nonparametric functions, especially if they are only "nuisance" functions and not of intrinsic interest.

Our specification of v_t implies that they are covariance stationary, so that the autocovariance matrix of the vector $v=(v_1,\ldots,v_T)'$ is of the Toeplitz form $\Gamma=(\gamma_{i-j})$, where

$$\gamma_i=E(v_t v_{t+i})=\sum_{j=0}^{\infty}\alpha_j\alpha_{j+i}$$

and where the variance of ϵ_t is equal to 1 with no loss of generality. The covariance matrix of $(u_1,\ldots,u_T)'$ is

$$\Omega=\sigma\Gamma\sigma$$

where

$$\sigma = \mathrm{diag}\{\sigma(x_1), \ldots, \sigma(x_T)\}.$$

Given knowledge of the α_j and σ^2, and thus of Γ and σ, we could form the G.L.S. estimate

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{2.3}$$

where

$$X = (x_1, \ldots, x_T)' \text{and } Y = (y_1, \ldots, y_T)'.$$

Given only a finite parameterization of Γ and σ we could insert estimates of the unknown nuisance parameters in (2.3) and obtain an estimate of β, denoted by $\bar{\beta}$, with the same asymptotic efficiency as $\tilde{\beta}$. We shall achieve the same asymptotic efficiency as $\tilde{\beta}$ and $\bar{\beta}$ by an estimate $\hat{\beta}$ that employs nonparametric estimation of Γ and σ. In fact we do not estimate the α_j or γ_i directly, but instead we adopt a frequently domain approach, entailing nonparametric estimation of the spectral density function of v_t. Again, we may term our estimate "semiparametric".

It must be stressed that (2.1) and (2.2) are only one way of describing serial correlation and heteroscedasticity. Other models are possible, for example, u_t might be modelled by a linear process where the innovations ϵ_t have variance depending on x_t. Mention must also be made of the ARCH models that describe conditional heteroscedasticity and serial dependence.

1.3.- PARAMETRIC ESTIMATES OF β

In order to place our work in some perspective it is appropriate to review the literature on the estimation of β in the cases of purely parametric modelling of the autocovariance structures of the u_t.

By far the most popular method of estimating β has been O.L.S.. This is in part due to its computational simplicity and in part due to its good statistical properties under suitable conditions. When the u_t are uncorrelated and homoscedastic, and X has full rank, O.L.S. is the best linear unbiased (BLU) estimate of β, attaining the Gauss-Markov bound. Moreover, when u_t is also normally distributed, O.L.S. is also normally distributed, which is a convenient property for hypothesis testing and interval estimation. When u_t is not assumed normal there is a need to develop asymptotic theory in order to allow approximate inferences. It is known under suitable conditions that O.L.S. is asymptotically efficient relative to all linear unbiased estimates of β, when the u_t are, say, i.i.d. and homoscedastic. Eicker (1963,1967) established asymptotic normality of O.L.S. under such circumstances, and proposed methods of consistently estimating the covariance matrix of the limiting distribution, and thus to be able to obtain approximately valid statistical inferences. This work has been developed in the more recent econometric literature by Domowitz and White (1982), White and Domowitz (1984), Newey and West (1987). Moreover, Grenander (1954) showed that O.L.S. is even asymptotically efficient in the presence of serially correlated u_t when the x_t are of certain type, polynomials and certain trigonometric functions of time.

In general, however, O.L.S. will not be asymptotically efficient when the u_t are serially correlated or heteroscedastic. There is evidence that O.L.S. can be very inefficient, see for instance Watson (1955) or Watson and Hannan (1956) with respect to G.L.S.. Moreover, though O.L.S. can be used in statistical inference with a covariance matrix estimate, the resulting inferences will not be asymptotically efficient. In general one needs to correct for the serial correlation and/or heteroscedasticity in order to obtain efficient point estimates and inferences.

1.3.1.- Weighted Least Squares

The estimation of β in the case of heteroscedasticity of known form is known as weighted least squares. It can be based on a two-step algorithm which can be iterated, although the asymptotic properties of the estimate of β are not altered. In the first step, the heteroscedasticity function $\sigma(x_t)$ is estimated by least squares regression of a certain transformation of the residuals obtained by $y_t-\bar{\beta}'x_t$, where $\bar{\beta}$ is an estimate of β, for example the O.L.S. residuals, on a transformation of the function $\sigma(x_t)$, see Judge et al. (1985). In the second step, a weighted least squares estimate of β is obtained, where the weights are the inverse of the estimated heteroscedasticity function. It is worth pointing out that although the least squares estimates of the parameters that the heteroscedasticity depends on are not the most efficient one, as far as the β estimates is not relevant, see for instance Theil (1971), Judge et al. (1985), Amemiya (1973a) or Harvey (1976) among others.

1.3.2.- Parametric Estimation with Serial Correlation

In the estimation of β when the residuals are autocorrelated, we will distinguish two cases. The first one is the pure autoregressive (AR) process and the second one is the autoregressive moving average (ARMA) process, since the inclusion of the MA factor induces certain difficulties or characteristics in the estimation procedure, which do not appear in the pure AR case.

In both cases if the variance-covariance matrix Γ of the residuals were known then we can apply G.L.S., i.e. equation (2.3) where σ is the identity matrix, obtaining the BLU estimate. Generally, the covariance matrix is unknown and therefore it has to be estimated.

1.3.2.1.- Estimation under AR(P) Errors

Three specifications are the most common. The first one is known as the feasible G.L.S. (F.G.L.S.) estimate. Using the transformation matrix Q, see Fuller (1976, p.423) for a general expression of Q, of the covariance matrix, such that $Q\Gamma Q'=I$, we can use a two step algorithm. In the first step, an estimate of the AR parameters is obtained and used in place of the true ones in Q, giving $\dot{Q}$. In the second step, the β estimate is obtained by regressing $\dot{Q}Y$ on $\dot{Q}X$. If in this step the first P rows are omitted, we obtain the Cochrane-Orcutt (1949) algorithm. Several procedures have been suggested for the estimates of the AR parameters. Among them are the least squares estimate (L.S.E.) of the residuals upon its P-period past values or Durbin's (1960) procedure.

The second specification is non-linear least squares (NLS). This procedure estimates both β and the AR parameters that minimize $(Y-X\beta)'Q'Q(Y-X\beta)$. In this case the estimator of β will still be of the same form as in the previous precedure, but the estimator of the AR parameters need no correspond to the Durbin or L.S.E. ones.

The third specification is M.L. assuming normality. In a series of papers Beach and MacKinnon (1978a,b) give an algorithm to estimate the β's and autoregressive parameter(s). Although the estimate for β is of the same type as the F.G.L.S., the estimate of the autoregressive parameters is different than for the above two specifications. For the linear AR(1) model, the M.L. estimate of the autoregressive parameter is obtained by solving a third order equation. They claim that this algorithm is not more expensive computationally than methods based on Cochrane-Orcutt and proved that it is more efficient in small samples. For the linear AR(2), Beach and MacKinnon (1978b) proposed a M.L. estimate which incorporates the stationary restriction, as in the AR(1) model, and the technique used is the fixed point or simplicial search algorithm. They also indicate that the algorithm can be extended to higher order autoregressive models. In the AR(2) model Fair (1973) discarded the first two observations in the specification of the likelihood function, and then it follows a direct generalization of the Cochrane-Orcutt procedure. It is worth noting that all the above methods give the same asymptotic efficiency for the estimate of β.

1.3.2.2.- Estimation under ARMA Errors

Due to the impossibility to obtain the Q transform, except for the MA(1) and ARMA(1,1) cases, see Jugde et al. and Tiao and Ali (1971) respectively, to obtain a two-step F.G.L.S estimate is not possible and therefore we will concentrate on M.L. and non-linear least squares estimates. However, Amemiya (1973b) obtains an asymptotically efficient estimate of the β parameters where the autocovariances are estimated from the L.S. regression of the calculated residuals upon its N-period past values, where N depends on the number of observations T. Also Harvey and Phillips (1979) give an algorithm to get the G.L.S. estimate of β by using the Kalman Filter algorithm and also, they give indications to obtain the exact M.L. estimate of β by the Kalman filter.

Pierce (1971) shows that there is no difference between M.L. and the various versions of NLS, which differ in the treatment of the presample observations, where two methods are the usual ones. The first one treats the presample observations as equal to zero, and the second one uses the backforcasting procedure to treat the presample observations, see Box and Jenkins (1976) for this method.

1.4.- SEMIPARAMETRIC G.L.S. ESTIMATE OF β

If in estimating β parametrically, we use a misspecified model for the autocovariance and conditional heteroscedasticity of u_t, then the estimate will still often be consistent and asymptotically normal. However, it will generally not be efficient, indeed there is evidence (Engle, 1974) that we can do even worse than using O.L.S. These circumstances motivate an estimate of β where the autocovariances and/or heteroscedasticity of u_t are estimated via nonparametric methods.

A variety of such semiparametric estimates of β have been proposed. Our discussion of these, and of the alternative semiparametric estimates which are in the next section is in part adapted from the survey article of Robinson (1988b)

Assuming that u_t is stationary and independent of all the x_t, Hannan (1963) proposed an estimate of β that has the same asymptotic efficiency as the best linear unbiased estimate, without employing a parametric model for the autocovariances of u_t. Using the fact that the Fourier transformation of the Toeplitz matrix of autocovariances approximately produces a diagonal matrix with diagonal elements equal to the u_t's spectral density, Hannan estimated a Fourier transformed version of (1.1) using weighted least squares, where the weights are the reciprocals of nonparametric spectral estimates. In a different fashion from Hannan's original version is to use of periodograms of I_{yx} and I_x instead of spectral weighting as in Hannan (1971), Robinson (1972), Hannan and Robinson (1973) or Hannan and Terrell (1973) among others.

More recently, a similar solution to the conceptually simple problem is when the u_t's are independent but conditionally heteroscedastic with heteroscedasticity function of unknown form. Assuming the model $u_t=\sigma(x_t)v_t$, where σ is of unknown form and the v_t are i.i.d. with zero mean and independent of all the x_t (i.e. a special case of (2.1), (2.2)), Carroll (1982) estimated the $\sigma^2(x_t)$ by nonparametric regression and established asymptotic efficiency of a weighted G.L.S. estimate of β. Robinson (1987a) considerably relaxed Carroll's conditions in several respects. One of these has bearing on efficiency. Robinson assumed that the relation between x_t and u_t was generated only by the conditional moment restriction

$$(4.1)\quad E[u_t|x_t]=0,\ V[u_t|x_t]=\sigma^2(x_t)$$

rather than independence of x_t and u_t. Under (4.1) Robinson's G.L.S. estimate is the most efficient possible one in the presence of an unknown σ^2, achieving the semiparametric efficiency bound defined and obtained by Begun et al. (1983), Chamberlain (1987) and others. Under Carroll's conditions, however, a better estimate than G.L.S. exists, because G.L.S. does not emploit his independence assumption. This has implications also for other estimates. Some extensions of semiparametric G.L.S. estimates to ARCH regression models, when the ARCH structure is nonparametric has been considered by Robinson (1987c), and applied empirically by Whistler (1988). Delgado (1989) has extended Robinson's work to non-linear regressions. Our treatment of (1.1), (1.3) is in a similar spirit to that of the above authors, except that our unknown function is the AR function ρ, rather than the heteroscedasticity function σ^2.

Some semiparametric work that allows for both serial correlation and heteroscedasticity simultaneously has been carried out. Harvey and Robinson (1988) assumed the model

$$u_t=\sigma_t v_t, \quad v_t-\sum_{j=1}^{p}\alpha_j v_{t-j}=\epsilon_t$$

where p is given and σ_t is an unknown function of t. Thus this model for the u_t's autocovariances is itself semiparametric rather than nonparametric, since it imposes a parametric time series model on v_t, unlike our model (2.2). Moreover, this specification that assumes that the variances σ^2 depend on t and not on x_t is a more restrictive specification than our (2.1). By estimating β under the semiparametric model (2.1), (2.2), we extend the work of Hannan (1963), Carroll (1982) and Robinson (1987a).

1.5.- MORE EFFICIENT SEMIPARAMETRIC ESTIMATES OF β

To place our work in full perspective it is worth discussing some literature that suggests that even more efficient estimates of β can be obtained than ours.

If the joint distribution of the u_t in (1.1) is of known parametric form, then at least in principle M.L. estimates of β can be obtained, and shown to be asymptotically efficient among the class of all regular estimates, achieving the asymptotic Cramer-Rao bound. Some versions of Huber's (1980) M-estimates have this property.

When the parametric form of the u_t's distribution is misspecified, then estimates of β will not only be inefficient but they may even be inconsistent. Following earlier work of Stone (1975) for a simple location model, Bickel (1982) considered a semiparametric estimation of β in (1.1), assuming that the disturbances u_t are i.i.d. with unknown distribution function. By employing nonparametric estimation of the probability density function of u_t, Bickel obtained estimates of β that he showed to be asymptotically as efficient under suitable conditions as ones based on a correct parametric distribution function for the u_t disturbances.

This work has been extended by a variety of authors, such as Manski (1984) and Kreiss (1987), who considered semiparametric estimation of non-linear regression models and semiparametric estimation of the coefficients of a parametric linear time series model in the presence of innovations whose distribution is of unknown form respectively. In this thesis we have not attempted to adapt for the full distribution of the u_t, but only to certain features relevant to G.L.S. estimation, so our efficiency goal is less ambitious. The possibility of extending our work by adapting more fully to the u_t distribution remains a challenging open question.

1.6.- PLAN OF THE THESIS

Our estimates require introducing a good deal of machinary on nonparametric estimation, consisting of nonparametric regression, probability density estimation and spectral density estimation. This is described in Chapter 2. In Chapter 2 it is also convenient to introduce the type of time series that will be used in deriving the limiting distributions of our estimates. Chapters 3 and 5 describe the methodology of semiparametric estimation of β under models (1.3) and (2.1), (2.2) respectively. Both estimates are shown to be asymptotically normal and consistent estimates of the limiting covariance matrix are presented. The proofs of these asymptotic results are in appendices. Chapters 4 and 6 describe Monte-Carlo evidence of the finite sample behaviour of our estimates.

CHAPTER 2

SOME NONPARAMETRIC METHODOLOGIES AND THE CONCEPT OF MIXING CONDITION

In this chapter we will discuss some general background about the methodologies that we will use throughout chapters 3 and 5, as well as the mixing conditions of weak dependence in time series which will play a fundamental role in this thesis. The methodologies to be used are kernel regression and spectral estimation. It is convenient also to specify alternative methodologies to the kernel regression estimators and describe why we will not proceed with such alternatives. The plan for this chapter is as follows: in section 1 we will introduce the kernel regression estimator and one of its competitors, in section 2 we will introduce the kernel spectral methodology in time series regression models; finally, in section 3 we will discuss the mixing conditions of weak dependence of a stationary stochastic process.

2.1.- THE KERNEL AND NEAREST NEIGHBOUR REGRESSION ESTIMATORS

Both these methodologies are known in the statistical literature as nonparametric estimators. But, what do we understand by nonparametric regression estimation? We will answer this question by means of an example. A practitioner has observations $(y_i,x_i)_{i=1,\ldots,T}$ from a population (Y,X). He may know certain conditional moments up to a set of parameters β, i.e. $E[Y|X]=p(X,\beta)$, and thus, based on the available data he will estimate the parameters β.

Another possibility arises when the practitioner is not able to parameterize the function p. In this situation, he may still use statistics, based on the available sample, to estimate the unknown function $p(\cdot)$ of a certain point x_0. These statistics are known as nonparametric estimators, their targets being not parameters but infinite dimensional-functions.

2.1.1.- Kernel Regression Estimation

Before introducing the kernel regression estimator, one has to point out that although this thesis is dealing with time series data, it is better to introduce it in the independent and identically distributed (i.i.d.) context, since it may then be easier to get the basic idea behind the estimator.

Because the kernel regression estimator is based on the kernel probability density estimator, we will introduce the latter first. Let $x_1,\dots,x_T$ be a random sample from a population with probability density function $f(x)$. Let us suppose that we are interested in the estimation of $f(x)$ at the point x_0. We know that a natural estimator of the distribution function $F(x)$ at $x=x_0$ is its sample analogue $F_T(x_0)$ defined by:

$$F_T(x_0)=\frac{1}{T}\text{ (n}^0\text{ of observations in the sample} \leq x_0).$$

If we pay a little attention to the mathematical definition of the probability density function in the point x_0, i.e. $f(x_0)$, this value can be approximated by the equation:

$$f(x_0)\approx \frac{F(x_0+a)-F(x_0-a)}{2a}$$

for an a sufficiently small and hence, a natural estimator of $f(x_0)$ turns out to be the above equation by substituting $F(x)$ for $F_T(x)$, i.e.:

$$f_T(x_0)= \frac{F_T(x_0+a)-F_T(x_0-a)}{2a} .$$

Let us now consider a function K defined by:

$$K(y)=\begin{cases} \frac{1}{2} & \text{if } |y|<1 \\ 0 & \text{if } |y|>1 \end{cases}$$

which turns out to be the uniform function in the interval $[-1,1]$. It is obvious that the estimator $f_T(x_0)$ can be rewritten as:

$$(1.1)\quad f_T(x_0)=\frac{1}{Ta}\sum_{j=1}^{T} K\left[\frac{x_0-x_j}{a}\right].$$

Consequently, if instead of having chosen the uniform distribution for the K function we would have chosen any other function, with the requirement that it integrates 1, we would get what it is known as the kernel estimator of the probability density function in the point x_o. We may think of the function K as a weighted function of the x_j data. In other words, consider a ball with centre x_o and radius a; then the importance that we give to the x_j observation in the estimation of $f(x_o)$ is measured by the quantity K_j, where K_j is equal to:

$$\frac{1}{a} K\left[\frac{x_0 - x_j}{a}\right].$$

Thus, when we use the uniform distribution function as the K function we give the same importance to all observations made inside the ball of radius a around the point x_o.

The initiative to estimate a probability density function in this way was first shown by Parzen (1962). It is to be noted that we do not assume the function K to be non-negative for reasons that will become clear in the next chapter, but, on the other hand, we need that this function to be highly concentrated around the origin, since more relevance is given to those observations which are closer to x_o.

This nonparametric method involves the feature of a "bandwidth" or "smoothing" parameter a, which is positive, selected by the practitioner and regarded as converging to 0 (not too slow or too fast) as T (the number of observations) approaches infinity. In fact, any choice of a in finite samples implies a trade-off between bias and variance, and this parameter a can be seen as a measure of the degree of smoothing of the kernel estimator in such a way that the larger a becomes the larger the degree of smoothing.

In the p-dimensional case, the natural generalization of the estimator given by the equation (1.1) is

$$(1.2)\quad \hat{f}(x_o)=\frac{1}{T|a|}\sum_{j=1}^{T} K[(x_o-x_j)'a^{-1}]$$

where $|a|$ stands for the determinant of the p-dimensional a matrix, that goes to 0 as T approaches to infinity. A special case arises when the a matrix is diagonal and K(u) is the product of p univariate functions; then the K function is the product of p-univariates K^i's, such that

$$K(u'a^{-1})=\prod_{i=1}^{p} K^i\left[\frac{u_i}{a_i}\right]$$

obtaining the estimator, for the case $a_i=a$ $\forall i=1,\dots,p$,

$$(1.3)\quad \frac{1}{Ta^p}\sum_{j=1}^{T}\left[\prod_{i=1}^{p} K^i\left[\frac{x_o^i-x_j^i}{a}\right]\right].$$

Definition 2.1.1

Given a sample $x_1,x_2,\dots,x_T$ from a p-dimensional population X, we will define the kernel density estimator of f(x) by (1.3).

It is worthwhile noting that there does not exist any unbiased estimator of the probability density function f(x) (see Rosenblatt, 1961).

Following this discussion a natural question which might arise is: given a random sample $(y_1, x_1), \ldots, (y_T, x_T)$ from a, say, bivariate population with probability density function $f(y,x)$, and assuming the existence of the conditional expectation of y on x, which is:

$$y(x) = \int_R y \, f_{Y|X}(y|x)\,dy$$

$$= \int_R y \, {f(x,y)}/{g(x)}\,dy \quad ; \qquad \text{where } g(x) = \int_R f(x,y)\,dy,$$

R is the real line and $f_{Y|X}(y|x)$ is the conditional distribution function of Y on X, is there any way to estimate the regression function $y(x)$?

The answer to this question was first shown by Watson (1964) and Nadaraya (1964) who independently proposed an estimator based on kernel functions and whose motivation is explained in the next paragraph.

We have seen that the probability density function was estimated by the equation (1.2). This leads us to think that a sensible candidate for an estimator of $y(x)$ is given by one in which we substitute the estimators of the probability density functions $f(y,x)$ and $g(x)$ in the regression equation $y(x)$, as given by definition 2.1.1. Therefore the estimator of $y(x)$ will be given by:

$$\hat{y}(x) = \int_R y\hat{f}(x,y)/\hat{g}(x)\,dy$$

$$-a \sum_{j=1}^{T} m\left[\frac{x-x_j}{a}\right] \Big/ \sum_{j=1}^{T} J\left[\frac{x-x_j}{a}\right]$$

$$+\sum_{j=1}^{T} y_j \, J\left[\frac{x-x_j}{a}\right] \Big/ \sum_{j=1}^{T} J\left[\frac{x-x_j}{a}\right]$$

after we integrate the R.H.S. of the first equation by parts, and where:

$$m(x)=\int_R yK(x,y)\,dy$$

and

$$J(x)=\int_R K(x,y)\,dy.$$

If, in addition, it is assumed that $K(x,y)=J(x)L(y)$ and $L(y)$ is a symmetric function, then the estimator of the regression function $y(x)$ will become:

$$(1.4) \quad \sum_{j=1}^{T} y_j \, J\left[\frac{x-x_j}{a}\right] \Big/ \sum_{j=1}^{T} J\left[\frac{x-x_j}{a}\right]$$

since $m(x)$ will be equal to zero.

The term (1.4) is the kernel regression estimator which was proposed by Watson (1964) and Nadaraya (1964) and which may be generally considered as:

Definition 2.1.2

Given a random sample $(y_1,x_1),\ldots,(y_T,x_T)$ from a population (Y,X) of dimension (1+p), we define the kernel regression estimator of the conditional expectation of h(y) given $x=x_0$, i.e. $E[h(y)|x=x_0]$, by:

$$(1.5)\quad \sum_{j=1}^{T} h(y_j)K\left[\frac{x_0-x_j}{a}\right] \Big/ \sum_{j=1}^{T} K\left[\frac{x_0-x_j}{a}\right]$$

where the K function is defined as in (1.3).

Frequently, the investigator is concerned with the estimation of functionals which involve derivatives of unknown functions, such as the score function or the information matrix of a density function, or the derivative of a regression function, whenever there exist. For example, in the model which we will present in the next chapter, the estimation procedure of the parametric part of the model involves the derivative of a regression function. If the functional form of the regression function is unknown, so is the derivative, and therefore this derivative has to be estimated via nonparametric methods.

It seems that a sensible estimator of the derivative of an unknown function in view of the definition 2.1.2 is given by:

Definition 2.1.3

Given a random sample $(y_1,x_1),\ldots,(y_T,x_T)$ from a population (Y,X) of dimension (1+p), the derivative kernel regression estimator of $\partial/\partial x_0 E[h(y)|x=x_0]$ is:

$$(1.6)\quad \frac{1}{a}\sum_{j=1}^{T} h(y_j)K'\left[\frac{x_0-x_j}{a}\right] \Big/ \sum_{j=1}^{T} K\left[\frac{x_0-x_j}{a}\right] - \frac{1}{a}\sum_{j=1}^{T} K'\left[\frac{x_0-x_j}{a}\right] \Big/ \left(\sum_{j=1}^{T} K\left[\frac{x_0-x_j}{a}\right]\right)^2$$

where K' is the derivative of the function K.

From this definition it is obvious that the kernel function K has to be differentiable.

Before finishing this subsection it is useful to realize that the implementation of the estimators involves two different aspects. Firstly, the choice of the kernel function K, and secondly, the bandwidth parameter a. Several K functions have been proposed in the literature. Some popular ones are:

a) K_e $\quad 3/4\ (1-1/5t^2)/\sqrt{5} \quad |t|<\sqrt{5}.$

b) Gaussian $\quad 1/\sqrt{2\pi}\ \exp(-\frac{1}{2}t^2) \quad t\in R.$

c) Rectangular $\begin{cases} \frac{1}{2} & |t|<1 \\ 0 & \text{otherwise.} \end{cases}$

Epanechnikov (1969) showed that K_e is the kernel function which minimizes (asymptotically) the function

$$\int_{-\infty}^{\infty} E\{\hat{f}(x)-f(x)\}^2 dx$$

that is, the mean integrated square error (M.I.S.E.), where this value can be seen as a global measure of discrepancy between f(x) and its estimator $\hat{f}(x)$. It can be seen that the relative efficiency of the other kernels to the Epanechnikov (K_e) is very close to 1.

Bearing in mind that this relative efficiency is an asymptotic result, it implies that the choice of the kernel function K will not have much importance on the basis of M.I.S.E.. Thus, it is perfectly legitimate, and also desirable, to choose the kernel function K based on other considerations such as differentiability or computational properties.

As far as the choice of the the bandwidth **a** is concerned, it is more crucial than the choice of the K function. While it may be possible to choose the bandwidth automatically based on some functional optimization, we are not going to proceed in this direction in choosing our **a**. In the simulations that we will report in chapters 4 and 6, we have decided to adopt the practice of doing the calculation over a grid of bandwidths by trial and error.

2.1.2.- Nearest Neighbour Estimation

The purpose of this section is to specify one of the most relevant alternative methodologies to the kernel regression one that has appeared in the nonparametric literature and to discuss its disadvantages with respect to the kernel methodology.

2.1.2.1.- Nearest Neighbour

The Nearest Neighbour(NN) methodology is, perhaps, the most important and exhaustively studied among the alternative methodologies to the kernel regression used in nonparametric estimation.

Let (X,Y) denote a random vector belonging to $R^p \times R$, and let $(x_i,y_i)_{i=1,\ldots,T}$ be a random sample from such a population. An alternative estimator to the kernel regression estimator given by definition 2.1.2 is:

$$(1.7) \qquad y_T(x)=\sum_{i=1}^{T} h(y_i)W_{T,i}(x)$$

if $E|h(Y)|<\infty$ and where $W_{T,i}(x)$ are weights following a probability function. The $W_{T,i}(x)$ weights those values of i for which the corresponding x_i is close to x (in a metric defined below) more heavily than those values of i for which x_i is further from x.

Because the explanatory variables x, in an econometric model, may not be measured in the same units, we divide each variable in the x vector by their sample standard deviation prior to applying the Euclidean metric, that is let s_m be:

$$s_m= (T-1)^{-1}\sum_{i=1}^{T}[x_{im}-(\sum_{i=1}^{T}x_{im}/T)]^2 \qquad \text{for all } 1\leq m\leq p$$

and the metric is defined by the equation:

$$\rho_i=\sum_{m=1}^{p}\frac{1}{s_m}[x_{im}-x_m]^2 \qquad i=1,\ldots,T.$$

The question which arises is how to apply the NN estimator of the conditional expectation of $h(y)$ on x?. Given x, the data (x_i,y_i) $1\leq i\leq T$ is rearranged according to the distances ρ_i. Let us define:

$$p_i = 1 + \sum_j I(\rho_j < \rho_i)$$

and

$$q_i = 1 + \sum_j I(\rho_j = \rho_i)$$

where the sums are from $j=1$ up to $j=T$, $j \neq i$, and $I(.)$ is the usual indicator function. Let $W^*_{T,i}(x)$ be defined by:

$$W^*_{T,i}(x) = q_i^{-1} \sum_{j=p_i}^{p_i+q_i-1} W_{T,j}(x)$$

then the NN estimator of $y_T(x)$ is defined as:

<u>Definition 2.1.4</u>

Given a random sample $(x_i, y_i)_{1,\ldots,T}$ from a population (X,Y), we define the NN estimator of $y(x)$ by:

$$\hat{y}_T(x) = \sum_{i=1}^{T} h(y_i) W^*_{T,i}(x).$$

When there are no ties, then $W^*_{T,i}(x)$ is equal to $W_{T,i}(x)$. One important subclass of NN estimators is the k-Nearest Neighbour (k-NN) which is such that $W_{T,i}(x)$ is equal to zero if $i>k$. Some examples of k-NN weights are:

the uniform: $W_{T,i}(x) = 1/k \qquad 1 \le i \le k,$

the triangular: $W_{T,i}(x) \propto k-i+1 \qquad 1 \le i \le k$

and

the quadratic: $W_{T,i}(x) \propto (k^2-(i-1)^2) \quad 1 \leq i \leq k.$

The objective of the triangular and quadratic k-NN estimators is to achieve greater smoothness in the k-NN estimator.

In fact, the NN estimator adapts the amount of smoothing to the "local" density of data. The degree of smoothing is conditional to the integer k, chosen to be considerably smaller than the sample size, although it goes to infinity as T approaches to infinity.

Once we have introduced the NN estimators, it is reasonable to discuss the reasons why we have not proceeded with this methodology, discussing its disadvantages with respect to the kernel regression estimator.

Most of the theory about the NN-estimators has been developed in an i.i.d. environment. There are some results on k-NN estimators under dependence, i.e. by Collomb (1985) and Yakowitz (1987), but nothing so far as we know which parallels the results of Stone (1977) for the independence case, which plays a fundamental role in all semiparametric estimation with the NN methodology. By contrast, all the theory available for kernel estimators in the i.i.d. case can be translated, under appropriate conditions, to the time series environment. Thus, because we are dealing with time series models, it is justifiable to use kernel regression estimation instead of NN.

Another drawback of the NN-estimator is that it is not differentiable. This is a serious problem since many semiparametric models involve the estimation of the derivatives of unknown conditional moments, as it is the case in the model of chapter 3.

2.2. KERNEL SPECTRAL ESTIMATOR

Before introducing the estimator of the coefficients of a time series regression model in the frequency domain, it will be convenient to give the definition of the kernel spectral estimator of the spectral density function of a stochastic process, and some of its asymptotic properties which will be used to prove the theorem of chapter 5.

2.2.1.- The Definition of the Spectral Density Function and its Estimator

In this subsection, we are going to give the definition of the spectral density function, assuming that it exists, of a real stochastic process x_t with autocovariance matrix function $\{\gamma(t)\}_{t=-\infty,\dots,\infty}$, where the ij-th element is $\gamma^{i,j}(t)=E(x_{is}x_{jt+s})$.

Definition 2.2.1

Let $\{x_t\}_{t=-\infty,\dots,\infty}$ be a (kxl) vector of a real stochastic process with autocovariance matrix function $\{\gamma(t)\}_{t=-\infty,\dots,\infty}$. We will define the spectral density function of the process x_t as the (kxk) matrix $f_{\overline{xx}}(\lambda)$, which ij-th element is defined as:

$$(2.1) \qquad f^{i,j}_{\overline{xx}}(\lambda)=\frac{1}{2\pi}\sum_{t=-\infty}^{\infty}\gamma^{i,j}(t)e^{it\lambda} \qquad \lambda\epsilon[-\pi,\pi].$$

The spectral density function and the autocovariance matrix function are two equivalent concepts, in the sense that, knowing one of them you can identify the other completely.

But, $\gamma(t)$ is, of unknown parametric form, and so is $f_{\overline{xx}}(\lambda)$. An obvious estimator of the spectral matrix density function, based on T observations, is the (kxk) matrix $f^*_{\overline{xx}}(\lambda)$, where the ij-th element is:

$$f^{*i,j}_{\overline{xx}}(\lambda)=\frac{1}{2\pi}\sum_{-T+1}^{T-1}\hat{\gamma}^{i,j}(t)e^{it\lambda}$$

where $\hat{\gamma}(t)$, i.e. the sample autocovariance, is an estimator of $\gamma(t)$, which ij-th element is defined as:

$$\hat{\gamma}^{i,j}(t)=\frac{1}{T}\sum_{s=1}^{T-|t|}x_{is}x_{js+|t|}.$$

It can be shown that $f^*_{\overline{xx}}$ is not a consistent estimate of the spectral density function, and therefore it is not useful. But a slight modification is consistent.

Let us introduce a function $K_1:[-1,1]\rightarrow R$, where K_1 is at least even and $K_1(0)=1$. Let also M be a nonnegative integer; the larger T is, the larger M should be. In spectral analysis the function K_1 is called the <u>lag-window</u> and M the <u>lag-number</u>.

Definition 2.2.2

Let $\{x_t\}_{t=-\infty,\ldots,\infty}$ be a (kx1) vector of a real stochastic process. We define the kernel spectral density estimator of such a process, based on T observations, as:

$$(2.2)\qquad \tilde{f}_{\overline{xx}}(\lambda)=1/2\pi \sum_{-M+1}^{M} K_1(t/M)\hat{\gamma}(t)e^{it\lambda} .$$

The lag-number marks the cutoff-point of the sample autocovariance to be used in equation (2.2).

Several K_1 functions have appeared in the literature. Some of the most popular ones are:

$$(2.3)\qquad \text{Truncated } K_1(s)=\begin{cases} 1 & |s|<M \\ 0 & |s|>M, \end{cases}$$

$$(2.4)\qquad \text{Barlett } K_1(s)=\begin{cases} 1-|s|/M & |s|<M \\ 0 & |s|>M, \end{cases}$$

$$(2.5)\qquad \text{Tukey-Hanning } K_1(s)=\begin{cases} 1-2a+2a\cos(\pi s/M) & |s|<M \\ 0 & |s|>M \end{cases}$$

and

$$(2.6)\qquad \text{Parzen } K_1(s)=\begin{cases} 1-6(|s|/M)^2+6(|s|/M)^3 & |s|<M/2 \\ 2(1-|s|/M) & M/2<|s|<M \\ 0 & |s|>M. \end{cases}$$

Like in the kernel density function of section 2.1.1, the selection of the lag-window is not relevant, what it is more important is the lag-number M that we choose.

In order to give some of the asymptotic properties of the kernel spectral density estimator, we have to introduce the concept of the characteristic exponent, which plays a fundamental role.

<u>Definition 2.2.3</u>

We define the characteristic exponent of K_1 as the greatest integer r for which

$$K^{(r)}=\lim_{u\to 0}\left[\frac{1-K_1(u)}{|u|^r}\right]$$

is finite.

Assuming that:

1) $\sum_{t=-\infty}^{\infty} |t|^q|\gamma(t)|<\infty$, $q<r$, where r is the characteristic exponent

and

2) K_1 is a bounded function

then (see Parzen, 1957 or Hannan, 1970 pp. 280-283) the mean square error of $\tilde{f}_{XX}(\lambda)$ is $O(T^{-1}M+M^{-2q})$.

2.2.2.- The Generalized Least Squares Estimator in The Frequency Domain

Let the following linear regression model be

$$
\begin{aligned}
&y_t = \beta' x_t + u_t \\
(2.7) \qquad & \qquad\qquad \epsilon_t \text{ i.i.d. with zero mean and variance 1.} \\
&u_t = \sum_{j=-\infty}^{\infty} \alpha_j \epsilon_{t-j} \qquad \sum_{j=-\infty}^{\infty} |\alpha_j| < \infty
\end{aligned}
$$

Here, we are going to describe two methods of estimating the coefficients β of the model (2.7) in the frequancy domain.

METHOD 1.

Hannan (1963) showed that in the absence of knowledge of the coefficients α_j, we can estimate the parameter β of equation (2.7) efficiently. His estimator is given by:

$$
(2.8) \qquad \tilde{\beta} = \left[\frac{1}{2M} \sum_{-M+1}^{M} \tilde{f}_{\hat{u}\hat{u}}^{-1}\left[\frac{\pi k}{M}\right] \tilde{f}_{\overline{xx}}\left[\frac{-\pi k}{M}\right]\right]^{-1} \left[\frac{1}{2M} \sum_{-M+1}^{M} \tilde{f}_{\hat{u}\hat{u}}^{-1}\left[\frac{\pi k}{M}\right] \tilde{f}_{\overline{yx}}\left[\frac{\pi k}{M}\right]\right]
$$

where $\tilde{f}_{\hat{u}\hat{u}}(\lambda)$ is as in definition 2.2.2., but the autocovariances are estimated with the estimated O.L.S. residuals instead of the true ones.

The intuition behind this estimator is given by the fact that, if outside a frequency interval (suppose x_t is scalar), x_t does not have spectral mass and the spectral density function of the residuals varies very little over such an interval, then the efficiency of the least squares of β will be nearly unity.

Then, if we are able to have a set of mutually exclusive filters, say r filters, whose union is $[0,\pi]$ and such that the variation of the spectral density of u_t is small over each interval, then the efficiency of the estimation of β over each interval is nearly unity. Therefore, if we combine each of this estimates of β optimally, we will obtain an estimate of β whose efficiency will not be less than for each of those estimates.

Although such filters are not available, we can still use an equivalent approximation. Since $f_{\overline{yx}}(\lambda)=\beta f_{\overline{xx}}(\lambda)$, then

$$\hat{\beta}(k)=\hat{f}^{-1}_{\overline{xx}}(\pi k/M)\ \hat{f}_{\overline{yx}}(\pi k/M)$$

will be a plausible estimate of β. This estimate of β has an asymptotic variance proportional to $f^{-1}{}_{\overline{xx}}(\pi k/M) f_{\overline{uu}}(\pi k/M)$. Hence, equation (2.8) will give the optimal weighting of the several $\hat{\beta}(k)$ if the correlations between the several $\hat{\beta}(k)$ are not taken into account.

METHOD 2.

The suggestion for this method comes from the fact that $\omega_u(\lambda_s)=\omega_y(\lambda_s)-\beta'\omega_x(\lambda_s)$ where ω_z is the discrete Fourier transform of $\{z_t\}_{t=1,\ldots,T}$ defined as

$$\omega_z(\lambda)=\frac{1}{\sqrt{2\pi T}}\sum_{j=1}^{T} z_j\ e^{i\lambda j}$$

where $\lambda_s = {}^{2\pi s}/_T$, has under regularity conditions an asymptotic distribution function $N(0, \overline{f_{uu}}(\lambda_s))$, and the $\omega_u(\lambda_s)$ are approximately independent for different λ_s.

Then, the G.L.S. principal will suggest as an estimate of β

$$(2.9) \qquad \bar{\beta} = \Big(\sum_{s=1}^{T} I_x(\lambda_s) g_{\hat{u}\hat{u}}(\lambda_s)^{-1} \Big)^{-1} \Big(\sum_{s=1}^{T} I_{xy}(\lambda_s) g_{\hat{u}\hat{u}}(\lambda_s)^{-1} \Big)$$

where $I_{zv}(\lambda) = \omega_z(\lambda)\omega_v'(-\lambda)$ and $g_{\hat{u}\hat{u}}(\lambda)$ is an estimator of the spectral density function defined as

$$(2.10) \qquad g_{\hat{u}\hat{u}}(\lambda) = (2\pi T)^{-1} \sum_{s=1}^{T} K_M(\lambda - \lambda_s) I_{\hat{u}\hat{u}}(\lambda_s)$$

where $K_M(\lambda) = M\sum_{-\infty}^{\infty} K_2(M(\lambda + 2\pi s))$ and K_2 is a real even function which integrates 1 and is absolutely integrable. This function is known as the spectral window.

Here M is not restricted to be an integer, as in method 1. One difference between the two methods is that method 2 uses (2.10), i.e. we weight the periodogram instead of weighting the autocovariances as in (2.2), apart from the formula for the β estimator. Another difference is that in this method, we use the periodograms in (2.9) instead of the smoothed spectral estimates as in (2.8)

2.3.- THE CONCEPT AND EXAMPLES OF MIXING PROCESSES

The purpose of this section is twofold. First, to give a description of some of the mixing conditions that have appeared in the statistical literature, some of which will be used frequently in chapters 3 and 5. Secondly, and perhaps more important, to give several sufficient conditions under which a given stationary stochastic process is mixing, as well as some examples; and therefore to be able to check that the models used in the simulations of chapters 3 and 5 satisfy the required conditions imposed in their respective theorems.

2.3.1.- Some Definitions of Mixing Processes.

The first question that comes to the mind is, what do we understand by a mixing process?

Definition 2.3.1

We say that a strictly stationary vector stochastic process $\{x_t\}$ is mixing if $\forall A,B\epsilon\mathcal{F}_{-\infty}^{\infty}$ (i.e. the σ-algebra generated by $\{x_t\}_{t=-\infty,\dots,\infty}$)

(3.1) $\lim\limits_{m\to\infty} P(A\cap T^{-m}B)=P(A)P(B)$

where T stands for the usual shift operator on events in $\mathcal{F}_{-\infty}^{\infty}$.

That is, loosely, although events determined by sets of x_t may be considerable dependent on recent events, they are almost independent provided they are far apart in time. Thus, mixing is a form of asymptotic independence.

From this definition, it is readily seen that the mixing property is a sufficient condition for ergodicity by taking $A \equiv B$-invariant where we obtain that $P(A)^2 = P(A)$ so that $P(A) = 1$ or 0.

Several measures of dependence have been proposed. The weakest one which we will consider is <u>strong-mixing</u> (Rosenblatt, 1956). Consider all events depending only upon x_t for $t \leq p$. These events constitute a σ-algebra $\mathcal{B}_p$. Similarly $\mathcal{A}_{p+r}$ is the σ-algebra of events determined by x_t $t \geq p+r$. Define $\alpha(\cdot)$ as a function such that:

$$(3.2) \qquad \sup_{B \in \mathcal{B}_p \; A \in \mathcal{A}_{p+r}} |P(B \cap A) - P(A)P(B)| < \alpha(r), \text{ where } \alpha(r) \geq 0 \text{ and } r > 0.$$

<u>Definition 2.3.2</u>

We say that the stationary stochastic process $\{x_t\}$ is strong-mixing if $\alpha(r)$ goes to zero as r goes to infinity.

It is obvious that strong-mixing implies mixing (i.e. (3.1)) and then the ergodicity of the process.

The next measure of dependence, stronger than strong-mixing is the <u>absolutely regular</u> condition, which was first studied by Volkonskii and Rozanov (1961) and who attributed it to Kolmogorov. Define $\mathcal{B}_p$ and $\mathcal{A}_{p+r}$ as we did above. Define $\zeta(r)$ a function such that:

(3.3) $\beta(r)=\sup_{p} E[\sup_{A\in A_{p+r}} \{|P(A|B_p)-P(A)|]$.

Definition 2.3.3

We say that the stationary stochastic process $\{x_t\}$ is absolutely regular if $\beta(r)$ goes to zero as r goes to infinity.

An alternative characterization given by Volkonskii and Rozanov (1961) is:

(3.4) $\beta(n)=\frac{1}{2} \mathrm{Var}[P_{o,n}-P_{1,n}]$

where $P_{o,n}$ is the measure induced by the process $\{x_i\}_{i=\ldots,-1,0,1,\ldots}$ on the σ-algebra $M_{-\infty}^{o}\cup M_{n}^{\infty}$ and $P_{1,n}$ is the measure defined for $A\in M_{n}^{\infty}$, and $B\in M_{-\infty}^{o}$ given by the equality:

$$P_{1,n}(A\cap B)=P_{o,n}(A)\ P_{o,n}(B).$$

As Yoshihara (1976) noted, this form of characterizing an absolutely regular process is more useful from an empirical point of view.

At this point, it will be worth reciting the fundamental lemma taken by Yoshihara (1976, lemma 1) and by Denker and Keller (1983, lemma 6); which is going to be extensively used along the proofs of theorem 1 of chapter 3 and the theorem of chapter 5.

Lemma 2.3.1

Let $\{x_t,\ t=0,\pm1,\pm2,\ldots\}$ be a stationary absolutely regular process with mixing coefficient $\beta(n)$. Let $t_1<t_2<\ldots<t_k$ be arbitrary integers. For any r $(1\leq r\leq k-1)$ put

$$P_r^{(k)}(E^{(r)}\times E^{(k-r)})=P((x_{t1},\ldots,x_{tr})\in E^{(r)})P((x_{tr+1},\ldots,x_{tk})\in E^{(k-r)})$$

and:

$$P_0^{(k)}(E^{(k)})=P((x_{t_1},\ldots,x_{t_k})\in E^{(k)}).$$

Let $h(x_1,\ldots,x_k)$ be a Borel function such that for any r:

$$\int_{R^{kp}}\int \left|h(x_1,\ldots,x_k)\right|^{1+\delta}dP_r^{(k)}\leq M<\infty \quad \text{for some } \delta>0$$

then:

$$\left|\int_{R^{kp}}\int h(x_1,\ldots,x_k)dP_0^{(k)}-\int_{R^{kp}}\int h(x_1,\ldots,x_k)dP_r^{(k)}\right|\leq$$

$$4M^{1/(1+\delta)}\beta^{\delta/(1+\delta)}(t_{r+1}-t_r).$$

<u>Proof</u>:

See Yoshihara (1976) lemma 1.

If the function $h(x_1,\ldots,x_k)$ is separable in the form $h_1(x_1,\ldots,x_m)h_2(x_{m+1},\ldots,x_k)$, $\beta(n)$ might be replaced by the corresponding strong-mixing coefficient and thus, we obtain essentially the inequality first given by Davydov (1968).

The next measure of dependence which has a close relationship with the absolutely regular one is the condition proposed by Gastwirth and Rubin (1975) (GR), and which is defined as follows. Let Y and Z be $(\ldots,x_{-1},x_0)$ and $(x_r,x_{r+1},\ldots)$ respectively. If the conditional distribution of Z given Y=y exists, define $\Delta_r(y)$ as

(3.5) $$\sup_{|h|\leq 1} \{\int h(z)P_r(dz|y)-\int h(z)P_+(dz)\}$$

where $P_r(dz|y)$ and $P_+(dz)$ denote respectively the conditional probability distribution of Z given Y=y and the probability distribution function of Z. As a measure of dependence between Y and Z we take the L^s norm of Δ_r denoted by $\|\Delta_r\|_s$ $(0<s<\infty)$.

Definition 2.3.4

We say that the stationary stochastic process $\{x_t\}$ satisfies the GR mixing condition if $\|\Delta_r\|_s \longrightarrow 0$ as $r\longrightarrow\infty$ for a given $0<s<\infty$.

It is clear that if $\|\Delta_r\|_1 \longrightarrow 0$ then also $\|\Delta_r\|_s \longrightarrow 0$ for every s, $0<s<\infty$, since $\Delta_r\leq 2$. Also it is worth noting that for s=1 this is exactly the definition of absolutely regular process.

It can be shown $\alpha(r)\leq\zeta(r)$, and $\alpha(r)\leq 4\|\Delta_r\|_1$.

Intuitively, $\alpha(r)$, $\zeta(r)$ or $\|\Delta_r\|_s$ measure the dependence of events in A_{p+r} on those of B_p in terms of how much the probability of the joint ocurrence of an event in each σ-algebra differs from the product of the probabilities of each event occurring.

2.3.2.- Examples and Sufficient Conditions for Mixing Processes

In the previous subsection we have given several measures of dependence on stationary processes. But, once, a particular process is presented, one may ask whether it satisfies one of the above mixing conditions and whether it is possible to know the rate at which the mixing coefficient tends to zero. This is particularly important in order to be able to give examples of processes where the conditions of the theorems in chapters 3 and 5 are satisfied.

The first and more trivial example is the independent identically distributed process. This process for its own definition is GR, absolutely regular and therefore strong-mixing.

The second process to consider is the MA(Q) (moving average) process. Since observations separated by more than Q periods are independent, it implies that the stochastic process is mixing in the sense of any of the above four definitions.

The next process to consider is the linear process, i.e.

$$x_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i} \quad \text{where } \epsilon_i \text{ are i.i.d.}$$

Pham and Tram (1985) studied this model, and gave sufficient conditions under which the above linear process is absolutely regular. In particular they show that

$$\zeta(n) = O\Big(\sum_{k=n}^{\infty} \Big[\sum_{j=k}^{\infty} |\alpha_j|\Big]^{\omega/(1+\omega)}\Big)$$

where $\omega > 0$, if:

i) $\int |g_t(v-u) - g_t(v)| dv < C\|u\|$ for all t, where $g_t(u)$ is the pdf of

ϵ_t,

ii) $\sum_{j=0}^{\infty} |\alpha_j|<\infty$ and $\sum_{j=0}^{\infty} \alpha_j \ z^j \neq 0$ for all $|z|<1$,

iii) $E|\epsilon_t|^{\omega}<\infty$ for all t and $\omega>0$, and

iv) $\sum_{k=1}^{\infty} [\sum_{j=k}^{\infty} |\alpha_j|]^{\omega/(1+\omega)})<\infty$.

Thus, if $|\alpha_j|=O(j^{-\{\alpha(1+\omega)+(1+2\omega)\}/\omega})$ then $\zeta(n)=O(n^{-\alpha})$. Note that for autoregressive moving average processes this is true for any $\alpha>0$, that is $\zeta(n)$ converges exponentially to zero. Thus, up to now, we have given a description of the mixing conditions for any linear process. But what can we say about non-linear models such as bilinears or non-linear autoregressive models? All the theory available about mixing conditions on these models is based on the ergodicity properties of Markov chains. Thus, we will only discuss the mixing conditions for Markov Chains.

Before giving sufficient conditions under which a Markov Chain is ergodic, it is important to ask ourselves: what do we understand by ergodicity in a Markov Chain?

Definition 2.3.5

A Markov Chain is said to be ergodic if there exists an invariant probability function π.

Another concept very useful is geometric ergodicity.

Definition 2.3.6

A Markov Chain is said to be geometrically ergodic if it is positive recurrent and there exists ψ $(0<\psi<1)$ such that:

$\|P^{(t)}(y,\cdot)-\pi\|=O(\psi^t)$ as t goes to infinity a.e. π, and $\forall y\in R$, where $P^{(t)}(y,A)$ stands for the t-th transition probability function from y to A.

This property plays a fundamental role in checking the mixing conditions, as it will be seen in the next theorem. Due to Nummelin and Tuominen (1982), in a Markov chain, geometric ergodicity and absolute regularity are two equivalent properties or concepts.

Theorem 2.3.1

If a stationary stochastic process is geometrically ergodic, then there exists $\psi<1$ such that:

$$\int \pi(dy)\|P^{(t)}(y,\cdot)-\pi(.)\|=O(\psi^t) \text{ as t goes to infinity.}$$

Proof:

See Nummelin and Tuominen (op. cit.).

Recalling the absolutely regular mixing condition, we can realize that it is in fact the above equation. Furthermore, this theorem tells us that the rate of convergence of the absolutely regular process is geometric. Also, we can conclude that the probability distribution function π is absolutely continuous.

Hence, it only remains to see under which circumstances we can say that a Markov chain is geometrically ergodic. Two sets of sufficient conditions will be given. The first set is a theorem by Feigin and Tweedie (1985). The second set is a theorem by Mokkadem (1987).

Theorem 2.3.2

Assume that the stochastic process x_t is a Feller Chain, that is if for every bounded continuous function $Q(x)$ $E[Q(x_{t+1})|x_t=x]$ is a continuous function of x, and that there exist a measure ξ and a compact set A with $\xi(A)>0$ such that:

(i) The chain is ξ-irreducible.

(ii) There exists a non-negative continuous function $g:R\to R$ satisfying $g(x)\geq 1$ $\forall x\epsilon A$. Also for some $\delta>0$

$$E[g(x_t)|x_{t-1}]\leq(1-\delta)g(x) \quad x\epsilon A^c$$

then the chain is geometrically ergodic.

For the next theorem, we suppose that the process is a non-linear autoregressive model defined by:

$$x_t=\rho(x_{t-1})+\epsilon_{t,x_{t-1}}$$

where we stand for $\epsilon_{t,x_{t-1}}$ a process which may depend on x_{t-1}, for instance in a bilinear model this process can be $\epsilon_t x_{t-1}$.

<u>Theorem 2.3.3</u>

Assume that the Markov Chain is aperiodic and μ-irreducible (μ stands for the Lebesgue measure). Also $\exists$ $M>0$, $\nu<1$ and $s>0$ such that:

$$E|\rho(x)+\varepsilon_{t,x}|^s<\nu|x|^s \qquad \forall |x|>M$$

and

$$\sup_{|x|\leq M} E|\rho(x)+\varepsilon_{t,x}|^s<\infty$$

then the chain is geometrically ergodic and π admits a moment of order s.

As a corollary, we can observe that if the function $\rho(\cdot)$ is bounded and the innovation $\varepsilon_{t,x}$ does not depend on x, the chain is geometrically ergodic (Doukhan and Ghindès, 1980).

CHAPTER 3

ADAPTIVE ESTIMATION WITH NON-LINEAR AUTOREGRESSIVE DISTURBANCES

3.1.- INTRODUCTION

In an econometric time series model, serial dependence in the residuals exists when the expectation of those residuals conditional on its own past is different than a constant, that is:

$$E[u_t | u_s,\ s<t] \neq \text{constant}.$$

In fact, serial dependence may be the rule rather than the exception in macro-economic models.

In this chapter, we will consider the multiple regression model:

$$(1.1) \qquad y_t = \beta' x_t + u_t$$

where β and x_t are k-dimensional vectors and y_t and u_t are scalars.

Also it is assumed that u_t has an autoregressive representation of known order P given by the equation(1) (footnotes are at the end of the chapter):

$$(1.2) \quad u_t = \rho(u_{t-1},\ldots,u_{t-P}) + \varepsilon_t,$$

where ε_t, $t=1,2,\ldots$, are innovations which are assumed to be an independent and identically distributed process and with a symmetric distribution function. In the econometric literature up to now, it has been assumed that the ρ function is a linear function, i.e.

$$\rho(u_{t-1},\ldots,u_{t-P}) = \rho_1 u_{t-1} + \ldots + \rho_P u_{t-P}.$$

Under this assumption, the parameters β of the model (1.1), (1.2) can be estimated via a non-linear least squares algorithm, given a generalized least squares (G.L.S.) type estimate. If, in addition, the innovations ε_t are normally distributed, then this estimate will be asymptotically equivalent to the maximum likelihood estimate and hence the Cramèr-Rao bound will be achieved.

A frequent assumption behind linear time series models is Gaussianity. In a Gaussian environment, linear models are appropriate. But once such an hypothesis of Gaussianity is abandoned, two possibilities emerge. The first one emerges when the autoregression function ρ in equation (1.2) is non-linear and the innovations ε_t are either Gaussian or not. The second possibility emerges when the autoregression function is linear but the innovations ε_t are not Gaussian. We share the opinion of Davies, Petruccelli and Watson (1987) by saying that once the hypothesis of Gaussianity is abandoned the first possibility is the most likely that it happens. Thus, non-linear time series models are called for.

As it was said in the first chapter, recently non-linear time series models have become popular among statisticians. These models, unlike linear ones, may capture characteristics like limit cycles and jumping behaviour which are observed in many time series. The jumping behaviour in a process may be viewed as a process which fluctuates around one of its stable singular points(2) and jumps from one stable singular point to another depending on the innovations of the model. As regards the limit cycle, it may be observed in time series which have a periodic character but the time series model is not constrained to be sinusoidal. The latter type of series implies that the system has a "self-exciting" mechanism, i.e. for small values of the process the system tends to "explode" while for large values the system tends to "damp down" towards its mean, say zero.

We know of no attempt to study the statistical properties of estimators of β in the model (1.1) where u_t may follow a possibly non-linear AR model. Hence, it would be interesting to study the statistical properties of the least squares estimators of β ($\hat{\beta}$) when the autoregression function ρ is given but not constrained to be linear, i.e.

$$\min_{\beta} \sum_{t=P+1}^{T} (y_t-\beta' x_t-\rho(y_{t-1}-\beta' x_{t-1},\ldots,y_{t-P}-\beta' x_{t-P}))^2=$$

$$\sum_{t=P+1}^{T} (y_t-\hat{\beta}' x_t-\rho(y_{t-1}-\hat{\beta}' x_{t-1},\ldots,y_{t-P}-\hat{\beta}' x_{t-P}))^2$$

where $\hat{\beta}$ may be obtained via a non-linear minimization algorithm procedure.

Once the hypothesis of linearity of the ARMA models is relaxed, allowing for further structures as the bilinear, smooth threshold autoregressive (S.T.A.R.), or exponential autoregressive models, to identify such structures may be a very difficult task. However, it is worthwhile noting that the hypothesis of linearity can be tested in an AR framework via nonparametric methods by using a test proposed by Robinson (1983). On the other hand, if the interest of the practitioner is not to model the residual structure itself, but to get estimates of β, one can ask himself if there is an estimation method whose asymptotic performance is as good as the estimates that we would get if the structure of the residuals were perfectly known.

To do so, we use nonparametric estimators of the non-linear autoregression function of the u_t's and, as will be shown in theorem 1 below, we obtain an estimator of β which is asymptotically as efficient (to first order) as the least squares estimator $\hat{\beta}$. Thus we would be able to adapt to the unknown autoregressive structure of the residuals. This semiparametric estimator would be, if the innovations ε_t were normal, Cràmer-Rao efficient because $\hat{\beta}$ would be the maximum likelihood estimator.

One question arises from all this discussion. In the linear framework it is very well known that when the true structure of the autoregression is mis-specified, then the estimator of the regression parameters may be inefficient (see for instance Hannan (1970) Ch. 7 for an extensive discussion on this point). Moreover, it may happen that this estimator is more inefficient than the O.L.S., see for instance Engle (1974). Hence, in our setting, if the model is estimated in the belief that the residuals follow a linear AR model, what can we say about their asymptotic covariance matrices of that mis-specified estimator and the least squares estimator $\hat{\beta}$?

For simplicity, let us suppose until the end of this section that there is no intercept, x_t is scalar and P=1.

The asymptotic variance of the β estimator will be given by:

$$(1.3) \qquad \sigma_\epsilon^2 \left[\text{plim } T^{-1} \sum_{t=2}^{T} (x_t - x_{t-1}\rho'(u_{t-1}))^2\right]^{-1}$$

where $\rho'(.)$ is the derivative of the function ρ, and σ_ϵ^2 is the variance of ϵ_t. In terms of the spectrum (1.3) is equal to:

$$(1.4) \qquad \sigma_\epsilon^2 \left\{ \int_{-\pi}^{\pi} f_{\overline{xx}}(\lambda)\{1+E(\rho'(u)^2)-2E\rho'(u) \cos\lambda\} \, d\lambda\right\}^{-1}$$

where $f_{\overline{xx}}(\lambda)$ is the spectrum of x_t.

One can mis-specify the ρ-function, and let us suppose that the investigator believes that the ρ-function is a linear function instead of the true non-linear one. Under this belief, he would estimate the parameters of the model via a Cochrane-Orcutt or Durbin algorithm, for example, that is we do a G.L.S. as if the u_t has spectrum $f_o(\lambda)$. Therefore, its asymptotic variance is:

$$(1.5) \qquad \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} f_x(\lambda) f_o^{-2}(\lambda) f(\lambda) d\lambda}{\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f_x(\lambda) f_o^{-1}(\lambda) d\lambda\right]^2}$$

where $f(\lambda)$ is the actual spectrum of u_t. By the Cauchy-Schwartz inequality we have that:

$$(1.6)\quad (1.5) \geq \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f_x(\lambda) f^{-1}(\lambda) d\lambda \right]^{-1}$$

where the RHS of (1.6) is the Gauss-Markov bound achieved when $f_o(\lambda)=f(\lambda)$ and also by Hannan's estimate.

Intuition tells us that the asymptotic variance of the parametric estimator of β, i.e. (1.4), is lower than (1.5) (the variance under the mis-specification). A sufficient condition for (1.5) $\geq$ (1.4) is:

RHS of (1.6) $\geq$ (1.4)

and a sufficient condition for the last expression is:

$$(1.7)\qquad \sigma_\varepsilon^2/2\pi \leq f(\lambda)\{1+E(\rho'(u)^2)-2E\rho'(u)\cos\lambda\}.$$

Unfortunately, I do not know how to get a formula for the spectrum of a non-linear AR model. Also, I do not see how the derivatives of the ρ-function enter. However, there is a possibility about the last point which consists in computing numerically these expectations using Jones' (1978) approach.

This point deserves much more attention, since it will be interesting to know under which circumstances equation (1.6) is satisfied. It would be more interesting and odd, if one finds an example where the mis-specified model performs better than the parametric estimator. This would be in view of the maximum likelihood interpretation of the least squares estimator very strange, since in this case we will achieve the Cràmer-Rao bound. Instead, we will present some Monte-Carlo evidence in the next chapter about this interesting point, where the M.S.E. of the estimates of β is studied under the true and false specification of the residuals u_t, as well as for the semiparametric estimator of β which will be introduced in section 3 below.

The remaining of the chapter is organized as follows: In §2 we will describe the kernel estimators which we are going to use. In §3 we will introduce the semiparametric estimator of the parameters β. Also, we will discuss a computational method for such estimators, based on a modification of the well known Cochrane-Orcutt algorithm for the linear AR(1) model. In §4 we will specify the concept of mixing condition that will be employed. In §5, we are going to describe the conditions on the model under which the semiparametric estimator of β is root-T consistent, and has the same asymptotic distribution as the least squares estimator, and finally in §6, a summary of the chapter and possible extensions are given. The lengthy proofs of the theorems are in appendices.

3.2.- THE KERNEL ESTIMATOR

Let z be a random variable and let g(z) be a Borel function such that $E|g(z)|<\infty$. Let K be a real bounded function such that:

$$\int_R K(u)du=1.$$

Let us also assume a positive number a which depends on T, and it goes to zero as T approaches to infinity. Then we define the

Watson-Nadaraya type estimator of the conditional expectation of $g(z_t)$ given $y_t=y$ by definition 2.1.2 of chapter 2, i.e.:

$$\frac{1}{Ta}\sum_{t=1}^{T} g(z_t)K\left[\frac{y-y_t}{a}\right] \Big/ \frac{1}{Ta}\sum_{t=1}^{T} K\left[\frac{y-y_t}{a}\right]$$

where $\sum K((y_t-y)/a)/Ta$ estimates the probability density function of the process y_t at the point y, i.e. f(y), and $\sum g(z_t)K(y_t-y/a)/Ta$ estimates $f(y)E[g(z_t)|y_t=y]$.

Thus, if $g(z_t)=u_t$ and $y_t=u_{t-1}$ we have that an estimator of $E[u_t|u_{t-1}=u]$, given u_t t=1,2,..., will be given by:

$$(2.1)\quad \frac{1}{Ta}\sum_{t=2}^{T} u_t\, K\left[\frac{u-u_{t-1}}{a}\right] \Big/ \frac{1}{Ta}\sum_{t=2}^{T} K\left[\frac{u-u_{t-1}}{a}\right] .$$

It is worth noting that the K function is not constrained to be a density function itself because, as it will be seen in section 5, we want to allow K to take negative values.

The above estimator of $\rho(u_{j-1})=E[u_t|u_{t-1}=u_{j-1}]$ is not generally antisymmetric, but a modified version is. We modify it as follows (and for technical reasons we also use the "leave-one-out" estimator):

$$(2.2)\quad \hat{\rho}(u_{j-1})=\frac{1}{2Ta}\sum_{\substack{t=2\\ t\neq j}}^{T} u_t\{K((u_{t-1}-u_{j-1})/a)-K((u_{t-1}+u_{j-1})/a)\}\hat{f}^{-1}(u_{j-1})$$

where $\hat{f}(u_{j-1})$ is the estimator of the probability density function of u_{j-1} as defined in definition 2.1.1. But this estimator is not generally symmetric, as the density function of u_{j-1} is by a result of Pemberton and Tong (1981), thus we modify it as follows:

$$\frac{1}{2Ta}\sum_{\substack{t=2\\t\neq j}}^{T}\{K((u_{t-1}-u_{j-1})/a)+K((u_{t-1}+u_{j-1})/a)\}-\hat{f}(u_{j-1}).$$

If we are interested in an estimator of the derivative of $\rho(u_{j-1})=E[u_t|u_{t-1}=u_{j-1}]$ with respect to u_{j-1}, this will be obtained by differentiating the equation (2.2) with respect to u_{j-1}, i.e. definition 2.1.3 of chapter 2.

3.3.- THE SEMIPARAMETRIC ESTIMATOR OF β

Let the observations $\{y_t,x_t\}_{t=1,...,T}$ be from a random variable (Y,X), where x_t is a k-dimensional vector. The least squares estimator of the parameters of the model (1.1) which takes into account that the residuals follow a possible non-linear AR(1) model, i.e.

(3.1) $u_t=\rho(u_{t-1},\theta)+\epsilon_t \qquad t=2,...,T$

where ρ is an antisymmetric function in u_{t-1} and θ is a qx1 vector of parameters, may be obtained, for instance, by the Gauss-Newton iteration procedure:

(3.2) $$\begin{bmatrix}\tilde{\beta}\\ \tilde{\theta}\end{bmatrix}=\begin{bmatrix}\bar{\beta}\\ \bar{\theta}\end{bmatrix}+(\sum\partial\epsilon_t/\partial\lambda\,\partial\epsilon_t/\partial\lambda')^{-1}\sum\partial\epsilon_t/\partial\lambda\,\epsilon_t|_{\lambda=[\bar{\beta}',\bar{\theta}']'}$$

where the summands from now on, unless otherwise specified, run from t=2 up to t=T, and $(\bar{\beta}',\bar{\theta}')'$ is a previous estimate of the parameters $(\beta',\theta')'$.

If the initial estimator of the parameters is root T-consistent then the second round in the iteration procedure is asymptotically equivalent to the least squares estimator $\hat{\beta}$ and $\hat{\theta}$ which minimizes the sum of squares

$$\sum(y_t-\beta' x_t-\rho(u_{t-1},\theta))^2.$$

As was said in the introduction, when the autoregression function of the residuals u_t is linear, i.e. $\rho(u_{t-1},\theta)=\theta u_{t-1}$, the Cochrane-Orcutt (1949) algorithm is asymptotically equivalent to the Gauss-Newton procedure described in equation (3.2). It is worthwhile writing down here the Cochrane-Orcutt algorithm, since the estimation method is based on a non-linear version of this algorithm:

1st step: The O.L.S. estimator of β is obtained, i.e $\beta^{o.l.s.}$.

2nd step: An estimator of θ, θ^1 is obtained by regressing $u_t^{o.l.s.}$ on $u_{t-1}^{o.l.s.}$, where $u_t^{o.l.s.}$ for t=1,2,... are the O.L.S. residuals.

3rd step: We regress $y_t-\theta^1 y_{t-1}$ on $x_t-\theta^1 x_{t-1}$, and obtaining and updating estimator of β.

Repeat steps 2 and 3 up to the convergence of the parameter estimates is obtained.

Step 3 can be rewritten as:

(3.3) $$\tilde{\beta}=\bar{\beta}+(\sum (x_t-x_{t-1}\bar{\theta})(x_t-x_{t-1}\bar{\theta})')^{-1}\sum (x_t-x_{t-1}\bar{\theta})(u_t-\bar{\theta}u_{t-1}).$$

This method (Cochrane-Orcutt) is an iterative technique which disregards the first observation. Therefore, it will be equivalent to the G.L.S. estimator only asymptotically. We could adopt a G.L.S. estimator by noting that the expected value of $\varepsilon_1{}^2$ conditional on u_1 is $(1-\theta^2)u_1{}^2$ and to minimize

$$\sum (y_t-\beta' x_t-\theta(y_{t-1}-\beta' x_{t-1}))^2+(y_1-\beta' x_1)^2(1-\theta^2)$$

which is known as the iterative Prais-Winsten (1954) algorithm, instead of minimizing the sum of squares

$$\sum (y_t-\beta' x_t-\theta(y_{t-1}-\beta' x_{t-1}))^2.$$

The Cochrane and Orcutt and Prais-Winsten are two particular algorithms. Instead, we can use more general non-linear methods, as the Gauss-Newton or Davidon-Fletcher-Powell iteration techniques, or searching algorithms, by selecting a number of values of θ over the interval $[-1,1]$, and to minimize simultaneously upon β and θ.

If the innovations ε_t were normal, the maximum likelihood function of β and θ after concentrating out $\sigma_\varepsilon{}^2$ (the variance of the disturbance ε_t) and ignoring constants would be:

$$[1-\theta^2]^{1/2}[\sum (y_t-\beta' x_t-\theta(y_{t-1}-\beta' x_{t-1}))^2+(y_1-\beta' x_1)^2(1-\theta^2)]^{-T/2}$$

or alternatively:

$$1/2\log(1-\theta^2)-T/2\log[\sum (y_t-\beta' x_t-\theta(y_{t-1}-\beta' x_{t-1}))^2+(y_1-\beta' x_1)^2(1-\theta^2)].$$

The M.L.E. would be obtained by maximizing the above equation in β and θ by iterative or search algorithms. Since the first term of the above equation does not depend on T, the G.L.S. and the M.L.E., i.e. the $\tilde{\beta}$ which maximizes the above term, will be asymptotically equivalent.

Also, we are aware that the estimation procedure for this simple AR(1) model can be generalized to the AR(P) or ARMA(P,Q) models, via iterative and search algorithms, see for example Harvey (1981) or Judge et al. (1985) for several estimation procedures.

Before introducing the semiparametric estimator of the parameters β, a word about the reason why we use a non-linear version of the Cochrane-Orcutt algorithm is worthwhile. The Cochrane-Orcutt method is a step-wise optimization algorithm. It is based on the fact that in the Gauss-Newton procedure (in the linear AR case) the asymptotic covariance matrix for the estimator of β on the one hand and the estimator of θ on the other is block diagonal, i.e. they are asymptotically independent.

In the non-linear framework, unlike the linear model, a closed formula for the expected value of $\epsilon_1{}^2$ conditional on u_1 is not generally possible. Hence, a "full" optimization algorithm where "all" the observations were taken into account, as in the Prais-Winsten (1954) algorithm, for instance, would be difficult to implement or impossible.

However, when the autoregression function of the residuals is non-linear then θ and θu_{t-1} in equation (3.3) have to be substituted by:

$$\frac{\partial\rho(u_{t-1},\theta)}{\partial u_{t-1}}$$

and $\rho(u_{t-1},\theta)$ respectively, given the announced non-linear version of the "Cochrane-Orcutt" algorithm.

By a result of Pemberton and Tong (1981), a necessary and sufficient condition for the probability density function of the process u_t to be symmetric is that the autoregression function is antisymmetric given that the disturbances of the model, i.e. ε_t, has a symmetric probability density function. This result will allow us to say that also in the non-linear framework, the estimator of β and θ are asymptotically independent. So, as long as our interest is focussed on the asymptotic behaviour of β, we can assume that the parameter θ is known throughout the remaining of the chapter, since this parameter may be considered as a nuisance parameter.

On the other hand, in our model no functional form of the autoregression function is assumed or specified and therefore neither $\rho(u,\theta)$ or its derivative with respect to u is known. Thus, we must substitute it by a nonparametric estimator as described in §2 and given by definitions 2.1.2 and 2.1.3 of chapter 2, which do not depend on any functional form, i.e. equation (2.2) and its derivative with respect to u_{j-1}. Since in the model the process u_t is not observed, we should replace the u_t by an estimator $\bar{u}_t$ given by

$$\bar{u}_t=y_t-\bar{\beta}'x_t$$

which is updated at each iteration. (From now on, we will always let $\bar{u}_t$ be the t-th residual computed from the estimator of β that we will consider).

From now on, we are going to restrict ourselves, for technical reasons, to discrete sequences of estimators $\bar{\beta}_T$ according to:

Let $\bar{\beta}$ be an initial estimator (for instance the O.L.S.) of β such that:

$$\sqrt{T}(\bar{\beta}-\beta)=O_{p_\beta}(1)$$

then the estimator $\bar{\beta}_T$ is given by one of the vertices of $\{\beta$ such that $\beta=\sqrt{T}^{-1}(i_1,i_2,\dots,i_k)'$, $i_1,\dots,i_k\epsilon Z\}$ nearest to $\bar{\beta}$.

Of course, $\bar{\beta}_T$ satisfies the following more general discreteness property.

Definition 3.3.1

A sequence $\{\bar{\beta}_T\}$ of estimates is called discrete if there exists a $C\epsilon N$ such that independently of $T\epsilon N$ $\bar{\beta}_T$ takes on at most C different values in:

$$Q_n=\{\beta_0\epsilon R^p \ ; \ \sqrt{T}|\beta_0-\beta|<c\} \qquad c>0 \text{ fixed.}$$

Lemma 3.3.1

Let $\{\xi_T(\beta), T\epsilon N\}$ be a sequence of random variables which depends on $\beta\epsilon\Theta_1$. If for each sequence $\{\beta_T\}\subseteq\Theta_1$ satisfying that $\sqrt{T}(\beta_T-\beta)$ is bounded by a constant $c>0$ and

$$\xi_T(\beta_T)=o_p(1) \qquad \text{when } \beta \text{ holds,}$$

then also $\xi_T(\bar{\beta}_T)=o_p(1)$ holds under β for discrete estimators $\{\bar{\beta}_T\}$ which are root T-consistent.

Proof:

See Kreiss (1987), lemma 4.4.

From this lemma, it will only be necessary to establish theorem 1 in section 5 for sequences of estimators of the form $\beta_T=\beta+h_T\sqrt{T}^{-1}$, where h_T is a bounded sequence, because of $\bar{\beta}_T$ is a root T-consistent estimator and its intersection with any sphere of radius $M\sqrt{T}^{-1}$ center in β is finite with cardinality bounded independently of T. Therefore β_T may be seen as a generalization of the discretizised estimator $\bar{\beta}_T$.

We are now going to introduce the concept of contiguity of two probability measures, see for instance the monograph by Roussas (1972), as a criterion of nearness of sequences of probability measures.

Definition 3.3.2

It is said that two measures P_T and Q_T are contiguous if and only if for every measurable sequence of events $A_T \subseteq R^T$, $T=1,\ldots,\infty$ $\int dQ_T \longrightarrow 0$ implies that $\int dP_T \longrightarrow 0$ for such an A_T as $T \longrightarrow \infty$ and the other way round.

Consequently, in order to prove theorem 1 of section 5, we will prove that the probability measures generated by the computed residuals $\bar{u}_t=y_t-\bar{\beta}_T'x_t$ and by the true ones $u_t=y_t-\beta'x_t$, that is the probability measures generated by $\bar{\beta}_T$ and by β are contiguous. The importance of this result is that if we are able to prove that theorem 1 of section 5 is true under the probability measure generated by the computed residuals then it is also for the probability measure generated by the true residuals. Thus, in view of lemma 3.3.1., the theorem 1 is also true for the discrete estimator $\bar{\beta}_T$.

Lemma 3.3.2

The measures generated by $\{\bar{u}_t\}_{t=1,2,\ldots}$ and $\{u_t\}_{t=1,2,\ldots}$ are contiguous.

Proof

See appendix C.

As in other uses of kernel regression, the estimator $\hat{\rho}(\bar{u}_{t-1})$ is technically difficult to handle owing to the random denominator of $\hat{\rho}(\bar{u}_{t-1})$ unless we assume that u_t has a compact support and its probability density function is bounded away from zero. Therefore, in order to avoid such an assumption, we "trim" out small values of $\hat{f}(\bar{u}_{t-1})$ as in Bickel (1982), Manski (1984) and Robinson (1988c). To do that, we define b>0 which tends to zero as T goes to infinity and $\bar{I}_{t-1}=I(|\hat{f}_{t-1}|>b)$ where I stands for the usual indicator function. Therefore, the semiparametric estimator of β will be:

$$(3.4)\quad \beta^{*}=\bar{\beta}_T+\Big(\sum\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big]\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big]'\Big)^{-1}$$

$$\times\sum\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big](\bar{u}_t-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1}).$$

where $\partial/\partial u_{t-1}\hat{\rho}(\bar{u}_{t-1})$ is obtained by differentiating equation (2.2) w.r.t. u_{t-1} and evaluating it at $\bar{u}_{t-1}$.

3.4.- MIXING CONDITION

As it was seen in chapter 2, the literature on mixing conditions is very vast. Also throughout the years many definitions of mixing processes have been discussed in the statistical literature.

For our purposes we are going to use the absolutely regular condition, that as it was seen in the earlier chapter is stronger than the strong mixing condition. On the other hand, this condition is very similar to the one proposed by Gastwirth and Rubin (1975), being in fact identical for s=1 (see chapter 2 for further details).

The reason why this mixing condition is used and not others is because it is easier to handle in relation with the U-statistics and V-statistics that we will encounter along the proof of theorem 1 in section 5, and to be able to use the results of Yoshihara (1976) on U-statistics for those processes. However, the strong mixing condition could be used by employing the characteristic version of the kernel function as it was employed by Robinson (1984) although we have, in that case, to strengthen the conditions of theorem 1.

3.5.- CONDITIONS AND THEOREMS

In the parametric case it was seen that the estimator of β was given by the equation:

$$(5.1)\quad \hat{\beta}=\bar{\beta}_T+\Big(\sum\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big]\times\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big]'\Big)^{-1}$$

$$\times\sum\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big](\bar{u}_t-\rho(\bar{u}_{t-1},\theta_T))$$

while in the nonparametric case was given by the equation:

$$(5.2)\quad \beta^*=\bar{\beta}_T+\Big(\sum\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big]\times\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big]'\Big)^{-1}$$

$$\times\sum\Big[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\Big](\bar{u}_t-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1}).$$

The adaptiveness of the semiparametric estimator β^* will be established in two steps. In the first one, it will be shown that $\sqrt{T}(\hat{\beta}-\beta^*)$ is $o_p(1)$. In the second one, it will be proved that the parametric estimator $\hat{\beta}$ is square root T consistent and moreover $\sqrt{T}(\hat{\beta}-\beta)$ converges in law to a Normal distribution with zero mean and variance-covariance matrix $\sigma_\epsilon^2V^{-1}$.

From equations (5.1) and (5.2) it is easily seen that $\sqrt{T}(\hat{\beta}-\beta^*)$ is equal to:

$$\Big(\frac{1}{T}\sum\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big]\times\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big]'\Big)^{-1}$$

$$\times\Big(\frac{1}{T^{\frac{1}{2}}}\sum\Big[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big](\bar{u}_t-\rho(\bar{u}_{t-1},\theta_T))\Big)-$$

$$(\frac{1}{T}\sum\left[x_t-x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\bar{I}_{t-1}\right]\times\left[x_t-x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\bar{I}_{t-1}\right]')^{-1}$$

$$\times\ (\frac{1}{T^{\frac{1}{2}}}\sum\left[x_t-x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\bar{I}_{t-1}\right]\ (\bar{u}_t-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1})).$$

In order to prove that $T^{\frac{1}{2}}(\hat{\beta}-\beta^*)=o_p(1)$, it will be sufficient to show that the next two expressions, namely:

$$(5.3)\ \frac{1}{T}\sum\left[x_t-x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\bar{I}_{t-1}\right]\times\left[x_t-x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\bar{I}_{t-1}\right]'-$$

$$\frac{1}{T}\sum\left[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\right]\times\left[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\right]'$$

and

$$(5.4)\ \frac{1}{T^{\frac{1}{2}}}\sum\left[x_t-\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}x_{t-1}\bar{I}_{t-1}\right][\bar{u}_t-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1}]$$

$$-\left[x_t-x_{t-1}\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\right][\bar{u}_t-\rho(\bar{u}_{t-1},\theta_T)]$$

are both $o_p(1)$.

Recalling that the probability measures under (β,θ) and $(\bar{\beta}_T,\theta)$, i.e. $P_{(\beta,\theta)}$ and $P_{(\beta,\theta)}$ at $\beta=\bar{\beta}_T$, are contiguous, to prove that (5.3) and (5.4) are $o_p(1)$, it will suffice to show that they are $o_p(1)$ under the probability measure induced by $(\bar{\beta}_T,\theta)$, e.g. by $\bar{u}_t=y_t-\bar{\beta}_T x_t$, since by lemma 3.3.1, it will be true that (5.3) and (5.4) are both $o_p(1)$ under the probability measure generated under (β,θ).

(5.4) can be split into the following three terms, i.e.

$$(5.5)\ 1/T^{\frac{1}{2}}\ \sum(\rho(\bar{u}_{t-1},\theta_T)-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1})(x_t-\rho'(\bar{u}_{t-1},\theta_T)x_{t-1}),$$

$$(5.6)\ \frac{1}{T^{\frac{1}{2}}}\sum x_t\bar{\epsilon}_t\Big(\Big[\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\,\bar{I}_{t-1}\Big]-\Big[\frac{\partial\rho(\bar{u}_{t-1},\theta_T)}{\partial u_{t-1}}\Big]\Big)$$

and

$$(5.7)\ \frac{1}{T^{\frac{1}{2}}}\sum x_{t-1}\frac{\partial\hat{\rho}(\bar{u}_{t-1})}{\partial u_{t-1}}\,\bar{I}_{t-1}(\rho(\bar{u}_{t-1},\theta_T)-\hat{\rho}(\bar{u}_{t-1})\bar{I}_{t-1}).$$

As long as interest is focussed on the asymptotic distribution of $\hat{\beta}$ and to show that $\sqrt{T}(\hat{\beta}-\beta^*)=o_p(1)$, in view of the discussion in section 3.3 it will be necessary to show that (5.3), (5.5)-(5.7) are $o_p(1)$, where instead of writing $\rho(u_{t-1};\theta_T)$ or its derivatives with respect to u_{t-1}, we can write $\rho(u_{t-1};\theta)$ or its derivatives with respect to u_{t-1} in order to prove theorem 1 below.

Also in theorem 2, we will give a proof for the case in which the ρ-function is perfectly known, i.e. its functional form and its parameters θ.

We will adopt the "higher-order" kernel approach to bias-reduction proposed by Barlett (1963). Since a sufficiently smooth function behaves locally like a polynomial of a certain degree, we can exploit this property by using a kernel function with enough zero moments, so that the bias can be reduced rapidly enough with a. For a construction of such kernel functions we refer to Robinson (1988c) or Prakasa-Rao (1983).

Definition 3.5.1

K_r $r>1$ is the class of functions $K:R\rightarrow R$ satisfying:

$$\int_R u^i K(u)du=\delta_{io} \quad i=0,\ldots,r-1$$

where δ_{ij} is the Kronecker delta.

The classes K_r confer increasingly small bias on kernel estimators as r increases, but also increasingly large variance, the latter varying directly with $\int K(u)^2 du$. However, the asymptotic distribution of β^* is independent of K, detecting no disadvantage in a K_r, with even arbitrary large r. Nevertheless, it would be surprising if, in finite samples, β^* did not to some extend inherit variance properties of the kernel estimators from which it is formed. Then, while increasingly r cannot shrink, and may widen, the band of a-sequences satisfying the theorem 1, we caution against choosing r too large.

Some further, practically unrestrictive, conditions on K will be imposed, such as boundedness and thin tails. Introduce:

Definition 3.5.2

$N_\mu=\{K:R\rightarrow R;\int |u|^\mu |K(u)|du<\infty\}$.

Definition 3.5.3

$S_{u\mu}=\{v:v\epsilon R/|u-v|<\gamma\}\ \forall u\epsilon R,\ \gamma>0$.

Choosing $K\epsilon K_r$ will not work unless the function ρ and f are collectively sufficiently smooth, and it seems reasonable to suppose that the smoother they are the better β^* will on average be.

Definition 3.5.4

$$\Lambda_\mu = \{K: R \to R; \sup_u |u|^\mu |K(u)| < \infty\}.$$

Definition 3.5.5

$\Xi^\alpha = \{g: R \to R; \int |g(u)|^\alpha f(u)du < \infty\}$ and f(.) is the probability density function of u.

Definition 3.5.6

X_μ^α is the class of g functions belonging to Ξ^α such that there exist $h \epsilon \Xi^\alpha$ and some $\gamma > 0$ such that:

$$\sup_{v \epsilon S_{u\gamma}} \frac{|g(u)-g(v)-Q(u,v)|}{|u-v|^\mu} \leq h(u), \text{ a.e. } (u)$$

for integer m such that $m-1 < \mu \leq m$

$$Q(u,v) = \sum_{r=1}^{m-1} \frac{1}{r!} \left[\frac{\partial^r g(u)}{\partial u^r}\right] (u-v)^r$$

and $\dfrac{\partial^r g(u)}{\partial u} \epsilon \Xi^\alpha$.

Loosely speaking, X_μ^α consists of functions g which can be expanded in a Taylor series of degree (m-1) and with a remainder term of degree μ, whose coefficients have a finite α-th moment as do g and its first m-1 partial derivatives.

ASSUMPTIONS

A1.- The process u_t is a stationary absolutely regular process with finite $2+\delta$-moments for some $\delta>0$, and where the mixing coefficient $\zeta(n)$, given in definition 2.3.3, satisfies $\zeta(n)=O(n^{-[(2+\delta')/\delta']})$ for some $\delta'>0$ and $0<\delta'<\delta$.

A2.- x_t are either stochastic or nonstochastic, and in the former is a stationary and ergodic vector process with finite second-moments.

A3.- x_t and u_s are independent for all t and s for the stochastic x_t.

A4.- the ρ function belongs to $X_\lambda^{2+\delta}$ where $\lambda>2$ and $\delta>0$.

A5.- $f(u)$, the probability density function of u_t, belongs to F_λ^∞ where $\lambda>2$.

A6.- $Ta^6b^4\to\infty$ as $T\to\infty$.

A7.- $T^{\frac{\gamma}{2}}a^{1+\left[\frac{1+\delta}{2+\delta}\right]}b\to\infty$ where:

$$\gamma=\frac{2(\delta-\delta')}{\delta'(2+\delta)}$$ with δ and δ' as in A1.

A8.- $a^{\lambda-2}b^{-2}\to 0$, where λ is as in A4 or A5.

A9.- $Ta^{4\lambda-2}b^{-4}\to 0$.

A10.- The kernel function K belongs to K_4, and has first continuous derivatives.

A11.- The characteristic function of K, $\kappa(v)$, defined by

$$\kappa(v)=\int_R K(u)e^{iuv}du$$

is such that $|v\kappa(v)|$ is integrable.

<u>Theorem 1</u>

Under A1 to A11:

$$\sqrt{T}\ (\beta^{*}-\hat{\beta})=o_p(1).$$

<u>Proof</u>:

See appendix A.

<u>Remark</u>:

In appendix A, we will use u_t instead of $\bar{u}_t$ to make easier the notation.

Before writing down the assumptions needed for theorem 2, a comment about the assumptions A6-A9 is necessary.

While A6-A9 prevent b from converging to 0 too fast, there is nothing to stop it converging arbitrary slowly. If the rate of convergence of $\zeta(n)$ is geometric then γ can be taken equal to 1. In that case, A7 is weaker than A6. This will be the case for all the examples that they will be studied in the next chapter.

For sake of argument consider the case γ=1. A sufficient condition for reconciling A6, A8 and A9 is: $b^{-1}=o(a^{2-\lambda/4})$, and thus, A6: $T^{-1}a^{-(4+\lambda)}\longrightarrow 0$ or finite, and A9: $Ta^{3\lambda}\longrightarrow 0$ or finite. Thus, to reconcile the last two, we can take $a=O(T^{-1/3\lambda})$.

A kernel function which satisfies A10 and A11 and also definitions 3.5.2 and 3.5.3 is:

$$(1.5-0.5u^2)K(u)$$

where $K(u)=1/\sqrt{2\pi}\exp(-1/2u^2)$.

For the asymptotic convergence of $\sqrt{T}(\hat{\beta}-\beta)$, it will be assumed the following conditions:

<u>ASSUMPTIONS</u>

B1.- Let ϵ_t be an i.i.d. stochastic process such that:

$E|\epsilon_t|^{2r}<\infty$ where r>1.

B2.- $$E\left[x_t-x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\right]\times\left[x_t-x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\right]'$$

exists and is positive definite.

B3.- $\left[\dfrac{\partial^2\rho(u_t)}{\partial u_t^2}\right]^2$ has finite 1+δ moments, where δ>0.

__Theorem 2__

Under B1-B3 and also A1-A4 we have that $\sqrt{T}(\hat{\beta}-\beta)$ is asymptotically $N(0,\sigma_\epsilon^2 V^{-1})$ where V is equal to $X'\Phi X$ and $\Phi=\mathrm{diag}(1,1+E\rho'(u_t)^2,\ldots,1+E\rho'(u_t)^2)-E\rho'(u_t)(J+J')$ where J is a matrix whose (i,j) element is 1 if j=i+1 and i=1,...,T-1 and zero otherwise. This V matrix is consistently estimated by:

$$\left[\frac{1}{T}\sum\left[x_t-x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\right]\times\left[x_t-x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\right]'\right]$$

and being σ_ϵ^2 the variance of the residuals ϵ_t.

__Proof__:

See appendix B.

3.6.- __EXTENSIONS AND CONCLUSIONS__

We have dealt with a very simple linear regression model where the residuals follow an AR(1) model, but where its functional form is unknown to the practitioner. A semiparametric estimator of the regression parameters has been proposed, whose asymptotic distribution function is the same as that of the parametric estimator, although we have posed an open question about the efficiency of the estimator, which we are only able to answer by a Monte-Carlo experiment given in the next chapter.

Several extensions could come to our minds in view of the studied model, but three of them are the most obvious.

The first extension to the model is when the residuals follow a non-linear AR(P) with P>1. In view of the proof of theorem 1, it is obvious that the proof for this case is only a mathematical complication in notation terms, although they are the same.

The second extension is when we have lagged dependent variables in the regression model, i.e. we have y_{t-j} in the x_t vector. This extension could be more interesting from an econometric point of view, since in a lot of applied work lagged dependent variables appear in the regression equation.

In this case, the problem of mis-specification might be more important, because we can have inconsistencies in the estimation of the parameters affecting the lagged dependent variables, by using estimation procedures as the Gauss-Newton algorithm, or Hatanaka (1974) algorithm.

The last possible extension of our model by inspecting the proofs of our theorems 1 and 2 is the non-linear regression model.

Like in most of the nonparametric literature there exists a problem in how to choose the "bandwidth" parameter a. This problem is far from being solved, although some attempts have been made. However, some possible cross-validation could be applied to our problem. This entails a huge computational burden, and whether for this choice of a theorem 1 still holds is a question that remains to be answered. Even the consistency of the semiparametric estimator with this choice of the bandwidth parameter may be difficult to prove.

Footnotes

(1) Although we are going to discuss the case of a general P in the introduction, we are going to discuss only the case for P=1 in the remainder of the chapter, mainly to ease the notation and arguments.

(2) We understand for stable singular point, a point in which in the absence of pertubations in the system, this, the system, will tend to that point. This is analogue to the concept of stable point in a differential equation system.

APPENDIX A

NOTATION

From now on we will denote $K_{j,z}$ and $K_{j,-z}$ by:

$$K_{j,z}=K\left[\frac{z_j-z}{a}\right] \quad \text{and} \quad K_{j,-z}=K\left[\frac{z_j+z}{a}\right].$$

The derivatives of the above equations with respect to z are going to be named $K'_{j,z}$ and $K'_{j,-z}$ respectively.

Also we write $\bar{K}_{j,z}$ and $\tilde{K}_{j,z}$ for:

$$\bar{K}_{j,z}=K\left[\frac{z_j-z}{a}\right]-K\left[\frac{z_j+z}{a}\right]$$

and

$$\tilde{K}_{j,z}=K\left[\frac{z_j-z}{a}\right]+K\left[\frac{z_j+z}{a}\right] \qquad \text{respectively.}$$

The first lemma is taken from Robinson (1988c), but owing to its importance we think that it is worthwhile reproducing it here.

Lemma 1

Let $f(u)g(u)$ be a function that belongs to X_λ^∞, and the kernel function $K\epsilon\Lambda_0\cap N_\lambda\cap K_r$. Then, if λ satisfies $r-1<\lambda<r$ where $r>1$

$$\text{(A.1)} \qquad \sup_z \left| \int \frac{1}{a} K_{u,z} g(u)f(u)du - f(z)g(z) \right| = O(a^\lambda).$$

Proof:

Call the function $g(u)f(u)=h(u)$. Since $h(u)\epsilon X_\lambda^\infty$, there exists a function $Q(u,z)$ which is equal to:

$$h(z)+\sum_{i=1}^{r-1} \frac{1}{i!} (u-z)^i \frac{\partial^i}{\partial u^i} h(z).$$

Define $R=E_{z\rho}\cup\bar{E}_{z\rho}$, where $\rho>0$ and $E_{z\rho}=\{u/|z-u|<\rho\}$. Define also $A(z;E)$ by:

$$\int_E a^{-1} K\left[\frac{u-z}{a}\right]\{h(u)-h(z)\}du.$$

Therefore, (A.1) is less than or equal to:

$$\sup_z |A_1(z;E_{z\rho})| + \sup_z |A_2(z;E_{z\rho})| + \sup_z |A(z;\bar{E}_{z\rho})|$$

where:

$$A_1(z;E_{z\rho}) = \int_{E_{z\rho}} a^{-1} K\left[\frac{u-z}{a}\right]\{Q(u,z)-h(z)\}du$$

and

$$A_2(z;E_{z\rho}) = \int_{E_{z\rho}} a^{-1} K\left[\frac{u-z}{a}\right]\{h(u)-Q(u,z)\}du.$$

Now then, $A_1(z;R)=0$ since the integrand has only elements of the form:

$$\int u^i K(u)du$$

and they are equal to zero since $K\epsilon K_r$. Therefore $A_1(z;E_{z\rho})\equiv -A_1(z;\bar{E}_{z\rho})$. Because r>1, we have that:

$$|Q(u,z)-h(z)|I(u\epsilon\bar{E}_{z\rho})\leq M\sum_{s=1}^{r-1}|u-z|^s I(u\epsilon\bar{E}_{z\rho})$$

$$\leq M\ |u-z|^\lambda I(u\epsilon\bar{E}_{z\rho}).$$

Recall that we are in the set $\bar{E}_{z\rho}$, and here $|u-z|>\rho>0$. This implies: $|u-z|<M|u-z|^2<\ldots<M|u-z|^\lambda$. Hence, applying lemma 1 of Robinson (1988c) we can conclude that:

$$\sup_z|A_1(z;E_{z\rho})|=O(a^\lambda).$$

Because $|h(u)-Q(u,z)|\leq M\ |u-z|^\lambda$, $u\epsilon E_{z\rho}$ it implies that:

$$\sup_z\ |A_2(z;E_{z\rho})|=O(a^\lambda) \text{ as well.}$$

Finally,

$$\sup_z\ |A(z;\bar{E}_{z\rho})|\ \leq 2\sup_z h(z)\ \int_{\bar{E}_{z\rho}} a^{-1}\left|K\left[\frac{u-z}{a}\right]\ \right|du$$

$$\leq M\ a^{-1}\rho^{-\lambda}\int\ |u-z|^\lambda\left|K\left[\frac{u-z}{a}\right]\right|du$$

$$=O(a^\lambda)$$

by lemma 1 of Robinson (1988c).

Lemma 2

Let z_t be a stationary strong mixing stochastic process as it was defined in definition 2.3.2., with strong mixing coefficient $\alpha(n)$ such that:

$$\sum_n \alpha(n)^{\delta/2+\delta} < \infty$$

and $g(z_t)$ has finite $2+\delta$ moments, and $\delta>0$. Let the function $g(z)f(z)$ be such that it belongs to X_λ^∞ and All, then:

$$\text{(A.2)} \qquad \sup_z \left| \frac{1}{Ta} \sum_{j=1}^{T} g(z_j) K\left[\frac{z_j - z}{a}\right] - g(z)f(z) \right| = O_p(a^\lambda + T^{-\frac{1}{2}} a^{-1}).$$

Proof:

In order to prove the lemma we are going to use the trick of employing the characteristic function of the K function, and hence to be able to avoid the term in z.

The left hand side of (A.2) is bounded by (A.3)+(A.4), where:

$$\text{(A.3)} \qquad \sup_z \left| \frac{1}{Ta} \sum_{j=1}^{T} g(z_j) K_{j,z} - E g(z_j) K_{j,z} \right|$$

and

$$\text{(A.4)} \qquad \sup_z \left| \frac{1}{Ta} \sum_{j=1}^{T} E g(z_j) K_{j,z} - f(z) g(z) \right|.$$

- For (A.4) we apply lemma 1 to conclude that it is $O(a^{\lambda})$.

- For (A.3), we have, by taking into account that $\kappa(z)$ is the characteristic function of K(z), that the expectation of this term is equal to:

$$E \sup_{z} \left| \int \kappa(av) e^{-iuz} \left[\frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{-iuz_j} - Eg(z_j) e^{-iuz_j} \right] dv \right|$$

$$\text{(A.5)} \quad < \int |\kappa(av)| \; E \left| \frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{iuz_j} - Eg(z_j) e^{iuz_j} \right| dv.$$

But, on the other hand we have that :

$$E \left| \frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{iuz_j} - Eg(z_j) e^{iuz_j} \right|^2 = O(T^{-1})$$

since the L.H.S. of this equation is equal to:

$$\frac{1}{T^2} \sum_{j=1}^{T} \mathrm{Var}[g(z_j) e^{iuz_j}] + \frac{2}{T^2} \sum_{m<j}^{T} \mathrm{Cov}(g(z_j) e^{iuz_j}, g(z_m) e^{iuz_m}).$$

By assumption, $g(z_j)$ has finite $2+\delta$ moments, so the first term is clearly $O(T^{-1})$. About the second term, its absolute value is bounded by:

$$\frac{2}{T} \sum_{j=1}^{T} \alpha(j)^{2/2+\delta} E|g(z_j) e^{iuz_j}|^{2+\delta}$$

by Davydov (1968) inequality, and as the function $g(z_j)$ has finite

$2+\delta$ moments we have that also this term is $O(T^{-1})$. Therefore the equation (A.5) is $O_p(a^{-1}T^{-\frac{1}{2}})$ by also using A11.

Q.E.D.

Lemma 3

Let z_t be a stationary strong mixing stochastic process as it was defined in definition 2.3.2., with strong mixing coefficient $\alpha(n)$ such that:

$$\sum_n \alpha(n)^{\delta/2+\delta} < \infty$$

and $g(z_t)$ has finite $2+\delta$ moments, and $\delta>0$. Let the function $\partial/\partial z\, g(z)f(z)$ be such that it belongs to X_λ^∞ then:

$$\text{(A.6)}\quad \sup_z \left| \frac{1}{Ta^2}\sum_{j=1}^{T} -g(z_j)K'\left[\frac{z_j-z}{a}\right] - \partial/\partial z\, g(z)f(z)\right| = O_p(a^\lambda + T^{-\frac{1}{2}}a^{-2}).$$

Proof:

We are going to prove the lemma by the usual trick of employing, once again, the characteristic function of the kernel function K, and so to avoid the term in z.

The left hand side of (A.6) is bounded by (A.7)+(A.8), where:

$$\text{(A.7)}\qquad \sup_z \left| \frac{1}{Ta^2}\sum_{j=1}^{T} -g(z_j)K'_{j,z} - Eg(z_j)K'_{j,z}\right|$$

and

$$\text{(A.8)} \quad \sup_{z} \left| \frac{1}{Ta^2} \sum_{j=1}^{T} Eg(z_j)K'_{j,z} - \partial/\partial z f(z)g(z) \right| .$$

- As far as (A.8) is concerned, this is equal to:

$$\sup_{z} \left| \int \frac{1}{a^2} K'_{j,z} g(z_j) f(z_j) dz_j - \partial/\partial z f(z)g(z) \right|$$

and calling $g(z)f(z)=h(z)$, we have that this term is equal, when we integrate by parts, to:

$$-\frac{1}{a} K\left[\frac{z_j - z}{a}\right] h(z_j) \Big|_{-\infty}^{\infty} + \int \frac{1}{a} K\left[\frac{z_j - z}{a}\right] h'(z_j) dz_j - h'(z)$$

and arguing as we did in (A.4) of lemma 2 we have that is $O(a^{\lambda})$.

- About (A.7), we have that the expectation of this term is equal to (taking into account that $\kappa(z)$ is the characteristic function of $K(z)$):

$$E \sup_{z} \left| \int iv\kappa(av) e^{-iuz} \left[\frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{-iuz_j} - Eg(z_j) e^{-iuz_j} \right] dv \right|$$

$$\text{(A.9)} \quad < \int |v\kappa(av)| \, E\left| \frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{iuz_j} - Eg(z_j) e^{iuz_j} \right| dv.$$

But, on the other hand we have that:

$$E\left| \frac{1}{T} \sum_{j=1}^{T} g(z_j) e^{iuz_j} - Eg(z_j) e^{iuz_j} \right|^2 = O(T^{-1})$$

if we argue as we did in the preceding lemma. Therefore (A.9) is

$O_p(a^{-2}T^{-\frac{1}{2}})$, and the lemma is proved.

Q.E.D.

For the next lemma we are going to use the following notation:

(A.10) $$K_{i,j}^{(r)} = \frac{1}{a^{r+1}} K^r\left[\frac{z_j - z_i}{a}\right]$$

where $K^r(u)$ stands for the r-th derivative of K(u).

Lemma 4

Let $\{z_j\}_{j=1,2,\ldots}$ be an absolutely regular stationary stochastic process, as it was defined in definition 2.3.3.. Then for r>1 and α>0, and assuming that the r-th derivative of the probability density function of z_j exists, then:

$$E|K_{i,j}^{(r)}|^{\alpha} = O(a^{(1-\alpha)-r\alpha}).$$

Proof:

Pretend for the moment that z_i and z_j are independent. Then we have that:

$$E|K_{ij}^{(r)}|^{\alpha} = a^{(1-\alpha)-r\alpha} \int |K^{(r)}(v)|^{\alpha} f(z_i) f(z_i + av) dz_i dv$$

$$< M a^{(1-\alpha)-r\alpha} \int |K^{(r)}(v)| dv$$

(A.11) $$= O(a^{(1-\alpha)-r\alpha})$$

where M is a generic constant.

But (A.11) also holds for absolutely regular processes in view of lemma 8 of Denker and Keller (1983).

Q.E.D.

Lemma 5

Let v be a random variable with probability density function f(x) such that $f(x)\neq 0$. Then:

$E\{I_v\}\longrightarrow 1$ as $b\longrightarrow 0$, where:

$$I_v=\begin{cases} 1 & \text{if } f(v)>b \\ 0 & \text{otherwise.} \end{cases}$$

Proof:

It is immediate.

Lemma 6

Let $\hat{g}(u)$, $\hat{h}(u)$ and $\hat{f}(u)$ be the kernel estimators of the functions g(u), h(u) and f(u) respectively such that:

(A.12) $\sup_u|\hat{g}(u)-g(u)|=O_p(T^{-\frac{1}{2}}a^{-\alpha}+a^{\lambda})$,

(A.13) $\sup_u|\hat{h}(u)-h(u)|=O_p(T^{-\frac{1}{2}}a^{-\beta}+a^{\lambda})$

and

(A.14) $\sup_u|\hat{f}(u)-f(u)|=O_p(T^{-\frac{1}{2}}a^{-1}+a^{\lambda})$

and assuming that $\lambda(u_t)$ is $O_p(1)$ and $I_t = I(f(u_{t-1})>b)$ then:

$$\text{(A.15)} \quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=1}^{T} \lambda(u_t)\frac{(\hat{g}(u_t)-g(u_t))(\hat{h}(u_t)-h(u_t))}{\hat{f}(u_t).\ f(u_t)} I_t$$

$$= O_p\left[\frac{T^{-\frac{1}{2}}a^{-(\alpha+\beta)}+a^{2\lambda}T^{\frac{1}{2}}+a^{\lambda-\alpha}+a^{\lambda-\beta}}{b^2-bc_\sigma[T^{-\frac{1}{2}}a^{-1}+a^{\lambda}]}\right].$$

<u>Proof</u>:

Using the facts that

1) $(f(u_t)-\sup_{u_t}|\hat{f}(u_t)-f(u_t)|)I_t \leq |\hat{f}(u_t)|I_t$

and

2) The L.H.S. of the above inequality, i.e. $(f(u_t)-\sup|\hat{f}(u_t)-f(u_t)|)I_t$ is a positive random variable except in a set S which measure tends to zero as T goes to infinity,

we have that the absolute value of the L.H.S. of equation (A.15) is less than or equal to, for T sufficiently large:

$$\frac{\sup_t|\hat{g}(u_t)-g(u_t)|\ \sup_t|\hat{h}(u_t)-h(u_t)|}{f^2(u_t)-f(u_t)\sup_t|\hat{f}(u_t)-f(u_t)|} I_t \frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\lambda(u_t)|I_t$$

but by equations (A.12)-(A.14) and because $|\lambda(u_t)|$ is $O_p(1)$ we have that this last term is in fact:

$$O_p\left[\frac{T^{-\frac{1}{2}}a^{-(\alpha+\beta)}+a^{2\lambda}T^{\frac{1}{2}}+a^{\lambda-\sigma}+a^{\lambda-\beta}}{b^2-bc_\sigma[T^{-\frac{1}{2}}a^{-1}+a^{\lambda}]}\right].$$

Q.E.D.

Lemma 7

Let u_t be a stationary absolutely regular stochastic process, as it was defined in definition 2.3.3.. Assume A1 and A6. Let $\gamma=2(\delta-\delta')/\delta'(2+\delta)>0$. Then:

$$\text{(A.16)} \quad \sum_{t=2}^{T} \Big[\frac{1}{f(u_{t-1})} \sum_{s=2} \varepsilon_s a^{-1}\bar{K}_{t-1,s-1} \Big] I_{t-1} = O_p\big[T^{(3/2-\gamma/2)} a^{-(1+\delta/2+\delta)} b^{-1}\big].$$

Proof:

We have that the L.H.S. of equation (A.16) is equal to:

$$\text{(A.17)} \qquad \sum_{t<s}^{T} \frac{1}{f(u_{t-1})} \varepsilon_s a^{-1}\bar{K}_{t-1,s-1} I_{t-1} + \frac{1}{f(u_{s-1})} \varepsilon_t a^{-1}\bar{K}_{s-1,t-1} I_{s-1}.$$

Calling $y_t=(\varepsilon_t, u_{t-1})$ we have that (A.17) can be written as:

$$\text{(A.18)} \qquad \sum_{t<s}^{T} g(y_s, y_t).$$

On the other hand, we have that the expectation of the function $g(y_s, y_t)$ with respect to y_s or y_t is equal to zero, since by construction the function $\bar{K}_{t-1,s-1}$ is antisymmetric in its arguments. This implies that as long as the function g is symmetric in its arguments we can apply the theory already done for U-statistics to conclude that the expected square of (A.18) is less than or equal to:

$$b^{-2}T^{3-\gamma}\left[E\left[\frac{1}{a}\bar{K}_{t-1,s-1}\right]^{2+\delta}\right]^{2/2+\delta}$$

and by lemma 4 this is in fact:

$$O_p\left[T^{3-\gamma}a^{-2\left[\frac{1+\delta}{2+\delta}\right]}b^{-2}\right].$$

Q.E.D.

<u>Lemma 8</u>

Let u_t be a stationary stochastic process satisfying the same conditions as in lemma 7. If in addition we assume A4 and A8 then:

$$\text{(A.19)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}^{T}\left[\frac{\hat{f}(u_{t-1})-f(u_{t-1})}{f(u_{t-1})}\right]\rho(u_{t-1})I_{t-1}=o_p(1).$$

<u>Proof</u>:

The L.H.S. of (A.19) is equal to:

$$\frac{1}{2T^{1\frac{1}{2}}}\sum_{t,s}^{T}\left[\frac{\tilde{K}_{t-1,s-1}}{a}-2f(u_{t-1})\right]\frac{\rho(u_{t-1})}{f(u_{t-1})}I_{t-1}=$$

$$\text{(A.20)}\quad \frac{1}{2T^{1\frac{1}{2}}}\sum_{t<s}^{T}\left[\left[\frac{\tilde{K}_{t-1,s-1}}{a}-\frac{1}{a}E_{s-1}\tilde{K}_{t-1,s-1}\right]\frac{\rho(u_{t-1})}{f(u_{t-1})}I_{t-1}+\right.$$

$$\left.\left[\frac{\tilde{K}_{s-1,t-1}}{a}-\frac{1}{a}E_{t-1}\tilde{K}_{s-1,t-1}\right]\frac{\rho(u_{s-1})}{f(u_{s-1})}I_{s-1}\right]$$

$$\text{(A.21)}\quad +\frac{1}{2T^{\frac{1}{2}}}\sum_{t=2}^{T}\left[\left[\int\frac{\tilde{K}_{t-1,s-1}}{a}f(u_{s-1})du_{s-1}\right]-2f(u_{t-1})\right]\frac{\rho(u_{t-1})}{f(u_{t-1})}I_{t-1}.$$

As regards (A.20), call $g(u_{t-1},u_{s-1})$ the term inside the square brackets. It is easy to show that the expectation of this term over each of the arguments when they are taken as independent is equal to zero. Also, as long as the function g(.,.) is symmetric in its arguments we can use lemma 2 of Yoshihara (1976) on U-statistics to conclude that it is $o_p(1)$. In fact this term is:

$$O_p\left[T^{-\gamma/2} a^{-\left[\frac{1+\delta}{2+\delta}\right]} b^{-1}\right].$$

- As far as (A.21) is concerned, it can be rewritten as:

$$\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}^{T} h(u_{t-1})$$

where $h(u_{t-1})$ is equal to the term inside the summand of (A.21) and it has zero expectation. Thus, as long as u_{t-1} is absolutely regular, we have that the second moments are bounded by:

$$\frac{1}{T}\sum_{t=2}^{T} Eh^2(u_{t-1}) + \sum_{n=1}^{T} \varsigma(n)^{\delta/2+\delta}(Eh(u_{t-1})^{2+\delta})^{2/2+\delta}$$

but we have that the expectation of the term in the square brackets of (A.21) is o(1) by lemma 2. Therefore the lemma is proved.

Q.E.D.

- **NOTATION**

$$\hat{g}(u)=f(u)\hat{\rho}(u)=\frac{1}{2Ta}\sum_{s=2}^{T} u_s \bar{K}_{s-1,u}$$

and $g(u)=f(u)\rho(u)$.

In view of the contiguity property, I will write always u_t instead of $\bar{u}_t$, where $\bar{u}_t=y_t-\beta_T'x_t$. Also I_t will stand for:

$$\begin{cases} 1 & \text{if } f(u_t)>b \\ 0 & \text{otherwise} \end{cases}$$

and $b\longrightarrow 0$ as $T\longrightarrow\infty$.

From now on, t=2 or s=2 below the summand sign means that the summand goes from t=2 up to t=T or s=2 up to s=T, unless otherwise specified.

Proposition 1

Let u_t be a stationary stochastic process satisfying A1, A4-A11. Then if $f(u)\rho(u)\epsilon X_X^{\infty}$:

$$\text{(A.22)} \quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} (\rho(u_{t-1})-\hat{\rho}(u_{t-1})I_{t-1})\rho'(u_{t-1})=o_p(1).$$

Proof:

The L.H.S. of (A.22) is equal to:

$$(A.23)\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\{\rho(u_{t-1})-\hat{\rho}(u_{t-1})\}\rho'(u_{t-1})I_{t-1}$$

$$(A.24)\quad +\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\rho(u_{t-1})\rho'(u_{t-1})\{1-I_{t-1}\}.$$

- About (A.24), we have that:

$$E((A.24)^2)=\frac{1}{T}\sum_{t=2}E\rho^2(u_{t-1})\rho'^2(u_{t-1})\{1-I_{t-1}\}^2$$

$$+\frac{2}{T}E\sum_{t<s}^{T}(\rho(u_{t-1})\rho'(u_{t-1})\{1-I_{t-1}\})(\rho(u_{s-1})\rho'(u_{s-1})\{1-I_{s-1}\})$$

which is $o_p(1)$ by lemma 5 and the mixing condition of the stochastic process u_t. Note that $E(\rho(u_t)\rho'(u_t)\{1-I_t\})=0$, by symmetry of the probability density function of u_t.

- As far as (A.23).

This term can be split up into two terms:

$$(A.25)\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\left[\rho(u_{t-1})-\frac{\sum_{s=2}\rho(u_{s-1})\bar{K}_{t-1,s-1}}{\sum_{s=2}\tilde{K}_{t-1,s-1}}\right]\rho'(u_{t-1})I_{t-1}$$

and

$$(\text{A.26})\qquad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\frac{\sum_{s=2}\epsilon_s\bar{K}_{t-1,s-1}}{\sum_{s=2}\tilde{K}_{t-1,s-1}}\,\rho'(u_{t-1})I_{t-1}.$$

(A.26) is the easiest one. This is equal to

$$(\text{A.27})\qquad \frac{1}{aT^{1\frac{1}{2}}}\sum_{t=2}\frac{\sum_{s=2}\epsilon_s\bar{K}_{t-1,s-1}}{f(u_{t-1})}\,\rho'(u_{t-1})I_{t-1}$$

$$(\text{A.28})\quad +\frac{1}{aT^{1\frac{1}{2}}}\sum_{t=2}\frac{(\sum_{s=2}\epsilon_s\bar{K}_{t-1,s-1})(\hat{f}(u_{t-1})-f(u_{t-1}))}{\hat{f}(u_{t-1}).f(u_{t-1})}\,\rho'(u_{t-1})I_{t-1}.$$

By lemma 7 and since $E|\rho'(u_{t-1})|<M<\infty$, we can conclude that (A.27) is:

$$O_p\left[T^{-\gamma/2}a^{-\left[\frac{1+\delta}{2+\delta}\right]}b^{-1}\right].$$

Using the facts that:

$$\sup_t\left|\frac{1}{Ta}\ (\sum_{s=2}\epsilon_s\bar{K}_{t-1,s-1})\right|=O_p(T^{-\frac{1}{2}}a^{-1})$$

and

$$\sup_t\ (\hat{f}(u_{t-1})-f(u_{t-1}))=O_p(T^{-\frac{1}{2}}a^{-1}+a^{\lambda})$$

by lemma 6, we conclude that (A.28) is:

$$=O_p\left[\frac{T^{-\frac{1}{2}}a^{-2}+a^{\lambda-1}}{b^2-bc_\sigma[T^{-\frac{1}{2}}a^{-1}+a^{\lambda}]}\right]=o_p(1)\ \text{if:}$$

$Ta^4b^4\longrightarrow\infty$, and $a^{\lambda-1}b^{-2}\longrightarrow 0$.

Thus, the term (A.26) is $o_p(1)$. Therefore, in order to prove the proposition it only remains to show that (A.25) is also $o_p(1)$. Now then, it is very easy to see that it is equal to:

$$\text{(A.29)} \quad \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \frac{(\hat{g}(u_{t-1})-g(u_{t-1}))(\hat{f}(u_{t-1})-f(u_{t-1}))}{\hat{f}(u_{t-1})\cdot f(u_{t-1})} I_{t-1}\rho'(u_{t-1})$$

$$\text{(A.30)} \quad + \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \rho(u_{t-1})\frac{(\hat{f}(u_{t-1})-f(u_{t-1}))^2}{\hat{f}(u_{t-1})\cdot f^2(u_{t-1})} I_{t-1}\rho'(u_{t-1})$$

$$\text{(A.31)} \quad + \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \rho(u_{t-1})\frac{(\hat{f}(u_{t-1})-f(u_{t-1}))}{f(u_{t-1})} I_{t-1}\rho'(u_{t-1})$$

$$\text{(A.32)} \quad + \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \frac{(\hat{g}(u_{t-1})-g(u_{t-1}))}{f(u_{t-1})} I_{t-1}\rho'(u_{t-1}).$$

About (A.29) and (A.30), we apply lemma 6 to conclude that they are $o_p(1)$. About (A.31) we call for lemma 8. Therefore it only remains to show that (A.32) is $o_p(1)$ to conclude the proposition. But this is equal to:

$$\frac{1}{2T^{1\frac{1}{2}}}\sum_{t<s}^{T} \Big[\Big[\rho(u_{s-1})\frac{\overline{K}_{t-1,s-1}}{a} - 2f(u_{t-1})\rho(u_{t-1})\Big] \frac{1}{f(u_{t-1})} I_{t-1}\rho'(u_{t-1})$$

$$\text{(A.33)} \quad +\Big[\rho(u_{t-1})\frac{\overline{K}_{s-1,t-1}}{a} - 2f(u_{s-1})\rho(u_{s-1})\Big] \frac{1}{f(u_{s-1})} I_{s-1}\rho'(u_{s-1}) \Big].$$

Let $h(u_{t-1},u_{s-1})$ be the term inside the summamd sign and $\overline{h}(u_{t-1})=E_{s-1}h(u_{t-1},u_{s-1})$

$$-\left(\int\rho(u_{s-1})\bar{K}_{t-1,s-1}f(u_{s-1})/a\,du_{s-1}\ -2f(u_{t-1})\rho(u_{t-1})\right)\frac{\rho'(u_{t-1})}{f(u_{t-1})}I_{t-1}.$$

We have that (A.33) is equal to:

$$\frac{1}{2T^{1\frac{1}{2}}}\sum_{t<s}^{T} h(u_{t-1},u_{s-1})-\bar{h}(u_{t-1})-\bar{h}(u_{s-1})$$

$$+\ \frac{1}{2T^{\frac{1}{2}}}\sum_{t=2}^{T}\bar{h}(u_{t-1}).$$

For the first term we apply once again lemma 2 of Yoshihara (1976). And for the second one, arguing as we did in (A.21) of lemma 8 we have that is equal to $o_p(1)$.

With this term we have concluded the proof of proposition 1.

NOTATION

For the next proposition, $g(u)={}^{\partial}/_{\partial u}f(u)\rho(u)$ and $\hat{g}(u)$ is the nonparametric estimator of this g function.

Proposition 2

Let u_t be a stationary absolutely regular process, as it was defined in definition 2.3.3., satisfying A1 to A11. Then:

$$\text{(A.34)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}^{T}\epsilon_t\left[\frac{\partial}{\partial u_{t-1}}\ \rho(u_{t-1})I_{t-1}-\ \frac{\partial}{\partial u_{t-1}}\rho(u_{t-1})\right]=o_p(1).$$

Proof:

The L.H.S. of (A.34) is equal to:

$$\text{(A.35)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t \rho'(u_{t-1})\{1-I_{t-1}\} +$$

$$\text{(A.36)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t\{\hat{\rho}'(u_{t-1})-\rho'(u_{t-1})\}I_{t-1}.$$

- About (A.35), only it has to be noted that ε_t is an i.i.d. process with finite second moments ($\rho'(u)$ may be considered as a bounded function) and using lemma 5 we conclude that this term is $o_p(1)$.

-About (A.36). This term is equal to:

$$\text{(A.37)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t \frac{\frac{1}{Ta}\sum_{s=2} \varepsilon_s \bar{K}'_{t-1,s-1}}{\hat{f}(u_{t-1})} I_{t-1}$$

$$\text{(A.38)}\quad + \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t \frac{\frac{1}{Ta}\sum_{s=2} \varepsilon_s \bar{K}_{t-1,s-1}}{\hat{f}(u_{t-1})} \quad \frac{\hat{f}'(u_{t-1})}{\hat{f}(u_{t-1})} I_{t-1}$$

$$\text{(A.39)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t \frac{\frac{1}{Ta}\sum_{s=2} \rho(u_{s-1})\bar{K}'_{t-1,s-1}}{\hat{f}(u_{t-1})} - \frac{g(u_{t-1})}{f(u_{t-1})} I_{t-1}$$

$$\text{(A.40)}\quad + \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} \varepsilon_t\left[\left[\frac{\frac{1}{Ta}\sum_{s=2} \rho(u_{s-1})\bar{K}_{t-1,s-1}}{\hat{f}(u_{t-1})} \cdot \frac{\hat{f}'(u_{t-1})}{\hat{f}(u_{t-1})}\right] - \right.$$

$$\frac{\rho(u_{t-1})}{f(u_{t-1})} f'(u_{t-1})]I_{t-1}.$$

- About (A.37). We are going to do the same trick as usual, i.e. to split it up into two terms:

$$\text{(A.41)} \quad \frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \varepsilon_t \frac{\frac{1}{Ta} \sum_{s=2} \varepsilon_s \bar{K}'_{t-1,s-1}}{f(u_{t-1})} I_{t-1}$$

and

$$\text{(A.42)} \quad \frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \varepsilon_t \frac{1}{Ta} \sum_{s=2} \varepsilon_s \bar{K}'_{t-1,s-1} \left[\frac{1}{\hat{f}(u_{t-1})} - \frac{1}{f(u_{t-1})}\right] I_{t-1}.$$

As regards (A.41), this term is equal to:

$$\frac{1}{aT^{1\frac{1}{2}}} \sum_{t<s}^{T} \left[\varepsilon_s \varepsilon_t \frac{\bar{K}'_{t-1,s-1}}{f(u_{t-1})} I_{t-1} + \varepsilon_s \varepsilon_t \frac{\bar{K}'_{t-1,s-1}}{f(u_{s-1})} I_{s-1}\right]$$

and, once again, a straightforward application of Yoshihara (1976) lemma 2, when we take $(\varepsilon_t, u_{t-1})=z_t$, and bearing on mind that E_s^z $h(z_s, z_t)=0$, then this term is $o_p(1)$. (By E_s^z, is meant the conditional expectation on z_s).

As far as (A.42) is concerned, it is just only an application of lemma 6 once again, and therefore we have that it is also $o_p(1)$. Hence (A.37) is $o_p(1)$. In fact (A.42) is:

$$\left[\frac{T^{-\frac{1}{2}}a^{-3}+a^{\lambda-2}}{b^2-bc_\sigma[T^{-\frac{1}{2}}a^{-1}+a^\lambda]}\right].$$

- About (A.38). Once again we are going to split up into two terms:

$$\text{(A.43)} \quad \frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \epsilon_t \frac{1}{Ta} \sum_{s=2} \epsilon_s \bar{K}_{t-1,s-1} \frac{f'(u_{t-1})}{\hat{f}^2(u_{t-1})} I_{t-1}$$

$$\text{(A.44)} \quad + \frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \epsilon_t \frac{1}{Ta} \sum_{s=2} \epsilon_s \bar{K}_{t-1,s-1} \frac{(f'(u_{t-1})-\hat{f}'(u_{t-1}))}{\hat{f}^2(u_{t-1})} I_{t-1}.$$

(A.44) is $o_p(1)$ as we did with (A.42) since:

$$\sup_t |\sum_s \epsilon_s \bar{K}_{t-1,s-1}| = O_p(T^{\frac{1}{2}})$$

and

$$\sup_t |f'(u_{t-1})-\hat{f}'(u_{t-1})| = O_p(T^{-\frac{1}{2}}a^{-2}+a^{\lambda-1})$$

and if $f \epsilon F_\lambda^\infty$ it implies that $f' \epsilon F_{\lambda-1}^\infty$. Hence (A.44) is

$$= O_p\left[\frac{T^{-\frac{1}{2}}a^{-3}+a^{\lambda-2}}{b^2-bc_\sigma[T^{-\frac{1}{2}}a^{-1}+a^\lambda]}\right] = o_p(1).$$

About (A.43), we have that it is equal to:

$$\frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \epsilon_t \frac{1}{Ta} \sum_{s=2} \epsilon_s \bar{K}_{t-1,s-1} \frac{f'(u_{t-1})}{f^2(u_{t-1})} I_{t-1} +$$

$$\frac{1}{T^{\frac{1}{2}}} \sum_{t=2} \epsilon_t \frac{1}{Ta} \sum_{s=2} \epsilon_s \bar{K}_{t-1,s-1} \left[\frac{f'(u_{t-1})}{\hat{f}^2(u_{t-1})} - \frac{f'(u_{t-1})}{f^2(u_{t-1})}\right] I_{t-1}.$$

Therefore, the first term is $o_p(1)$ by U-statistic theory (see Yoshihara, 1976), and using assumption A7 for the rate of convergence, and the second by taking the supremums as it was done in (A.44) for instance, and using assumptions A8 and A9. Thus, (A.38) is also $o_p(1)$.

- As regards (A.39). As we did in equation (A.25) of proposition 1, this term can be written down as follows:

$$\text{(A.45)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}^{T} \varepsilon_t \frac{(\hat{g}(u_{t-1})-g(u_{t-1}))(\hat{f}(u_{t-1})-f(u_{t-1}))}{\hat{f}(u_{t-1}).f(u_{t-1})} I_{t-1}+$$

$$\text{(A.46)}\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}^{T} \varepsilon_t \frac{g(u_{t-1})(\hat{f}(u_{t-1})-f(u_{t-1}))^2}{\hat{f}(u_{t-1}).f^2(u_{t-1})} I_{t-1} +$$

$$\text{(A.47)}\quad \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \varepsilon_t \frac{g(u_{t-1})\ (\hat{f}(u_{t-1})-f(u_{t-1}))}{f^2(u_{t-1})} I_{t-1} +$$

$$\text{(A.48)}\quad \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \frac{(\hat{g}(u_{t-1})-g(u_{t-1}))}{f(u_{t-1})} \varepsilon_t I_{t-1}.$$

- For (A.45) and (A.46) we use lemma 6 to conclude that they are $o_p(1)$, by assumptions A8 and A9.

- As regards (A.48). It is easy to show that this term is equal to:

$$\frac{1}{T^{1\frac{1}{2}}}\sum_{t\neq s}^{T} \frac{\varepsilon_t}{f(u_{t-1})} \left[\frac{1}{a}\, \rho(u_{s-1})\bar{K}_{t-1,s-1}^{i} - \left[\frac{\partial}{\partial u_{t-1}} f(u_{t-1})\rho(u_{t-1})\right] I_{t-1}\right.$$

and in a more compact notation to:

$$\text{(A.49)} \quad \frac{1}{T^{1\frac{1}{2}}} \sum_{t<s} h(y_t, y_s)$$

where $y_t=(\epsilon_t, u_{t-1})$ and $h(y_t, y_s)$ is equal to:

$$\frac{\epsilon_t}{f(u_{t-1})}\left[\frac{1}{a}\rho(u_{s-1})\bar{K}'_{t-1,s-1}-\left[\frac{\partial}{\partial u_{t-1}} f(u_{t-1})\rho(u_{t-1})\right] I_{t-1}+\right.$$

$$\frac{\epsilon_s}{f(u_{s-1})}\left[\frac{1}{a}\rho(u_{t-1})\bar{K}'_{s-1,t-1}-\left[\frac{\partial}{\partial u_{s-1}} f(u_{s-1})\rho(u_{s-1})\right] I_{s-1}.\right.$$

Therefore (A.49) can be written as:

$$\text{(A.50)} \quad \frac{1}{2T^{1\frac{1}{2}}} \sum_{t<s}^{T} h(y_{t-1}, y_{s-1})-\bar{h}(y_{t-1})-\bar{h}(y_{s-1}),$$

$$\text{(A.51)} \quad \frac{1}{2T^{\frac{1}{2}}} \sum_{t=2} \bar{h}(y_{t-1})$$

and

$$\bar{h}(y_{t-1})=\frac{\epsilon_t}{f(u_{t-1})}\left[\frac{1}{a} E_{y_s}(\rho(u_{s-1})\bar{K}'_{t-1,s-1})-(\frac{\partial}{\partial u_{t-1}} f(u_{t-1})\rho(u_{t-1})\right] I_{t-1}.$$

Thus, the term (A.51) is $o_p(1)$ since $\bar{h}(y_{t-1})$ is a martingale difference and by lemma 3

$$\sup_t \left|\frac{1}{a} E_{y_s}(\rho(u_{s-1})\bar{K}'_{t-1,s-1})-(\frac{\partial}{\partial u_{t-1}} f(u_{t-1})\rho(u_{t-1})\right| I_{t-1}=o_p(1)$$

and as far as the term (A.50) is concerned, once again, we apply the U-statistic methodology to say that it is

$$o_p\left[T^{-\gamma/2} a^{-\left[\frac{1+\delta}{2+\delta}\right]-1} b^{-1}\right].$$

- As far as (A.47) is concerned, this term is equal to:

$$\frac{1}{2T^{1\frac{1}{2}}} \sum_{t\neq s}^{T} \varepsilon_t \frac{g(u_{t-1})}{f^2(u_{t-1})} \left[\frac{1}{a} \tilde{K}_{t-1,s-1} - 2f(u_{t-1})\right] I_{t-1}$$

which it can be written into the sum of

$$\text{(A.52)} \quad \frac{1}{2T^{1\frac{1}{2}}} \sum_{t<s}^{T} h(y_{t-1}, y_{s-1}) - \bar{h}(y_{t-1}) - \bar{h}(y_{s-1})$$

and

$$\text{(A.53)} \quad \frac{1}{2T^{\frac{1}{2}}} \sum_{t=2} \bar{h}(y_{t-1})$$

where $h(y_{t-1}, y_{s-1}) =$

$$\varepsilon_t \frac{g(u_{t-1})}{f^2(u_{t-1})} \left[\frac{1}{a} \tilde{K}_{t-1,s-1} - 2f(u_{t-1})\right] I_{t-1} +$$

$$\varepsilon_s \frac{g(u_{s-1})}{f^2(u_{s-1})} \left[\frac{1}{a} \tilde{K}_{s-1,t-1} - 2f(u_{s-1})\right] I_{s-1}$$

and $\bar{h}(y_{t-1}) = \frac{\varepsilon_t}{f^2(u_{t-1})} \left[\frac{1}{a} E_{y_s}(\tilde{K}_{t-1,s-1}) - 2f(u_{t-1})\right] g(u_{t-1})\ I_{t-1}$.

Obviously, there is nothing new in this two terms with respect to what we have said in (A.50) and (A.51) respectively. Thus, these two terms (A.52) and (A.53) are both $o_p(1)$.

With this, we conclude that the term (A.39) is $o_p(1)$. Therefore, in order to prove the proposition, it only remains to show that the term (A.40) is $o_p(1)$.

- As far as (A.40) is concerned, i.e.:

$$(A.40)\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\varepsilon_t\Bigg[\left[\frac{\frac{1}{Ta}\sum\limits_{s=2}\rho(u_{s-1})\bar{K}_{t-1,s-1}}{\hat{f}(u_{t-1})}\cdot\frac{\hat{f}'(u_{t-1})}{\hat{f}(u_{t-1})}\right]-$$

$$\frac{\rho(u_{t-1})}{f(u_{t-1})}f'(u_{t-1})\Bigg]I_{t-1}$$

can be written as:

$$(A.54)\quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\varepsilon_t\frac{f'(u_{t-1})\ (\hat{\rho}(u_{t-1})-\rho(u_{t-1}))}{f(u_{t-1})}I_{t-1}$$

$$(A.55)\ +\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\varepsilon_t(\hat{\rho}(u_{t-1})-\rho(u_{t-1}))\Bigg[\frac{(f'(u_{t-1})-\hat{f}'(u_{t-1}))}{f(u_{t-1})}+$$

$$\frac{(\hat{f}(u_{t-1})-f(u_{t-1}))^2}{\hat{f}(u_{t-1})f^2(u_{t-1})}+\frac{\hat{f}(u_{t-1})-f(u_{t-1})}{f(u_{t-1})}+$$

$$\frac{(\hat{f}'(u_{t-1})-f'(u_{t-1}))(\hat{f}(u_{t-1})-f(u_{t-1}))}{f(u_{t-1})}\Bigg]I_{t-1}$$

$$(A.56)\ +\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\varepsilon_t\rho(u_{t-1})\Bigg[\frac{(f'(u_{t-1})-\hat{f}'(u_{t-1}))}{f(u_{t-1})}+$$

$$\frac{(\hat{f}(u_{t-1})-f(u_{t-1}))^2}{\hat{f}(u_{t-1})f^2(u_{t-1})}+\frac{\hat{f}(u_{t-1})-f(u_{t-1})}{f(u_{t-1})}+$$

$$\frac{(\hat{f}'(u_{t-1})-f'(u_{t-1}))(\hat{f}(u_{t-1})-f(u_{t-1}))}{f(u_{t-1})}\Bigg]I_{t-1}.$$

Each term from (A.54)-(A,56) is $o_p(1)$ since their terms are nothing different from what we have done in proposition 1 or in proposition 2 up to now.

Q.E.D.

Proposition 3

Let u_t be as in proposition 2. Let also assume the same conditions as in proposition 2. Then:

$$\text{(A.57)} \quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} x_{t-1}(\hat{\rho}(u_{t-1})-\rho(u_{t-1}))\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}} I_{t-1}=o_p(1).$$

Proof:

The L.H.S. of equation (A.57) is equal to:

$$\text{(A.58)} \quad \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} x_{t-1}(\hat{\rho}(u_{t-1})-\rho(u_{t-1}))[\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}} - \rho'(u_{t-1})]I_{t-1}$$

$$\text{(A.59)} \quad + \frac{1}{T^{\frac{1}{2}}}\sum_{t=2} x_{t-1}(\hat{\rho}(u_{t-1})-\rho(u_{t-1})) \; \rho'(u_{t-1})I_{t-1}.$$

The most easy term to show that it is $o_p(1)$ is (A.59), since in fact it is nothing different than what we have done in proposition 1. The only difference is that we have x_{t-1} in (A.58) and (A.59), but it does not make any difference at all since it is independent of u_s with finite second moments.

- About (A.58) is evident that it is $o_p(1)$ since the only thing to do is to split it up in

$$\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}} - \rho(u_{t-1}) \quad \text{and} \quad \rho(u_{t-1})-\hat{\rho}(u_{t-1})$$

as we have done in proposition 1 and 2 and to take the supremums as we did with the terms (A.46) and (A.47) for instance. (See remark at the end of proposition 4).

Proposition 4

Let u_t be a stationary stochastic process as in proposition 1 or 2. Let also assume that A1, A2 and A3 are satisfied. Then:

$$\text{(A.60)} \quad \frac{1}{T}\sum_{t=2}\Big[(x_t-x_{t-1}\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}I_{t-1})(x_t-x_{t-1}\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}I_{t-1})'-$$

$$(x_t-x_{t-1}\rho'(u_{t-1}))(x_t-x_{t-1}\rho'(u_{t-1}))'\Big]=o_p(1).$$

Proof:

The L.H.S. of (A.60) is equal to:

$$\frac{1}{T}\sum_{t=2}[x_{t-1}x_t'+x_tx_{t-1}'](\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}I_{t-1}-\rho'(u_{t-1}))]$$

$$+\frac{1}{T}\sum_{t=2}x_{t-1}x_{t-1}'(\rho'(u_{t-1})^2-\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}^2I_{t-1})$$

$$-\frac{1}{T}\sum_{t=2}[x_{t-1}x_t'+x_tx_{t-1}'](\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}-\rho'(u_{t-1}))]\,I_{t-1}$$

$$+\frac{1}{T}\sum_{t=2}[x_{t-1}x_t'+x_tx_{t-1}']\rho'(u_{t-1})\{1-I_{t-1}\}$$

$$+\frac{1}{T}\sum_{t=2}x_{t-1}x_{t-1}'(\rho'(u_{t-1})^2-\frac{\partial\hat{\rho}(u_{t-1})^2}{\partial u_{t-1}})I_{t-1}$$

$$+\frac{1}{T}\sum_{t=2}x_{t-1}x'_{t-1}\rho'(u_{t-1})^2\{1-I_{t-1}\}.$$

Now the second and fourth terms involve only an application of lemma 5, and the first and third ones are $o_p(1)$ by lemma 6. Although we do not have here ε_t, there is nothing to worry about since we have $1/T$ instead of $1/T^{\frac{1}{2}}$. Thus, only we are going to need that the difference to be $o_p(1)$, i.e. its consistency.

__Remark__:

Take into account that although we have the x_t's inside the sums, it does not cause any problem in order to apply propositions 1 and 2, since by assumption A3, the process x_t is independent of u_t and by assumption A2 it has finite second moments.

__Theorem 1__

Under the same conditions of proposition 4, we have that:

$$T^{\frac{1}{2}}(\beta^*-\hat{\beta})=o_p(1).$$

__Proof__:

From (5.1) and (5.2) as we have said in section 5 of the chapter we only need to show that (5.3) and (5.5)-(5.7) are $o_p(1)$.

In order to prove the theorem, we are going to use a device by Härdle and Stoker (1987) of proving that (5.3) and (5.5)-(5.7) are $o_p(1)$ for an uncomputable estimator, i.e. when the trimming $\bar{I}_{t-1}$ is substituted by I_{t-1}, i.e.

$$\bar{I}_{t-1}=\begin{cases} 1 & \text{if } |\hat{f}(u_{t-1})|>b \\ 0 & \text{otherwise} \end{cases}$$

is substituted by

$$I_{t-1}=\begin{cases} 1 & \text{if } f(u_{t-1})>b \\ 0 & \text{otherwise} \end{cases}$$

and after that, to prove that in fact it does not make any difference at all asymptotically between the computable (when we use $\bar{I}_{t-1}$) and the uncomputable one.

Now from proposition 1 through proposition 4 we have seen that (recall the remark at the end of proposition 4):

1) $$\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}(\rho(u_{t-1})-\hat{\rho}(u_{t-1})I_{t-1})(x_t-x_{t-1}\rho'(u_{t-1}))$$

2) $$\frac{1}{T^{\frac{1}{2}}}\sum_{t=2}\epsilon_t\, x_{t-1}\left[\frac{\partial}{\partial u_{t-1}}\,\hat{\rho}(u_{t-1})I_{t-1} - \frac{\partial}{\partial u_{t-1}}\rho(u_{t-1})\right]$$

3) $$\frac{1}{T^{\frac{1}{2}}}\sum_{t=2} x_{t-1}(\hat{\rho}(u_{t-1})-\rho(u_{t-1}))\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}}\, I_{t-1}$$

4) $$\frac{1}{T}\sum_{t=2} (x_t - x_{t-1}\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}} I_{t-1}) \times (x_t - x_{t-1}\frac{\partial\hat{\rho}(u_{t-1})}{\partial u_{t-1}} I_{t-1})'$$

$$-(x_t - x_{t-1}\rho'(u_{t-1})) \times (x_t - x_{t-1}\rho'(u_{t-1}))'$$

are all $o_p(1)$, but also this is still true if b, the trimming parameter, is substituted by:

$$b - c_\sigma(T^{-\frac{1}{2}}a^{-1} + a^{2+\lambda}) = b - c_T.$$

Recall that this $b-c_T$ still $O(b)$, since $b^{-1}c_T \to 0$, so for this trimming parameter $b-c_T$ 1),2),3) and 4) are still $o_p(1)$.

But by construction:

$$P\{ \left|\frac{1}{T^{\frac{1}{2}}}\sum x_t(\rho(u_{t-1}) - \hat{\rho}(u_{t-1})\bar{I}_{t-1})\right| > \epsilon,\ |\hat{f}(u_{t-1}) - f(u_{t-1})| < c_T,\ \forall t\} <$$

$$P\{ \left|\frac{1}{T^{\frac{1}{2}}}\sum x_t(\rho(u_{t-1}) - \hat{\rho}(u_{t-1}) I_{t-1})\right| > \epsilon,\ |\hat{f}(u_{t-1}) - f(u_{t-1})| < c_T,\ \forall t\}.$$

But, on the other hand, we have that:

$$P\{ \sup_t |\hat{f}(u_{t-1}) - f(u_{t-1})| > c_T\}$$

goes to zero as T goes to infinity, and consequently we can conclude that:

$$P\{ \left|\frac{1}{T^{\frac{1}{2}}}\sum x_t(\rho(u_{t-1}) - \hat{\rho}(u_{t-1})\bar{I}_{t-1})\right| > \epsilon,\ |\hat{f}(u_{t-1}) - f(u_{t-1})| < c_T, \forall t\}$$

goes to zero and so:

$$P\{ \left|\frac{1}{T^{\frac{1}{2}}}\sum x_t(\rho(u_{t-1}) - \hat{\rho}(u_{t-1})\bar{I}_{t-1})\right| > \epsilon \}$$ goes to zero.

Therefore 1) is also true when we substitute I_{t-1} by $\overline{I}_{t-1}$. With 2), 3) and 4) we can argue in the same way to conclude that:

$$T^{\frac{1}{2}}(\beta^{*}-\hat{\beta})=o_p(1).$$

and theorem 1 of section 5 is proved.

APPENDIX B

In this appendix we are going to give the proof of theorem 2. Since, as we have shown elsewhere (see section 3 of this chapter), that the estimation of the parameters β is asymptotically independent of the parameter(s) θ of the ρ-function, we are going to prove the theorem as if these parameters were known to the practitioner as well.

Before proving the Central Limit Theorem (C.L.T.) of the β estimates, we are going to prove a previous lemma which turns out to be a corollary of a theorem by Brown (1971).

Lemma 1

Let v_t be a sequence of martingale difference variables such that they have $2+\delta$ finite moments ($\delta>0$). Then we have that:

$$(B.1) \qquad \frac{1}{T^{\frac{1}{2}}} \sum v_t$$

has the Central Limit Property, i.e. its limit distribution is a $N(0,\sigma)$, where $\sigma^2 = \lim\limits_{T} {}^1/_T \sum E v^2_t$.

Proof:

We only need to prove that

$$\text{i) plim } \frac{1}{T} \sum_t Ev_t^2|\text{past}=\sigma^2$$

and

$$\text{ii) lim } Ev_1^2 I(|v_1|>\epsilon T\sigma^2)=0.$$

Let us show first i).

Due to the fact that we do not assume existence of the fourth moments of v_t we do the usual trick of truncating the v_t's.

Calling $\bar{v}_t=v_t I(|v_t|<a)$

and

$$\tilde{v}_t=v_t I(|v_t|>a).$$

To deal with Ev_t^2 or $E\bar{v}_t^2$ no really matters for our purpose since the difference is equal to

$$\text{(B.2)} \quad \int_{|v_t|>a} v_t^2 dF_t <$$

$$\frac{1}{a^\delta} \int_{|v_t|>a} |v_t|^{2+\delta} dF_t <$$

$$\frac{1}{a^\delta} \int |v_t|^{2+\delta} dF_t <$$

$$\frac{1}{a^{\delta}} M$$

and for sufficiently large a this is less than or equal to some ϵ.

On the other hand:

(B.3) $$E\{\frac{1}{T}\sum_t E_{t-1}\bar{v}_t^2 - \frac{1}{T}\sum_t \bar{v}_t^2\}^2 = \frac{1}{T^2}\sum E(E_{t-1}\bar{v}_t^2 - \bar{v}_t^2)^2$$

$$= \frac{1}{T} E(E_{t-1}\bar{v}_t^2 - \bar{v}_t^2)^2$$

$$< \frac{a^4}{T}$$

which goes to zero if $a = o(T^{\frac{1}{4}})$.

By (B.2) and (B.3)

$$\text{plim}\left[\frac{1}{T}\sum \bar{v}_t^2 - \frac{1}{T}\sum E_{t-1}v_t^2\right] \text{ is equal to zero.}$$

So, if we prove that:

(B.4) $$\text{plim}\ \frac{1}{T}\sum \bar{v}_t^2 = \sigma^2$$ then i) will be proved.

But, on the other hand, we have that the L.H.S. of (B.4) is equal to:

$$\text{plim}\ \frac{1}{T}\sum v_t^2 - \text{plim}\ \frac{1}{T}\sum \tilde{v}_t^2$$

and by Chebyshev's theorem and (B.2) we have that the last term of this difference goes to zero. Hence, part i) is proved.

Part ii) is obvious since v_t^2 is uniformly integrable. Q.E.D.

Lemma 2

Assume B1, B2, B3 and A2. Then:

$$\frac{1}{T}\sum_t \frac{\partial^2\rho(u_{t-1})}{\partial u^2_{t-1}} x_{t-1}x'_{t-1}\varepsilon_t \to 0 \text{ in probability.}$$

Proof:

It follows immediately by noting that ε_t is an i.i.d. process and that:

$$\frac{\partial^2\rho(u_{t-1})}{\partial u_{t-1}{}^2} x_{t-1}x'_{t-1}$$

which by assumption B3 has finite $2+\delta$ moments, where $\delta>0$.

This lemma will allow us to say that:

$$\text{(B.5)} \qquad \frac{1}{T}\sum_t \Big(x_t - x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\Big)\Big(x_t - x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\Big)'$$

and

$$\frac{1}{T}\sum_t \Big(x_t - x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\Big)\Big(x_t - x_{t-1}\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}}\Big)' +$$

$$\Big(\frac{\partial^2\rho(u_{t-1})}{\partial u^2_{t-1}} x_{t-1}x'_{t-1}\varepsilon_t\Big)$$

which is the second derivative w.r.t. β of

$$\frac{1}{T}\sum_t (u_t - \rho(u_{t-1}))^2$$

have the same probability limit, and which by assumption B2 is finite and definite positive.

Lemma 3

$$\text{Let } \frac{1}{T^{\frac{1}{2}}} \sum_t (x_t - x_{t-1} \frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})\varepsilon_t = \frac{1}{T^{\frac{1}{2}}} \sum_t v_t$$

then under B1, B2, B3 and A2

$$\frac{1}{T} \sum_t v_t \rightarrow N(0, \sigma^2 C)$$

where 0<C is the probability limit of (B.5).

Proof:

$E(v_t | v_{t-1}, \ldots, v_1) = 0$ by B2 and therefore v_t is a martingale difference. Now by B2 and A2 $E|v_t|^{2+\delta} < \infty$ uniformly, so we apply lemma 1 to obtain the required result.

Q.E.D.

Therefore by lemma 3 and B2 we have that

$$\text{(B.6)} \quad (\frac{1}{T} \sum_t (x_t - x_{t-1} \frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})(x_t - x_{t-1} \frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})')^{-1}$$

$$\times \frac{1}{T^{\frac{1}{2}}} \sum_t \varepsilon_t (x_t - x_{t-1} \frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})$$

converges to a $N(0, \sigma^*)$ where σ^* is given by $C^{-1} E\varepsilon_t^2$ because the ε_t are i.i.d., and where the matrix C is given by (B.5)

In our Gauss-Newton framework we have:

$$\hat{\beta}=\beta_T+\left(\sum_t (x_t - x_{t-1}\,\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})(x_t - x_{t-1}\,\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})'\right)^{-1}$$

$$\sum_t \varepsilon_t (x_t - x_{t-1}\,\frac{\partial\rho(u_{t-1})}{\partial u_{t-1}})\Big|_{\beta_T^{\cdot}}$$

Now if β_T is consistent then the limit distribution of $\sqrt{T}(\hat{\beta}-\beta)$ has the same distribution as in (B.6) because in general if we have that $\nu_T \longrightarrow \nu_0$ in probability and $|Q_T(\nu)-Q(\nu)| \to 0$ uniformly in ν in probability then $Q_T(\nu_T) \longrightarrow Q(\nu_0)$ also, by lemma 4 of Amemiya (1973).

Therefore Theorem 2 of section 5 has been proved.

APPENDIX C

Here, we are going to prove that the measures generated by $\{\bar{u}_t\}$ and $\{u_t\}$ are contiguous. To do that, we have to check all the assumptions listed in Roussas (1972) chapter 2, i.e.:

$$\text{define } q(u_1;\beta,\beta^*)=\left[\frac{dP_1(u_1,\beta)}{dP_1(u_1,\beta^*)}\right]$$

and

$$q(u_2,u_1;\beta,\beta^*)=\left[\frac{dP_2(u_1,u_2,\beta)}{dP_2(u_1,u_2,\beta^*)}\right].$$

(A.1) For each $\beta\epsilon\Theta_1$, the Markov process $\{u_t\}$, $t\geq 1$ is stationary and ergodic.

(A.2) The probability measures $\{P_{t,\beta},\beta\epsilon\Theta_1\}$ are mutually absolutely continuous for all $t\geq 1$.

(A.3) Let $\varphi_1(\beta,\beta^*)=\left[\frac{q(u_2,u_1;\beta,\beta^*)}{q(u_1;\beta,\beta^*)}\right]^{\frac{1}{2}}$ be , then

(i) For each $\beta\epsilon\Theta_1$, the random function $\varphi_1(\beta,\beta^*)$ is differentiable in quadratic mean (q.m.) $[P_\beta]$ with respect to β^* at

$\beta^*\!\to\!\beta$.

Let $\dot{\varphi}_1(\beta)$ be the derivative in q.m of $\varphi_1(\beta,\beta^*)$ w.r.t. β^* at β. Then:

(ii) $\dot{\varphi}_1(\beta)$ is a measurable function.

Let $\Gamma(\beta)=4E_\beta[\dot{\varphi}(\beta)\dot{\varphi}(\beta)']$. Then:

(iii) $\Gamma(\beta)$ is a positive definite matrix for all $\beta\epsilon\Theta_1$.

(A.4) (i) For each $\beta\epsilon\Theta_1$, $q(u_1,u_2;\beta,\beta^*)\longrightarrow 1$ in $P_{1,\beta}$ as $\beta^*\longrightarrow\beta$.

(ii) $\forall\beta\epsilon\Theta_1$, $q(u_1;\beta,\beta^*)$ and $q(u_1,u_2;\beta,\beta^*)$ are measurable.

In the model what we have to check is (A.3), because of (A.1) is satisfied by assumption and (A.2) is satisfied providing that the probability density function of the disturbances ϵ_t is different than zero. About (A.4), if (A.3) is true then the conditional probability distribution of the residuals under β^* converges to the one under β and the probability density function under β^* converges in probability to the probability measure under β. Thus, the only problem is to check condition (A.3).

<u>Lemma 1</u>

Assume that the random variable ϵ_t in (3.1) has a continuously differentiable probability density function g such that:

(i) $g(x)\neq 0 \;\forall x\epsilon R$

(ii) $\int_R \left[\frac{g'(x)}{g(x)}\right]^2 g(x)dx<\infty$

and

(iii) $\int_R x^2 g(x)dx<\infty$

and the common distribution of the u_t's admits the following decomposition:

$$f_1(u_1,\ldots,u_T)=f(u_1)\prod_{t=2}^{T} g(u_t-\rho(u_{t-1}))$$

where u_t, $t=1,2,\ldots,T$, are the residuals obtained from $y_t-\beta' x_t$, $t=1,2,\ldots,T$,. Then the condition (A.3) is satisfied.

<u>Proof</u>:

Let $u_t^*=y_t-\beta^{*\prime}x_t$ be. It can be seen that:

$$\varphi_1(\beta,\beta^*)=\left[\frac{g(u_2^*-\rho(u_1^*))}{g(u_2-\rho(u_1))}\right]^{\frac{1}{2}}.$$

Then, the derivative of this function with respect to β^* evaluated at β is equal to:

$$\frac{g'(u_2-\rho(u_1))}{2g(u_2-\rho(u_1))}(x_2-x_1\dot{\rho}(u_1)).$$

Hence, we have that:

$$E\left[\frac{\partial^2}{\partial\beta^*\partial\beta^{*\prime}}\Big|\beta^*=\beta\right] = \int \frac{g'(u_2-\rho(u_1))^2}{4g(u_2-\rho(u_1))}\tilde{x}\tilde{x}' f(u_1)du_1du_2,$$

where $\tilde{x}=(x_1-x_2\dot{\rho}(u_1))$.

Therefore, what we have to prove is that:

$$E_\beta\Big[\frac{1}{v}\Big[\frac{g^{\frac{1}{2}}(\varepsilon_2-vh'x_2-(\rho(u_1-vh'x_1)-\rho(u_1)))}{g^{\frac{1}{2}}(u_2-\rho(u_1))}-1\Big]-\frac{g'(\varepsilon_2)}{2g(\varepsilon_2)}h'\{x_2-x_1\dot{\rho}(u_1)\}\Big]^2 \xrightarrow[v\to 0]{} 0.$$

But the L.H.S. of the above equation is less than or equal to 2 times

$$E_\beta\Big[\frac{1}{v}\Big[\frac{g^{\frac{1}{2}}(\varepsilon_2-vh'x_2-(\rho(u_1-vh'x_1)-\rho(u_1)))-g^{\frac{1}{2}}(\varepsilon_2-vh'x_2-vh'x_1\dot{\rho}(u_1))}{g^{\frac{1}{2}}(u_2-\rho(u_1))}\Big]^2+$$

$$E_\beta\Big[\frac{1}{v}\Big[\frac{g^{\frac{1}{2}}(\varepsilon_2-v(h'x_2-h'x_1\dot{\rho}(u_1)))}{g^{\frac{1}{2}}(u_2-\rho(u_1))}-1\Big]-\frac{g'(\varepsilon_2)}{2g(\varepsilon_2)}h'\{x_2-x_1\dot{\rho}(u_1)\}\Big]^2.$$

Thus, if each of this two terms goes to zero then we have proved part (i) of (A.3). The proof of the second term is identical to theorem 3.3. of Kreiss (1987) but instead of having $\lambda h^T Z(j-1,\beta,\beta^*)$ we have $v(h'x_2-h'x_1\dot{\rho}(u_1))$.

About the first term, this is equal (applying the mean value theorem and $\dot{\varepsilon}_2$ standing for a point lying between $\varepsilon_2-vh'x_2-(\rho(u_1-vh'x_1)-\rho(u_1))$ and $\varepsilon_2-vh'x_2-vh'x_1\dot{\rho}(u_1))$ to:

$$\frac{1}{4}\int\Big[\frac{g'(\dot{\varepsilon}_2)}{g(\dot{\varepsilon}_2)}\Big]^2\big[\{h'x_1\dot{\rho}(u_1)\}-\frac{1}{v}\,(\rho(u_1-h'vx_1)-\rho(u_1))\}\big]^2\ g(\varepsilon_2)df(u_1)du_1d\varepsilon_2.$$

Now then:

$$\int\frac{1}{4}\ \Big[\frac{g'(\dot{\varepsilon}_2)}{g(\dot{\varepsilon}_2)}\Big]^2 g(\dot{\varepsilon}_2)d\varepsilon_2 \quad \text{converges to} \int\frac{1}{4}\ \Big[\frac{g'(\varepsilon_2)}{g(\varepsilon_2)}\Big]^2 g(\varepsilon_2)d\varepsilon_2<\infty$$

since g and its derivative are continuous and also:

$$\lim_{v\to 0} \{ h'x_1\dot{\rho}(u_1)-1/_v(\rho(u_1-vh'x_1)-\rho(u_1))=0.$$

Therefore we can conclude that (A.3) (i) is satisfied by applying the Lebesgue dominated convergence theorem since we are able to introduce the **lim** inside the integrand sign.

Also $\Gamma(\beta)>0$ since by hypothesis the information matrix is less than infinity (see assumption B2), and the measurability is also true since the limit of measurable functions is a measurable function as well.

Q.E.D.

CHAPTER 4

SOME FINITE SAMPLE EVIDENCE ABOUT THE ADAPTIVE ESTIMATOR OF CHAPTER 3

In this chapter, we provide some finite sample evidence about the semiparametric estimator which we have studied in chapter 3.

We simulated a wide variety of models. We report only few of them, since the findings on the other models are of a similar kind. Throughout the whole set of simulations the regression model has been:

$$(1.1) \qquad y_t=\alpha+\beta x_t+u_t$$

where $x_t=0.9x_{t-1}+v_t$, and v_t are independent and identically distributed zero mean normal with variance equal to 4. We have chosen $\alpha=1$ and $\beta=2$.

We consider seven different models for the autoregression function of the residuals u_t. These are as follows:

Model 1. $u_t=\dfrac{u_{t-1}}{1+u_{t-1}^2}+\varepsilon_t$,

Model 2. $u_t=u_{t-1}(.25+5\exp(-u_{t-1}{}^2))+\varepsilon_t$,

Model 3. $u_t = u_{t-1}(.25-5\exp(-^1/_9 u_{t-1}{}^2))+\epsilon_t$,

Model 4. $u_t = u_{t-1}\exp(-^1/_9 u_{t-1}{}^2))+\epsilon_t$,

Model 5. $u_t = 5\sin u_{t-1}+\epsilon_t$,

Model 6. $u_t = 5(\sin u_{t-1}+0.5\sin(2u_{t-1}))+\epsilon_t$,

Model 7. $u_t = .5u_{t-1}+\epsilon_t$

where the ϵ_t's are independent zero mean normals with a variance of 4.

All of these seven models considered above satisfy the assumptions of theorem 1 of chapter 3, the view associated with theorems 3.2.2 and 3.2.3 of chapter 2 and by Pham and Tram's (1985) result for model 7.

By parametric estimator of α and β we will understand the least squares estimator of these parameters when the autoregression function of u_t is perfectly known, i.e. the functional form and the parameters that the autoregression function may have. On the other hand, by MIS we will understand the estimator of α and β that we obtain when we estimate the regression model in the belief that the autoregression function of the residuals follows a linear AR(1) model.

In section 4.1, we will see the performance, in terms of the Mean Square Error (M.S.E.) of the parametric, semiparametric and the MIS estimators of the model studied in chapter 3. In section 4.2, we examine the power of a Wald test for the parameters of that model and for the different estimators, e. g. parametric, semiparametric and MIS.

4.1.- NUMERICAL RESULTS IN FINITE SAMPLES OF THE ADAPTIVE ESTIMATOR

In this section we provide some finite sample evidence about the performance of the semiparametric estimator of α and β, and give some evidence through a Monte-Carlo experiment of the open question posed in the introduction of chapter 3. This turns out to be very important in view of the inability to give, in mathematical terms, a proof for the inequality (1.4) > (1.5) or the inequality given in equation (1.7) of that chapter.

All the computations were carried out in double-precision FORTRAN on the University of London's Amdahl computer, using a random generator from the NAG library and the Marquardt subroutine taken from Press et al. (1986) to minimize the non-linear least squares. For each of the seven models we generated 1000 bivariates time series ε_t and x_t of lengths T=100 and 200.

In tables 4.1-4.7 below, we only report the M.S.E. for α and β since the bias is much the same for the parametric, MIS, and the semiparametric estimator, and its contribution to the M.S.E. is negligible.

TABLE 4.1

M.S.E. for model 1 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0570	.0029		.0270	.0013
MIS	.0650	.0035		.0316	.0016
a=.65	.0670	.0034	a=.35	.0340	.0017
a=.95	.0640	.0033	a=.65	.0310	.0015
a=1.25	.0640	.0034	a=.95	.0310	.0015

TABLE 4.2

M.S.E. for model 2 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0180	.0013		.0058	.0003
MIS	.1600	.0071		.0750	.0031
a=.35	.0900	.0031	a=.35	.0300	.0006
a=.65	.0790	.0027	a=.65	.0260	.0007
a=.95	.0850	.0033	a=.95	.0320	.0010

TABLE 4.3

M.S.E. for model 3 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0082	.0004		.0034	.0002
MIS	.1162	.0066		.0522	.0027
a=.65	.0154	.0005	a=.35	.0120	.0002
a=.95	.0157	.0006	a=.65	.0100	.0002
a=1.25	.0180	.0007	a=.95	.0100	.0002

TABLE 4.4

M.S.E. for model 4 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0750	.0035		.0330	.0014
MIS	.1200	.0055		.0570	.0025
a=.65	.0950	.0045	a=.35	.0480	.0022
a=.95	.0970	.0042	a=.65	.0430	.0018
a=1.25	.0990	.0044	a=.95	.0570	.0018

TABLE 4.5

M.S.E. for model 5 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0037	.0002		.0017	.00009
MIS	.1500	.0096		.0689	.0040
a=.35	.0175	.0006	a=.35	.0110	.0001
a=.65	.0147	.0005	a=.65	.0100	.0001
a=.95	.0194	.0007	a=.95	.0122	.0002

TABLE 4.6

M.S.E. for model 6 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.0022	.0003		.0008	.00005
MIS	.2700	.0130		.1200	.0058
a=.35	.0750	.0015	a=.35	.0170	.00008
a=.65	.0590	.0007	a=.65	.0160	.0001
a=.95	.0730	.0014	a=.95	.0250	.0005

TABLE 4.7

M.S.E. for model 7 with sample sizes of T=100 and T=200, and 1000 replicates.

	T=100			T=200	
	α	β		α	β
Parametric	.1800	.0070		.0900	.0031
MIS	-----	-----		-----	-----
a=.65	.2100	.0083	a=.35	.1000	.0041
a=.95	.1900	.0075	a=.65	.1000	.0036
a=1.25	.1800	.0073	a=.95	.0920	.0034

From the above tables, we can distinguish three different kinds of results, as far as the M.S.E. is concerned. For model 1, we observe that there does not seem to be a loss of efficiency when we estimate the parameters of equation (1.1) in the belief that the residuals u_t follow a linear AR(1) instead of the true non-linear one. For model 4, we observe that the ratio between the M.S.E.'s of the MIS and the parametric estimator is near 2, and this ratio gets better as the sample size increases. And finally, for models 2, 3, 5 and 6, we observe that this ratio is quite enormous, for example in model 6 it is as big as 116, and once again this ratio increases with the sample size.

This difference in the results may be explained by the severity of the non-linearity in the autoregression function ρ. In all of these models, the autoregression function has one singular point at the mean of the process, that is at zero. But the characteristic which differentiates models 1 and 4 from models 2, 3, 5 and 6 is that the latter models have more than one singular point, as it is the case for model 2, or that the autoregression function induces a cyclical pattern in the residuals u_t, as it is the case for models 3, 5 and 6. These last four models are such that the non-linearity is likely to be more "severe", and thus its approximation by a linear model may be very poor, a fact reflected in the M.S.E. of the estimators of α and β for the MIS and parametric estimators.

Although models 1 and 4 have only one singular point and they do not have a cyclical behaviour, their difference is that for model 4 the non-linearity may be a little more severe than for model 1, especially for small values of u_t, where the distribution of the u_t is more concentrated. This may explain why the results for model 4 are better, as far as the behaviour of the parametric estimator is concerned.

As a whole, the results are very encouraging, showing that there is going to be a loss of efficiency for the estimators of α and β if we estimate the model in the manner of a linear AR(1) model, and this loss of efficiency depends on the severity of the non-linearity of the autoregression function.

As far as the bandwidth parameter **a** is concerned, we observe that the **a**, for which the semiparametric estimate proposed in the previous chapter performs better in terms of the M.S.E., is quite stable throughout the models. Also, it seems that the range of "best" **a**'s is quite wide. This "best" bandwidth parameter **a** seems to be between 0.65 and 0.95, but with the important characteristic that the difference in the M.S.E. when the bandwidth parameter **a** is 0.65 or 0.95 is quite negligible, except in model 6, although for both a=0.65 and a=0.95 the performance of the semiparametric estimator is much better than the MIS one. For the sample size of 200, the stability seems greater, and although the bandwidth parameter is between .35 and .65, their difference in performance is very small. It is worthwhile pointing out that we are aware that this parameter **a** may be influenced by the variance of the process u_t, although this influence may be difficult to determine in view of the difficulty to obtain a close expression for the variance of a non-linear AR model. But, we can point out that as a rule of thumb, we might choose **a** to be a value around $T^{-0.1}$ which is perfectly legitimate in view of the assumptions of theorem 1 of chapter 3. All this discussion excludes model 7, but even in this model, the M.S.E. of the semiparametric estimator is near identical for all the bandwidths, and very close to the parametric one.

As it was pointed out earlier, we simulated other models to see especially the performance of the parametric versus the MIS estimator of α and β, in view of the inability to show the inequality of (1.4)$<$(1.5) or the inequality given by equation (1.7) in chapter 3. An important fact that we found, as we might expect, is that the gap between the MSE's of the parametric and MIS decreases when the autocorrelation goes to 1. This fact is expected since as the autocorrelation goes to 1, u_t may be written as a linear combination of u_{t-1}, i.e. the model becomes linear. We have also studied the performance of the parametric and MIS estimators with a sample size of 1000. Although this sample size is unrealistic for most time series data sets available, we have done it to see their performance. We observed two facts. First, for the estimated asymptotic covariance matrix given by the M.S.E. times the sample size is "identical" for T=200 and T=1000. And secondly, the performance of both estimators are more similar than for samples sizes of T=100 or T=200. These last two points may corroborate the conjecture that the inequalities given in equation (1.7) or (1.4) > (1.5) are in fact true.

4.2.- THE BEHAVIOUR IN FINITE SAMPLES OF THE ADAPTIVE ESTIMATOR IN HYPOTHESIS TESTING

In this section we explore finite sample evidence of test statistics for the parameters of the regression model using the parametric, semiparametric and MIS estimators. The reason for this study is the same as in the previous section. If the inequalities (1.7) or (1.4) > (1.5) were true, as intuition tells us, then although we can still form consistent tests under the MIS model, these tests would have less power than the ones under the true specification.

Thus, as it was already argued in the previous section, it would be interesting to observe the finite sample behaviour of test statistics of the model for the different estimators.

We have used two sample sizes of lengths T=100 and 200, and replicated the model 1000 times for both sample sizes. We have used the same seven models as in the previous section. All the computations were carried out on the University of London's Amdahl computer.

We have to mention that we have changed the regression model slightly. As we have seen in the tables of section 1, the estimated variances for the different models were rather small. Thus, this will imply that unless the difference between the true parameter and the hypothesized one is rather small, the power of the test-statistics for the MIS, the parametric and the semiparametric estimators will be near 1. There are two options whereby the variance of the estimator could be increased. The first one is to increase the variance of the innovations ε_t. The second one is to decrease the variance of the regressors. We have chosen the second alternative, and we have chosen for the x_t the following model:

$$x_t = .9x_{t-1} + v_t$$

where the v_t's are independent zero mean normal with a variance of 0.16.

We also have to point out that we have calculated the power of the test-statistics in a different way, although it makes no difference at all, as far as the conclusions that we can draw from it are concerned. Normally simulation studies of test-statistics are carried out by varying the α's and β's of the model and keeping the null hypothesis constant. By contrast, we have studied the power of the tests by keeping the parameters of the model fixed and varying the null hypothesis. Our test-statistics are given by:

$$\hat{\pi}=T(\hat{\alpha}-\alpha)V(1,1)^{-1}(\hat{\alpha}-\alpha)\xrightarrow{d}\chi^2_1$$

and

$$\hat{\pi}=T(\hat{\beta}-\beta)V(2,2)^{-1}(\hat{\beta}-\beta)\xrightarrow{d}\chi^2_1$$

where $V(i,i)$ $i=1$,and 2 stands for an estimator of the asymptotic variance of $\sqrt{T}(\hat{\alpha}-\alpha)$ and $\sqrt{T}(\hat{\beta}-\beta)$ respectively.

Tables 4.8-4.14 show the proportion $\hat{\pi}$ of rejections in 1% and 5% tests for the different estimators. The numbers .35 or .65 or .95 or 1.25 stand for the chosen bandwidth parameter a of the semiparametric estimator.

TABLE 4.8

$\hat{\pi}$'s for model 1, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.35	a=.65	G.L.S.	MIS	a=.65	a=.95
1%	α=1.00	.020	.033	.029	.018	.023	.023	.022	.015
	α=1.25	.092	.095	.102	.072	.180	.167	.174	.149
	α=1.50	.399	.370	.364	.316	.739	.684	.694	.678
	α=2.00	.968	.933	.928	.916	1.00	1.00	.998	.998
	β=2.00	.027	.023	.023	.021	.021	.015	.020	.018
	β=2.25	.112	.062	.091	.064	.181	.128	.167	.145
	β=2.50	.373	.271	.353	.287	.678	.593	.642	.622
	β=3.00	.908	.822	.861	.843	.995	.995	.992	.992
5%	α=1.00	.082	.099	.095	.067	.072	.070	.073	.063
	α=1.25	.228	.231	.215	.185	.380	.339	.360	.336
	α=1.50	.621	.564	.591	.543	.882	.833	.854	.839
	α=2.00	.994	.981	.975	.967	1.00	1.00	1.00	1.00
	β=2.00	.075	.064	.096	.064	.073	.067	.077	.064
	β=2.25	.233	.184	.218	.195	.352	.294	.325	.316
	β=2.50	.591	.478	.543	.506	.842	.770	.805	.795
	β=3.00	.971	.926	.938	.945	1.00	.999	.999	.999

TABLE 4.9

$\hat{\pi}$'s for model 2, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.35	a=.65	G.L.S.	MIS	a=.35	a=.65
1%	α=1.00	.044	.048	.103	.035	.018	.026	.037	.009
	α=1.25	.459	.067	.343	.088	.833	.091	.600	.205
	α=1.50	.978	.182	.793	.456	.999	.316	.974	.867
	α=2.00	.999	.631	.963	.939	1.00	.906	.998	.999
	β=2.00	.060	.019	.092	.021	.031	.014	.028	.006
	β=2.25	.401	.040	.310	.078	.711	.080	.494	.168
	β=2.50	.905	.146	.721	.393	.996	.281	.965	.806
	β=3.00	.998	.542	.968	.914	1.00	.880	.998	.999
5%	α=1.00	.106	.110	.168	.063	.061	.086	.082	.027
	α=1.25	.698	.164	.536	.275	.943	.199	.772	.425
	α=1.50	.994	.342	.886	.710	.999	.519	.983	.970
	α=2.00	.999	.804	.976	.964	1.00	.976	.998	1.00
	β=2.00	.118	.064	.160	.055	.085	.067	.085	.021
	β=2.25	.609	.136	.486	.201	.871	.180	.703	.377
	β=2.50	.964	.310	.847	.635	.999	.518	.985	.933
	β=3.00	.999	.744	.979	.965	1.00	.955	1.00	1.00

TABLE 4.10

$\hat{\pi}$'s for model 3, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.65	a=.95	G.L.S.	MIS	a=.35	a=.65
	α=1.00	.030	.028	.015	.007	.019	.011	.036	.016
	α=1.25	.680	.056	.483	.329	.957	.067	.918	.897
	α=1.50	.998	.160	.978	.945	1.00	.356	.998	.999
	α=2.00	1.00	.644	1.00	1.00	1.00	.961	1.00	1.00
1%									
	β=2.00	.030	.005	.011	.004	.016	.007	.023	.008
	β=2.25	.587	.032	.396	.276	.865	.055	.850	.790
	β=2.50	.978	.120	.931	.864	1.00	.250	1.00	1.00
	β=3.00	1.00	.522	.999	.999	1.00	.851	1.00	1.00
	α=1.00	.091	.083	.057	.034	.064	.058	.100	.048
	α=1.25	.861	.139	.697	.593	.999	.178	.972	.964
	α=1.50	1.00	.344	.988	.985	1.00	.590	.999	.999
	α=2.00	1.00	.862	1.00	1.00	1.00	.988	1.00	1.00
5%									
	β=2.00	.084	.044	.047	.029	.073	.034	.085	.048
	β=2.25	.750	.107	.627	.504	.962	.153	.936	.925
	β=2.50	.994	.287	.972	.951	1.00	.498	1.00	1.00
	β=3.00	1.00	.724	1.00	1.00	1.00	.956	1.00	1.00

TABLE 4.11

$\hat{\pi}$'s for model 4, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.65	a=.95	G.L.S.	MIS	a=.65	a=.95
1%	α=1.00	.025	.038	.036	.024	.019	.028	.022	.011
	α=1.25	.087	.074	.090	.058	.140	.096	.140	.106
	α=1.50	.324	.211	.361	.221	.627	.393	.573	.520
	α=2.00	.907	.745	.847	.809	.997	.964	.990	.990
	β=2.00	.025	.018	.050	.030	.019	.013	.021	.015
	β=2.25	.094	.059	.099	.063	.157	.091	.146	.107
	β=2.50	.332	.170	.295	.226	.601	.359	.555	.495
	β=3.00	.860	.649	.813	.760	.995	.938	.984	.973
5%	α=1.00	.086	.114	.110	.081	.063	.084	.080	.057
	α=1.25	.184	.160	.208	.154	.318	.230	.312	.259
	α=1.50	.542	.395	.502	.426	.809	.622	.752	.717
	α=2.00	.973	.872	.934	.912	.999	.991	.997	.996
	β=2.00	.084	.067	.103	.069	.067	.064	.072	.048
	β=2.25	.218	.136	.211	.167	.323	.214	.297	.260
	β=2.50	.535	.344	.482	.331	.794	.604	.748	.703
	β=3.00	.939	.813	.908	.896	.999	.977	.996	.995

TABLE 4.12

$\hat{\pi}$'s for model 5, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.35	a=.65	G.L.S.	MIS	a=.35	a=.65
1%	α=1.00	.023	.023	.043	.011	.024	.025	.038	.019
	α=1.25	.968	.040	.831	.678	1.00	.058	.988	.973
	α=1.50	1.00	.120	.987	.993	1.00	.253	1.00	.999
	α=2.00	1.00	.551	1.00	1.00	1.00	.894	1.00	1.00
	β=2.00	.023	.010	.025	.004	.023	.011	.011	.002
	β=2.25	.868	.032	.737	.573	.999	.047	.983	.926
	β=2.50	1.00	.097	.989	.985	1.00	.193	1.00	1.00
	β=3.00	1.00	.388	.998	1.00	1.00	.734	1.00	1.00
5%	α=1.00	.071	.078	.102	.032	.064	.055	.088	.019
	α=1.25	.993	.116	.994	.855	1.00	.155	.990	.989
	α=1.50	1.00	.291	.994	.997	1.00	.479	.997	.999
	α=2.00	1.00	.771	1.00	1.00	1.00	.963	1.00	1.00
	β=2.00	.098	.057	.078	.023	.073	.061	.053	.019
	β=2.25	.950	.101	.865	.774	.999	.134	.998	.983
	β=2.50	1.00	.218	.992	.995	1.00	.374	1.00	1.00
	β=3.00	1.00	.591	1.00	1.00	1.00	.873	1.00	1.00

TABLE 4.13

$\hat{\pi}$'s for model 6, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.35	a=.65	G.L.S.	MIS	a=.35	a=.65
1%	α=1.00	.030	.034	.072	.032	.010	.023	.047	.045
	α=1.25	1.00	.049	.886	.610	1.00	.051	.990	.933
	α=1.50	1.00	.108	.972	.964	1.00	.172	.996	.993
	α=2.00	1.00	.379	.991	.990	1.00	.683	.999	.999
	β=2.00	.038	.014	.047	.007	.012	.008	.007	.003
	β=2.25	.979	.022	.818	.506	.999	.040	.991	.904
	β=2.50	1.00	.058	.970	.942	1.00	.167	1.00	.999
	β=3.00	1.00	.264	.993	.992	1.00	.569	1.00	1.00
5%	α=1.00	.081	.093	.114	.057	.063	.077	.093	.088
	α=1.25	1.00	.136	.941	.800	1.00	.146	.994	.973
	α=1.50	1.00	.248	.981	.979	1.00	.343	.997	.995
	α=2.00	1.00	.601	.993	.991	1.00	.855	.999	.999
	β=2.00	.088	.048	.092	.030	.067	.053	.033	.027
	β=2.25	.989	.078	.900	.724	1.00	.113	.997	.968
	β=2.50	1.00	.164	.979	.976	1.00	.311	1.00	1.00
	β=3.00	1.00	.479	.995	.997	1.00	.765	1.00	1.00

TABLE 4.14

$\hat{\pi}$'s for model 7, with sample sizes of T=100 and 200, and 1000 replications.

		T=100				T=200			
		G.L.S.	MIS	a=.95	a=1.25	G.L.S.	MIS	a=.65	a=.95
1%	α=1.00	.039	-	.052	.047	.028	-	.062	.048
	α=1.25	.056	-	.087	.089	.078	-	.134	.114
	α=1.50	.151	-	.190	.202	.248	-	.330	.327
	α=2.00	.531	-	.591	.622	.820	-	.873	.873
	β=2.00	.026	-	.041	.046	.015	-	.047	.040
	β=2.25	.145	-	.081	.082	.070	-	.140	.117
	β=2.50	.140	-	.207	.214	.280	-	.388	.359
	β=3.00	.530	-	.624	.632	.883	-	.922	.912
5%	α=1.00	.110	-	.139	.142	.083	-	.151	.125
	α=1.25	.153	-	.196	.197	.171	-	.251	.233
	α=1.50	.283	-	.333	.337	.437	-	.524	.508
	α=2.00	.723	-	.763	.773	.934	-	.948	.953
	β=2.00	.077	-	.119	.116	.070	-	.113	.098
	β=2.25	.134	-	.190	.192	.190	-	.269	.230
	β=2.50	.298	-	.367	.375	.500	-	.574	.560
	β=3.00	.736	-	.783	.788	.956	-	.958	.954

From Tables 4.8-4.14 we observe that the $\hat{\pi}$'s decrease with a, although their sensitivity to the increase in a seems surprisingly small. The test seems to reject the null hypothesis too frequently, especially in model 2, without any plausible explanation for this phenomena in this model. Also, as the sample size increases, it seems that the number of times that we reject the null hypothesis when it is true goes to the value predicted by the asymptotic theory (as we may expect). As far as the power of the tests is concerned, the power for the MIS estimator is surprisingly small compared with the parametric and the semiparametric estimators. This is especially true in those models where the "loss of efficiency", measured by the M.S.E., is bigger, as we may expect. We also note that the power of the test, when the semiparametric estimator was used, tends to be very close to the power for the parametric estimator. Only in model 1 is the power of the tests for the different estimators very close one from the each other, although the power for the parametric estimator seems slightly better.

Also, it is worth mentioning two further points. Firstly, the sensitivity of the $\hat{\pi}$'s to the bandwidth parameter used for the semiparametric estimator is less as the sample size increases. Secondly, the power of the test of the MIS estimator compared with the parametric will depend presumingly on the severity of non-linearity of the autoregression function, as we may suspect from the results obtained in tables 4.1-4.7 of the previous section.

CHAPTER 5

ADAPTIVE ESTIMATION IN A HETEROSCEDASTIC TIME SERIES REGRESSION MODEL

5.1.- INTRODUCTION

In econometric regression models serial correlation and heteroscedasticity are two familiar non-standard features of disturbance behaviour. In econometric time series models, serial correlation in the disturbance is a rule rather than the exception. On the other hand, heteroscedasticity appears when the variance across the observations varies. Although heteroscedasticity has been identified more for cross-section data, there is no reason to restrict it to these sort of observations. In fact, in pure time series modelling where no regressors appear, Box-Cox transformation may be a way to eliminate heteroscedasticity. An example is that the variance depends on the level of the data. A way to stabilize the variance is to take logarithms of the series. Thus, an econometric time series model where both serial correlation and heteroscedasticity are present is not unrealistic.

There is a large literature concerning efficiency under serial correlation or heteroscedasticity. The case in which both "problems" are present has led some research, see for instance Harrison and McCabe (1975). In a semiparametric set up Harvey and Robinson (1988) allow the residuals to be heteroscedastic and serial correlated, being the serial correlation generated by an AR(P) process with known P and the heteroscedasticity an unknown function depending on t.

Consider the multiple regression model:

$$(1.1) \qquad y_t = \beta' x_t + u_t$$

$$u_t = \sigma(x_t) v_t$$

where β and x_t are k-dimensional column vectors and y_t, $\sigma(x_t)$ and u_t are scalars. Let $\{v_t\}_{t=1,2,\ldots}$ be an unobservable strictly stationary stochastic process with zero mean and variance 1, defined as:

$$v_t = \sum_{j=0}^{\infty} \alpha_j \, \eta_{t-j} \qquad \text{and} \qquad \sum_{j=0}^{\infty} |\alpha_j| < \infty$$

where η_t is i.i.d process. It follows that:

$$E[y_t | x_t] = \beta' x_t$$

and

$$\mathrm{Var}[y_t | x_t] = \mathrm{Var}(\sigma(x_t) v_t | x_t)$$

$$= \sigma^2(x_t).$$

On the basis of T observations y_t and x_t t=1,...,T, the generalized least squares estimator of β has an asymptotic covariance matrix

$$[E[X'\sigma^{-1}\Gamma^{-1}\sigma^{-1}X]]^{-1}$$

where σ=diag($\sigma(x_1),\dots,\sigma(x_T)$), Γ is the covariance matrix of the process v_t and X is the regression matrix defined as $(x_1,x_2,\dots,x_T)'$, or in the frequency domain the above covariance matrix can be written as:

$$(1.2) \quad \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{\overline{vv}}^{-1}(\lambda) f_{\overline{xx}}(-\lambda) d\lambda \right]^{-1}$$

where $f_{\overline{vv}}(\lambda)$ is the spectral density function of v_t and which is going to be assumed, from now on, bounded away from zero, i.e. $\exists$ d such that $f_{\overline{vv}}(\lambda) \geq d > 0$ $\forall \lambda \epsilon [-\pi,\pi]$ and $f_{\overline{xx}}(\lambda)$ is the spectral density function of the process $x/\sigma(x)$, which unlike $f_{\overline{vv}}(\lambda)$ is not needed to be bounded away from zero.

Remark

Here, it is interesting to mention that the spectral distribution function of $x/\sigma(x)$ need not be absolutely continuous for the results to follow, although for expositional purposes it will be assumed that this is the case.

Since $\sigma(x_t)$ and Γ are unknown to the practitioner, often he assumes a parametric form for them and estimates the resulting parameters based on the O.L.S. residuals of equation (1.1).

But, an incorrect parameterization of $\sigma(x_t)$ or Γ, although it will not prevent consistency of β in (1.1), will produce inconsistent estimates of their asymptotic covariance matrix. It is, of course, possible to estimate consistently the asymptotic covariance matrix of the O.L.S. estimator in the presence of both serial dependence and heteroscedasticity as first shown by Eicker (1963,1967). But on the other hand, while this leads to asymptotically valid inferences, they will not be asymptotically efficient, and any test based on O.L.S. estimates will be asymptotically suboptimal compared with the G.L.S. estimator.

Unless the parameterization of both serial dependence and heteroscedasticity is of interest on its own, it will be better to avoid some parameterization of them, and to use instead some nonparametric estimator of Γ and $\sigma(x)$. Those nonparametric estimators are inserted in the estimator of the β parameters which would be obtained if Γ and $\sigma(x)$ were known. The goal is to see that the semiparametric estimator of β does not entail any loss in asymptotic first-order efficiency relative to the case where Γ and $\sigma(x)$ are known.

This adaptive estimator extends the one studied by Harvey and Robinson (1988) by allowing the covariance matrix Γ to be unknown, and by not requiring $\sigma(x_t)$ to depend only on t, and that of Robinson (1987a) and Carroll (1982) by additionally allowing serial dependence instead of i.i.d. Also our semiparametric estimate extends Hannan's estimate in the opposite way, by allowing the disturbances to be heteroscedastic instead of homoscedastic.

The estimator that we are going to use for the parameters β is of the same kind as that of Hannan (1963; 1970, chapter 7), i.e. we use spectral techniques rather than time domain ones, but the unknown function $\sigma(x_t)$ is replaced by a Watson-Nadaraya kernel type estimator.

5.2.- THE SEMIPARAMETRIC ESTIMATOR OF β AND NOTATION

Before describing the semiparametric estimator of β and the natural extension of Hannan's (1963) estimator to the case of unknown heteroscedasticity, it is worthwhile introducing some notation and the kernel estimator of the conditional variance $\sigma^2(x_t)$.

Let K_1 be a function defined as in section 2.2.1 of chapter 2 with characteristic exponent q>0 defined as in definition 2.2.3 of such a chapter.

Let K be a function $:R^k \rightarrow R$, that it non-negative and integrates 1. Also, this function is assumed to be even. We define the Watson-Nadaraya kernel estimator of the conditional variance as it was defined in definition 2.1.2. of chapter 2 by:

$$(2.1) \qquad \hat{\sigma}_t^2 = \sum_{j=1}^{T} \hat{u}_j^2 \, K\left[\frac{x_j - x_t}{a}\right] I_t \Big/ \sum_{j=1}^{T} K\left[\frac{x_j - x_t}{a}\right]$$

where I_t is the indicator function defined as:

$$I_t = \begin{cases} 1 & \text{if } \hat{g}(x_t) > b \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{g}(x_t)$ stands for the kernel probability density estimator of the probability density function $g(x_t)$ of the regressors x_t, $\hat{u}_j$ stands for the O.L.S. residuals $y_t - x_t'\bar{\beta}^{o.l.s.}$ of the model (1.1), and b is a parameter which goes to zero as $T \to \infty$. Let us also define the estimator of $\sigma(x_t)$ as the positive square root of $\hat{\sigma}(x_t)^2$. (Recall that the estimator of the conditional variance is greater than or equal to zero since the kernel function is greater than or equal to zero).

Note that $\hat{\sigma}_t^2$ does use all the $\hat{u}_t^2$, so there is not any element of "sampling splitting" in $\hat{\beta}$ as in Carroll (1982), a feature not employed by Bickel (1982), Manski (1984) in different adaptive estimator or Robinson (1987a) where he discards $\hat{u}_t^2$ for the estimation of $\sigma(x_t)^2$.

Before introducing the semiparametric estimator, some notation will be needed.

$$\tilde{f}_{\overline{vv}}(\lambda) = \frac{1}{2\pi} \sum_{-M+1}^{M} K_1\left(\frac{j}{M}\right) e^{ij\lambda} \; \frac{1}{T} \sum \frac{\hat{u}_t \hat{u}_{t+j}}{\sigma_t \sigma_{t+j}} .$$

$$\tilde{f}_{\hat{v}\hat{v}}(\lambda) = \frac{1}{2\pi} \sum_{-M+1}^{M} K_1\left(\frac{j}{M}\right) e^{ij\lambda} \; \frac{1}{T} \sum \frac{\hat{u}_t \hat{u}_{t+j}}{\hat{\sigma}_t \hat{\sigma}_{t+j}} \; I_t I_{t+j} .$$

$$\hat{f}_{\hat{v}\hat{v}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{u_t u_{t+j}}{\hat{\sigma}_t\hat{\sigma}_{t+j}}\ I_t I_{t+j}\ .$$

$$\hat{f}_{\overline{vv}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{u_t u_{t+j}}{\sigma_t\sigma_{t+j}}\ .$$

$$\hat{f}_{\overline{yx}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{y_t x_{t+j}}{\sigma_t\sigma_{t+j}}\ .$$

$$\hat{f}_{\hat{y}\hat{x}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{y_t x_{t+j}}{\hat{\sigma}_t\hat{\sigma}_{t+j}}\ I_t I_{t+j}\ .$$

$$\hat{f}_{\hat{x}\hat{x}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{x_t x'_{t+j}}{\hat{\sigma}_t\hat{\sigma}_{t+j}}\ I_t I_{t+j}\ .$$

$$\hat{f}_{\overline{xx}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{x_t x'_{t+j}}{\sigma_t\sigma_{t+j}}\ .$$

$$\hat{f}_{\hat{v}\hat{x}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{u_t x_{t+j}}{\hat{\sigma}_t\hat{\sigma}_{t+j}}\ I_t I_{t+j}\ .$$

$$f_{\overline{vx}}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum \frac{u_t x_{t+j}}{\sigma_t\sigma_{t+j}}\ .$$

In the above equations, the second summand runs over all terms for which the products are defined.

We already know from Hannan (1963) (see also Hannan (1970) chapter 7) that if the $\sigma(x_t)^2$'s were known, then an asymptotically efficient estimator of β in equation (1.1) is given by:

$$\beta^* = \left[\frac{1}{2M}\sum_{-M+1}^{M} \hat{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}}\left[\frac{-\pi k}{M}\right]\ \right]^{-1}\left[\frac{1}{2M}\sum_{-M+1}^{M} \hat{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xy}}\left[\frac{-\pi k}{M}\right]\ \right]$$

which may be rewritten as:

$$(2.2)\qquad \beta^* = \beta + \left[\frac{1}{2M}\sum_{-M+1}^{M} \hat{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}}\left[\frac{-\pi k}{M}\right]\ \right]^{-1}\left[\frac{1}{2M}\sum_{-M+1}^{M} \hat{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]f_{\overline{xv}}\left[\frac{-\pi k}{M}\right]\ \right].$$

We also know that $\sqrt{T}(\beta^*-\beta)$ converges, under regularity conditions, to a normal distribution with zero mean and covariance matrix equal to the inverse of:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}_{\overline{vv}}(\lambda) f_{\overline{xx}}(-\lambda)d\lambda\ .$$

Observed that the only difference between the β^* estimator and the one proposed by Hannan (op. cit.) is that instead of the spectral density function of x_t we have the spectral density function of the process $x_t/\sigma(x_t)$.

<u>Remark 2.1</u>

It is worthwhile noting that Hannan (1963 and 1970 chapter 7) assumes that:

$$v_t = \sum_{j=-\infty}^{\infty} \alpha_j \; \eta_{t-j}$$

but by an extension to the well known Paley-Wiener theorem to the discrete-time case as long as the spectral density function is bounded away from zero and continuous, then v_t can be written as

$$v_t = \sum_{j=0}^{\infty} \alpha_j \; \eta_{t-j}$$

and thus, it does not represent a real restriction to the model.

Of course the $\sigma(x_t)$'s are unknown to the practitioner and therefore the estimator of β given by equation (2.2) is not feasible. Thus, we have to substitute this unfeasible "G.L.S." estimator of β by its corresponding feasible one, where the $\sigma^2(x_t)$'s are substituted by some estimated values. In this case the functional form of the heteroscedasticity is unknown and thus, a nonparametric estimator is called for. Thus, a sensible estimator of β will be

$$(2.3) \qquad \hat{\beta} = \left[\frac{1}{2M} \sum_{-M+1}^{M} \tilde{f}_{\hat{v}\hat{v}}^{-1}\left[\frac{\pi k}{M}\right] \hat{f}_{\hat{x}\hat{x}}\left[\frac{-\pi k}{M}\right] \right]^{-1} \left[\frac{1}{2M} \sum_{-M+1}^{M} \tilde{f}_{\hat{v}\hat{v}}^{-1}\left[\frac{\pi k}{M}\right] \hat{f}_{\hat{y}\hat{x}}\left[\frac{-\pi k}{M}\right] \right].$$

The aim will be to see that $\hat{\beta}$ is an adaptive estimator of β, i.e. $\sqrt{T}(\hat{\beta}-\beta)$ is asymptotically normal with zero mean and covariance matrix given by equation (1.2).

But to show the adaptation of the semiparametric estimator is enough to show that:

$$\sqrt{T}(\beta^{*}-\hat{\beta})=o_p(1)$$

because of $\sqrt{T}(\beta^{*}-\beta)$ is asymptotically normal with zero mean and covariance matrix given by (1.2). Noting that equation (2.3) can be written as:

$$\hat{\beta}-\beta+\left[\frac{1}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{x}\hat{x}}\left[\frac{-\pi k}{M}\right]\right]^{-1}\left[\frac{1}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{v}\hat{x}}\left[\frac{\pi k}{M}\right]\right].$$

Then it suffices to note that the following two expressions:

$$(2.4)\qquad \left[\frac{1}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{x}\hat{x}}\left[\frac{-\pi k}{M}\right]\right]-\left[\frac{1}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}}\left[\frac{-\pi k}{M}\right]\right]$$

and

$$(2.5)\qquad \left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{v}\hat{x}}\left[\frac{-\pi k}{M}\right]\right]-\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]f_{\overline{vx}}\left[\frac{-\pi k}{M}\right]\right]$$

are both $o_p(1)$.

To finish this section it is worth pointing out that several alternatives to the conditional variance estimator (2.1) have appeared in the literature. By noting that $E[u^2_t|x_t]=E[y^2_t|x_t]-(E[y_t|x_t])^2$, a natural alternative estimator to (2.1) is given by:

$$(2.6)\qquad \sigma^{*}(x_t)^2=\frac{\sum_{s=1}^{T} y_s^2\, K_{s,t}\, I_t}{\sum K_{s,t}}-\left[\frac{\sum_{s=1}^{T} y_s\, K_{s,t}\, I_t}{\sum K_{s,t}}\right]^2.$$

Another two alternatives are:

$$(2.7) \qquad \tilde{\sigma}(x_t)^2 = \sum_{j=1}^{T} \left[\hat{u}_j^2 / \dot{\sigma}_j^2\right] K\left[\frac{x_j - x_t}{a}\right] I_t \Big/ \sum_{j=1}^{T} K\left[\frac{x_j - x_t}{a}\right]$$

and

$$(2.8) \qquad \bar{\sigma}(x_t)^2 = \sum_{j=1}^{T} (\hat{u}_j^2 - \dot{\sigma}_j^2) K\left[\frac{x_j - x_t}{a}\right] I_t \Big/ \sum_{j=1}^{T} K\left[\frac{x_j - x_t}{a}\right]$$

where $\dot{\sigma}_s^2$ are nonnegative given proxies or estimated finite parameterization of $\sigma(x_t)^2$, for example the L.S. estimate on, say, a polynomial of x_t and $K_{s,t}$ stands for $K((x_s - x_t)/a)$.

5.3.- CONDITIONS AND ITS IMPLEMENTATION

In order to establish the asymptotic efficiency of $\hat{\beta}$ we introduce the following conditions.

CONDITIONS

C1.- Let $\{v_t\}$ be a stationary absolutely regular stochastic process with zero mean and variance 1 and the fourth cumulant is absolutely summable in all of its arguments. The absolutely regular coefficient $\zeta(n)$ satisfies $\zeta(n) = O(n^{-(2+\delta)/\delta})$, and v_t has finite first $6+3\delta'$ moments where $0<\delta<\delta'$. Also:

$$\sum_{s=0}^{\infty} |\gamma(s)||s|^{q}<\infty$$

where $\gamma(s)$ is the autocovariance function of the stochastic process v_t and $q>\frac{1}{2}$.

C2.- Let $\{x_t\}$ be a vector stationary strong-mixing stochastic process with finite first $6+3\lambda$ moments, for some $\lambda>0$, and where the strong mixing coefficient $\alpha(n)$ satisfies:

$$\sum_{n} \alpha(n)^{\lambda/(2+\lambda)}<\infty.$$

C3.- x_t and v_s are independent for all t and s.

C4.- $T^{-1-\gamma}M^{3}a^{-k}b^{-1}\longrightarrow 0$ where $M \longrightarrow\infty$, $a\longrightarrow 0$ and $b\longrightarrow 0$

as $T\longrightarrow\infty$ and $\gamma=\frac{2(\delta'-\delta)}{\delta(2+\delta')}$.

C5.- $M\, a^{4}b^{-2}\longrightarrow 0$.

C6.- $T^{-2}M^{3}a^{-4k}b^{-2}\longrightarrow 0$.

C7.- $M\, T^{-\frac{1}{2}}\longrightarrow 0$.

C8.- The regressors x_t weighted by the heteroscedascity, i.e. $x_t/\sigma(x_t)$, satisfy

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_t\, x'_t}{\sigma(x_t)^2}\xrightarrow{P}\Xi_{kx_k} \text{ positive definite (p.d.)}$$

although to some extent C2 will be responsible for this.

Observe that this condition does not permit trending behaviour in the x_t's variables, as Hannan (1963; 1970, chapter 7) or Harvey and Robinson (1988) does, although that case may be handled if $\sigma(x_t)$ would not depend on the trending variables, and $^1/_T$ is substituted appropriately. Also condition C8 is a sufficient condition so that the covariance matrix of the β estimator, i.e. expression (1.2), is p.d., since the matrix Ξ is p.d. and the spectral density function of the process v_t ($f_{\overline{vv}}(\lambda)$) is bounded away from zero $\forall\lambda$ (see Hannan (1970), pp 427-428).

C9.- $K_1(s)$ is a bounded even function defined on (-1,1) and $K_1(0)=1$ with characteristic exponent, see definition 2.2.3. of chapter 2, of order $r>q$, where q is as in C1.

C10.- $\sigma(x_t)^2>c>0 \quad \forall t$.

This condition will prevent us from having infinite weights in equation (2.2).

C11.- $\sigma(x_t)^2 g(x_t)$ is twice differentiable with respect to x_t.

This condition will allow the bias of our kernel estimator of $\sigma^2(x_t)$ to tend sufficiently fast to zero.

C12.- The kernel function K has absolutely integrable characteristic function $\kappa(z)$.

Having written down some sufficient conditions to establish the theorem given in the next section, we now discuss its implementation and regularity conditions.

<u>Comments on implementation and regularity conditions</u>

The semiparametric estimator $\hat{\beta}$ involves two bandwidths or smoothing parameters which we have called a and M. These bandwidth's choices are more crucial than the kernel and lag window functions K and K_1 respectively. The nonparametric literature discusses several automatic bandwidth determination procedures based on function optimization, although reproducing the asymptotic theory with a data-driven a and M would not be easy to implement.

The choice of the lag window K_1 has been exhaustively studied in the spectral estimation literature (see e.g. Hannan, 1970, Ch. 5). Eicker (1967), Levine (1983) White and Domowitz (1984) use $K_1 \equiv 1$, although it is quite unpopular since it does not sufficiently protect against influence of spectral peaks at distant frequencies. Newey and West (1987) used the modified Barlett window $K_1(m)=1-|m|$, i.e equation (2.4) of chapter 2, and Robinson (1987b) used a Hanning-Tukey type given by equation (2.5) of chapter 2.

In spectral estimation, an automatic method to choose the lag number M has been justified by Beltrao and Bloomfield (1987). In a time series regression model Robinson (1988a) has established the asymptotic normality and efficiency of a modification of Hannan's

(1963 and 1970, chapter 7) estimator, when the parameter M optimizes a form of cross-validated Gaussian pseudo-likelihood, i.e. Robinson allows M to be data-dependent.

The absolutely regular condition imposed on α_j is stronger than the condition $\sum|j|^q|\gamma(j)|<\infty$, since a sufficient condition for the latter to be true is $\alpha_j=o(j^{-q-1})$, as we can conclude for Pham and Tram (1985), see section 2.3.1 of chapter 2 $\alpha_j=o(j^{-2-\alpha})$, being $q<2$ and $\alpha=^{\delta+2}/_{\delta}>1$ as usual. Also note that for the case in which η_t is a Gaussian process ω may be any positive number and thus $\alpha_j=O(j^{-\alpha-2})$.

Assumption C7 is a standard one in spectral density estimation. C5 is true if M is taken to be $O(a^{-4}b^2)$. This leads us to say that C6 is true if $T^{-1}a^{-4(k+1)}$ is finite or goes to zero, and hence C4 will be a stronger assumption than C6 depending on whether $\gamma>\frac{1}{4}$ or $\gamma<\frac{1}{4}$. So far the case $\gamma>\frac{1}{4}$ and $k=1$, a can be chosen to be of order $O(T^{-\frac{1}{8}})$. Then $M=O(T^{\frac{1}{2}}b^2)$ and C7 is satisfied as well. Also condition C5 could be relaxed if higher order kernels were used, but this would imply that $\hat{\sigma}_t{}^2$ can be negative and this would not be very useful since in the proofs of propositions 1 and 2 we need an estimate of the standard deviation instead of the variance.

Finally, one point to mention is that in the semiparametric estimator $\hat{\beta}$ we have not used higher order kernel functions to estimate the conditional variances, even in the case of large k. This is quite unusual in semiparametric problems where higher order kernels are frequently used to reduce the bias of the function estimates especially when the dimension of the nonparametric function, i.e. k, is large, although it should be noted that neither

Carroll (1982) or Robinson (1987a) use them.

5.4.- THEOREM

In this section it will be established in the theorem below the adaptiveness of the semiparametric estimator $\hat{\beta}$. Due to the lengthy proof, we will prove two propositions which will allow us to follow the proof of the theorem more easily.

Proposition 1

Under C1 to C12:

$$(4.1)\quad \left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{v}\hat{x}}\left[-\frac{\pi k}{M}\right]\right]-\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]f_{\overline{vx}}\left[-\frac{\pi k}{M}\right]\right]=o_p(1).$$

Proof:

The L.H.S. of (4.1) is equal to:

$$(4.2)\quad \left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{v}\hat{x}}\left[-\frac{\pi k}{M}\right]\right]-\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}I}\left[\frac{\pi k}{M}\right]f_{\overline{vx}I}\left[-\frac{\pi k}{M}\right]\right]$$

$$(4.3)\quad -\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}(1-I)}\left[\frac{\pi k}{M}\right]f_{\overline{vx}I}\left[-\frac{\pi k}{M}\right]\right]$$

$$(4.4)\quad -\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right] f_{\overline{vx}(1-I)}\left[\frac{\pi k}{M}\right]\right]$$

where:

$$f_{\overline{vx}I}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|j|} x_{t+j}\ \frac{u_t}{\sigma_t\sigma_{t+j}} I_t I_{t+j}$$

and

$$f_{\overline{vx}(1-I)}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|j|} x_{t+j}\ \frac{u_t}{\sigma_t\sigma_{t+j}}(1-I_t I_{t+j}).$$

- About (4.4):

This term is $o_p(1)$ by lemma 2 of appendix A.

- About (4.3)

This term may be rewritten as:

$$\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M}\left[\tilde{f}^{-1}_{\overline{vv}(1-I)}\left[\frac{\pi k}{M}\right]-f^{-1}_{\overline{vv}(1-I)}\left[\frac{\pi k}{M}\right]\right] f_{\overline{vx}I}\left[\frac{\pi k}{M}\right]$$

$$+\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} f^{-1}_{\overline{vv}(1-I)}\left[\frac{\pi k}{M}\right]\ f_{\overline{vx}I}\left[\frac{\pi k}{M}\right]$$

where $\tilde{f}^{-1}{}_{\overline{vv}(1-I)}=\tilde{f}^{-1}{}_{\overline{vv}}-\tilde{f}^{-1}{}_{\overline{vv}I}$ and

$$\tilde{f}_{\overline{vv}I}(\lambda)=\frac{1}{2\pi}\sum_{-M+1}^{M} K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|j|}\frac{\hat{u}_t\hat{u}_{t+j}}{\sigma_t\sigma_{t+j}}\ I_t I_{t+j}.$$

The proof that the first term of the above summand is $o_p(1)$ is identical to the proof of lemma 2. Also it must be taken into account that:

$$\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M}\left[\hat{f}^{-1}_{\overline{vv}}\left(\frac{\pi k}{M}\right)- f^{-1}_{\overline{vv}}\left(\frac{\pi K}{M}\right)\right]f_{\overline{vx}}\left(-\frac{\pi k}{M}\right)=O_p(M^2T^{-1}+M^{1-2q})$$

see for instance Hannan (1963).

The second term is more cumbersome.

First note that:

$$f^{-1}_{\overline{vv}(1-I)}=f^{-1}_{\overline{vv}}-f^{-1}_{\overline{vv}I}$$

where

$$f_{\overline{vv}I}=\frac{1}{2\pi}\sum_{j=-T+1}^{T-1}\gamma(j)\; e^{ij\lambda}\left[\frac{1}{T}\sum_{t=1}^{T-|j|}I_tI_{t+j}\right].$$

Let the term in brackets be $h(j)={}^1/_T\sum I_iI_{i+j}$, and call $\bar{\gamma}(j)=\gamma(j)h(j)$. If $f_{\overline{vv}I}$ were greater than or equal to $d>0$, we could use a theorem due to Wiener (see Naimark 1960, pg-205), and conclude that the inverse of $f_{\overline{vv}I}$ has an absolutely convergent Fourier series, i.e. their coefficients belong to the L_1 space. But to prove that $f_{\overline{vv}I}$ is greater than or equal to $d>0$, all that is needed is to show that:

$$\sup_{\lambda}\left|\frac{1}{2\pi}\sum_{|j|<T}[\gamma(j)-\bar{\gamma}(j)]e^{ij\lambda}+\frac{1}{2\pi}\sum_{|j|>T}\gamma(j)e^{ij\lambda}\right|=o_p(1).$$

The expectation of the L.H.S. of that equation is less than or equal (except for the 2π factor) to:

$$\sum_{|j|<T} |\gamma(j)|\, E|1-h(j)| + \sum_{|j|>T} |\gamma(j)|.$$

But $E|1-h(j)|=o(1)$ $\forall j$ by lemma 6 of Robinson (1988c) and lemma 2 of chapter 3, or by using proposition 4 of Robinson (1988c), since

$$\sup_z |\hat{g}(z)-g(z)|=o_p(1).$$

Hence, $f_{\overline{vv}I}>d>0$ and thus we can conclude that:

$$f^{-1}_{\overline{vv}I}=\sum_j \bar{\Delta}(j)e^{ij\lambda}$$

where $\bar{\Delta}(j)$ belongs to the L_1-space.

On the other hand we have by definition that:

$$\Big[\sum_j \gamma(j)e^{ij\lambda}\Big]\Big[\sum_j \Delta(j)e^{ij\lambda}\Big]=1$$

and

$$\Big[\sum_{|j|<T} \bar{\gamma}(j)e^{ij\lambda}\Big]\Big[\sum_j \bar{\Delta}(j)e^{ij\lambda}\Big]=1$$

and also that $\bar{\gamma}(j)-\gamma(j)=o_p(1)$, $\forall j$.

From the above equations we can straightforwardly conclude that:

1) $$E\Big|\Big(\sum_j \gamma(j)e^{ij\lambda}\Big)\Big(\sum_j \bar{\Delta}(j)e^{ij\lambda}\Big)-1\Big|=o(1)$$

and

$$2)\quad E|(\sum_j \{\Delta(j)-\bar{\Delta}(j)\}e^{ij\lambda}|= E|\sum_j \tilde{\Delta}(j)e^{ij\lambda}|=o(1)$$

since the spectral density function of the process v_t is bounded away from zero by assumption.

The next question will be to show that $\Delta(j)-\bar{\Delta}(j)=\tilde{\Delta}(j)$ goes to zero in probability. But since $\tilde{\Delta}(j)$ belongs to L_1, we can say by the inverse of the Fourier transform that $\tilde{\Delta}(s)$ is equal $\forall s$ to:

$$\tilde{\Delta}(s)=\Delta(s)-\bar{\Delta}(s)=\frac{1}{2\pi}\int_{-\pi}^{\pi}(\sum_j\tilde{\Delta}(j)e^{ij\lambda})e^{-is\lambda}d\lambda.$$

Therefore

$$E|\tilde{\Delta}(s)|\leq\frac{1}{2\pi}\int_{-\pi}^{\pi}E|\sum\tilde{\Delta}(j)e^{ij\lambda})|\ d\lambda$$

$$\leq(2\pi)^{-1}\int_{-\pi}^{\pi}\varepsilon\ d\lambda$$

because the expectation inside the integrand is less than or equal to ε and therefore the above equation is equal to ε.

Since it is $\forall\varepsilon>0$, we conclude that $\tilde{\Delta}(j)\longrightarrow 0$ in probability.

With those ingredients we can show that:

$$\frac{1}{2\pi}\sum_{-M+1}^{M}\tilde{\Delta}(j)\ K_1\left[\frac{j}{M}\right]e^{ij\lambda}\ \left(\frac{1}{T^{\frac{1}{2}}}\sum_{t=1}^{T-|j|}x_{t+j}\ \frac{u_t}{\sigma_t\sigma_{t+j}}\ I_tI_{t+j}\right)$$

is $o_p(1)$ by the following argument. We already know that $\tilde{\Delta}(j)$ is

$o_p(1)$ and the term in parenthesis is $O_p(1)$. But the term $\tilde{\Delta}(j)$ also belongs to the space L_1 thus, we can apply Fatou lemma in order to be able to introduce the **lim** operation after the expectation has been taken, that is:

$$\lim \frac{1}{2\pi} \sum_{-M+1}^{M} E\left|\tilde{\Delta}(j)\ K_1\left[\frac{j}{M}\right] e^{ij\lambda} \left(\frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^{T-|j|} x_{t+j} \frac{u_t}{\sigma_t \sigma_{t+j}} I_t I_{t+j}\right)\right|$$

which is less than or equal, by Cauchy inequality, to

$$C\left(\sum_{-M+1}^{M} \lim(E|\tilde{\Delta}(j)|^2)^{\frac{1}{2}} \lim(E|\frac{1}{T^{\frac{1}{2}}} \sum_{t=1}^{T-|j|} x_{t+j} \frac{u_t}{\sigma_t \sigma_{t+j}} I_t I_{t+j}|^2)^{\frac{1}{2}}\right.$$

but $\lim E|\tilde{\Delta}(j)|^2=0$ since $E|\tilde{\Delta}(j)|^2 < \sup|\tilde{\Delta}(j)|\ E|\tilde{\Delta}(j)|$ and then, the above term is less than or equal to

$C \sum 0 = 0$, and so (4.3) is $o_p(1)$.

Hence, (4.1) will be true if

$$(4.5)\quad \left[\frac{\sqrt{T}}{2M} \sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right] \hat{f}_{\hat{v}\hat{x}}\left[-\frac{\pi k}{M}\right]\ \right] - \left[\frac{\sqrt{T}}{2M} \sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vvI}}\left[\frac{\pi k}{M}\right] f_{\overline{vxI}}\left[-\frac{\pi k}{M}\right]\right] = o_p(1).$$

But in order to prove this fact, it will be sufficient to show that the next three terms are in fact $o_p(1)$, namely

$$(4.6)\qquad \frac{T^{\frac{1}{2}}}{2M} \sum_{-M+1}^{M} \left[\tilde{f}^{-1}_{\hat{v}\hat{v}} - \tilde{f}^{-1}_{\overline{vvI}}\right] (\hat{f}_{\hat{v}\hat{x}} - f_{\overline{vxI}}),$$

$$(4.7)\qquad \frac{T^{\frac{1}{2}}}{2M}\sum_{-M+1}^{M}\left[\tilde{f}^{-1}_{\hat{v}\hat{v}}-\tilde{f}^{-1}_{\overline{vvI}}\right] f_{\overline{vxI}}$$

and

$$(4.8)\qquad \frac{T^{\frac{1}{2}}}{2M}\sum_{-M+1}^{M}\tilde{f}^{-1}_{\overline{vvI}}\,(\hat{f}_{\hat{v}\hat{x}}-f_{\overline{vxI}}).$$

By lemmas 8 and 9 is already known that:

$$E\left\|\tilde{f}^{-1}_{\hat{v}\hat{v}}-\tilde{f}^{-1}_{\overline{vvI}}\right\|^2< C\ O(T^{-2}M^3a^{-k}b^{-2}+T^{-2}M^2a^{-4k}b^{-2}+a^4b^{-2})$$

except on a set S whose measure tends to zero as T approaches to infinity.

On the other hand, by lemmas 1 and 3 and since $\sum K_1(s/_M)=O(M)$;

$$T\ E\left\|\hat{f}_{\hat{v}\hat{x}}-f_{\overline{vxI}}\right\|^2< O_p(T^{-\gamma}M^2a^{-k}b^{-2}+T^{-1}M^2a^{-4k}b^{-1}+M^2a^4b^{-1})$$

and finally, we know that:

$$\left\|T^{\frac{1}{2}}f_{\overline{vxI}}\right\|^2= O_p(M)$$

by the same argument as those used by Hannan (1963) pg. 27.

As far as (4.8) is concerned, we can split this term into two terms:

$$(4.8.1)\qquad \frac{T^{\frac{1}{2}}}{2M}\sum_{-M+1}^{M}f^{-1}_{\overline{vvI}}\,(\hat{f}_{\hat{v}\hat{x}}-f_{\overline{vxI}})$$

and

$$(4.8.2)\quad + \frac{T^{\frac{1}{2}}}{2M} \sum_{-M+1}^{M} \left[\tilde{f}^{-1}_{\overline{vvI}} - f^{-1}_{\overline{vvI}} \right] (\hat{f}_{\hat{v}\hat{x}} - f_{\overline{vxI}}).$$

- Consider now (4.8.1):

All what we need to show is that

$T^{\frac{1}{2}}(c_{\hat{v}\hat{x}} - c_{\overline{vxI}})$ goes to zero in probability where

$$c_{\overline{vxI}} = \frac{1}{T} \sum_{t=1}^{T-|j|} x_{t+j} \frac{u_t}{\sigma_t \sigma_{t+j}} I_t I_{t+j}$$

and

$$c_{\hat{v}\hat{x}} = \frac{1}{T} \sum_{t=1}^{T-|j|} x_{t+j} \frac{u_t}{\hat{\sigma}_t \hat{\sigma}_{t+j}} I_t I_{t+j}$$

but by lemma 3 of appendix A, it is of probability order:

$$O_p(T^{-\gamma/2} a^{-k/2} b^{-1} + a^2 b^{-1}).$$

- About (4.8.2), it will be of order:

$$O_p(T^{-1-\gamma} M^3 a^{-k} b^{-1} + T^{-2} M^3 a^{-4k} b^{-1} + T^{-1} M^3 a^4 b^{-2}),$$

by lemmas 1 and 3 of appendix A and since $\sum K_1(s/M) = O(M)$.

And thus, (4.8) is $o_p(1)$.

- About (4.7). We already know that:

$$\frac{1}{2M}\sum_{-M+1}^{M} \left\| T^{\frac{1}{2}} f_{\overline{vxI}} \right\|^2 = O_p(M)$$

which will imply that (4.7) is of probability order:

$$O_p(T^{-2}M^3a^{-4k}b^{-2}+T^{-2}M^4a^{-k}b^{-2}+M\ a^4b^{-2})$$

by lemma 8 of appendix A. Therefore, it only remains to prove (4.6) to conclude the proposition.

- About (4.6):

If (4.7) and (4.8) are both $o_p(1)$ evidently (4.6) is also $o_p(1)$.

Q.E.D.

The next proposition will prove that (2.4) is $o_p(1)$.

<u>Proposition 2</u>

Under C1 to C12

$$(4.9)\qquad \left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{x}\hat{x}}\left[\frac{-\pi k}{M}\right]\right]-\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}}\left[\frac{-\pi k}{M}\right]\right]=o_p(1).$$

<u>Proof</u>:

The L.H.S. of (4.9) is equal to the following three terms:

$$(4.10)\qquad \left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left[\frac{\pi k}{M}\right]\hat{f}_{\hat{x}\hat{x}}\left[\frac{-\pi k}{M}\right]\right]-\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vvI}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xxI}}\left[\frac{-\pi k}{M}\right]\right]$$

$$(4.11)\quad -\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}(1-I)}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}I}\left[\frac{-\pi k}{M}\right]\right]$$

$$(4.12)\quad -\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}}\left[\frac{\pi k}{M}\right]\hat{f}_{\overline{xx}(1-I)}\left[\frac{-\pi k}{M}\right]\right].$$

In order to prove that (4.11) and (4.12) are both $o_p(1)$, note that the method follows the same lines as the proof of (4.3) and (4.4). Also observe that in this case, we do not have the factor $T^{\frac{1}{2}}$ as previously.

Now, we will show that the expression (4.10) is $o_p(1)$. To show that in fact this expression is $o_p(1)$, it is equivalent to show that each of the next three terms are in fact $o_p(1)$, namely

$$(4.13)\quad \frac{1}{2M}\sum_{-M+1}^{M} \left[\tilde{f}^{-1}_{\hat{v}\hat{v}} - \tilde{f}^{-1}_{\overline{vv}I}\right](\hat{f}_{\hat{x}\hat{x}} - \hat{f}_{\overline{xx}I})$$

$$(4.14)\quad + \frac{1}{2M}\sum_{-M+1}^{M} \left[\tilde{f}^{-1}_{\hat{v}\hat{v}} - \tilde{f}^{-1}_{\overline{vv}I}\right] \hat{f}_{\overline{xx}I}$$

$$(4.15)\quad + \frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}I}\ (\hat{f}_{\hat{x}\hat{x}} - \hat{f}_{\overline{xx}I}).$$

Noting that the basic difference with the terms (4.6)-(4.8) is that here we have:

$\hat{f}_{\hat{x}\hat{x}}$ instead of $T^{\frac{1}{2}}\ \hat{f}_{\hat{v}\hat{x}}$ and

$\hat{f}_{\overline{xxI}}$ instead of $T^{\frac{1}{2}}\ f_{\overline{xvI}}$.

In order to show that (4.15) is $o_p(1)$, we repeat the same arguments as we did for the term (4.8), but where lemma 1 of appendix A it is substituted $^1/\sqrt{T}$ by $^1/_T$ and u_t by x_t, and in lemma 3 of appendix A we substitute $^1/\sqrt{T}$ by $^1/_T$ and v_t by x_t.

As far as (4.14) is concerned, note that it is identical to what it was explained for (4.7), with the above remark. Finally, (4.13) is also $o_p(1)$ since (4.14) and (4.15) are.

Q.E.D.

THEOREM 1

Under C1 to C12

a) $\sqrt{T}(\beta^{*}-\hat{\beta})$ is $o_p(1)$

and

b) $\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}_{\hat{v}\hat{v}}^{-1}\left(\frac{\pi k}{M}\right)\hat{f}_{\hat{x}\hat{x}}\left(\frac{-\pi k}{M}\right)\ \right]^{-1}$ is a consistent estimator

of the covariance matrix (1.2).

Proof:

In order to prove part a) of the theorem it suffices to show that:

$$\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}_{\hat{v}\hat{v}}^{-1}\left(\frac{\pi k}{M}\right)\hat{f}_{\hat{v}\hat{x}}\left(-\frac{\pi k}{M}\right)\ \right]-\left[\frac{\sqrt{T}}{2M}\sum_{-M+1}^{M} \tilde{f}_{\overline{vv}}^{-1}\left(\frac{\pi k}{M}\right)f_{\overline{vx}}\left(-\frac{\pi k}{M}\right)\right]$$

and

$$\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left(\frac{\pi k}{M}\right)\hat{f}_{\hat{x}\hat{x}}\left(\frac{-\pi k}{M}\right)\right]-\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}}\left(\frac{\pi k}{M}\right)\hat{f}_{\overline{xx}}\left(\frac{-\pi k}{M}\right)\right]$$

are both $o_p(1)$, but they are $o_p(1)$ by propositions 1 and 2.

As regards part b) of the theorem, we already know that:

$$\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\overline{vv}}\left(\frac{\pi k}{M}\right)\hat{f}_{\overline{xx}}\left(\frac{-\pi k}{M}\right)\right]$$

converges in probability to the inverse of (1.2), and so by part a)

$$\left[\frac{1}{2M}\sum_{-M+1}^{M} \tilde{f}^{-1}_{\hat{v}\hat{v}}\left(\frac{\pi k}{M}\right)\hat{f}_{\hat{x}\hat{x}}\left(\frac{-\pi k}{M}\right)\right] \tag{4.16}$$

also converges in probability to the inverse of (1.2). But on the other hand we know that this matrix is p.d. and therefore the inverse of (4.16) converges in probability to (1.2).

Q.E.D.

5.5.- CONCLUSIONS

In this chapter we have studied first order efficiency in estimating the parameters of a linear regression model, when both heteroscedasticity and serial correlation are present but neither are parameterized. The conditions about the regressors and the stochastic process v_t are not much stronger than the conditions needed in a parametric setting. The semiparametric estimator of β turns out to be a natural extension of works by Hannan (1963 and 1970, chapter 7), Robinson (1987a) or Harvey and Robinson (1988).

Some possible extensions come to the mind. The first one, in view of the works by Hannan (1963) is the multivariate regression model. A second possible extension, which from an econometric point of view is more relevant, is the linear simultaneous equation model with exogenous regressors. And a third one is the non-linear multivariate regression model, in view of the work of Robinson (1973).

APPENDIX A

In this appendix we will prove a series of lemmas which are used in several places in the proof of the propositions 1 and 2 of the chapter.

Before attacking the following lemmas, some notation will be introduced which is used thereafter. We will define $\bar{\sigma}^2(x_t)$ as:

$$\bar{\sigma}_t^2 = \sum_{j=1}^{T} \sigma_j^2 K_{t,j} \Big/ \sum_{j=1}^{T} K_{t,j}$$

where $K_{j,t}=K(^{(x_j-x_t)}/_a)$.

Lemma 1

Let the conditions C1 to C3 and C9 hold and $\gamma = ^{2(\delta'-\delta)}/_{\delta(2+\delta')}$.

Then:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{u_t\, x_t}{\bar{\sigma}_t^3\, \hat{g}(x_t)} \left[\frac{1}{Ta^p} \sum_{j=1}^{T} (\hat{u}_j^2-\sigma_j^2)\, K_{t,j}\right] I_t = O_p(T^{-\gamma/2} a^{-k/2} b^{-1}).$$

Proof

The L.H.S. of the above equation can be rewritten as:

$$\text{(A.1)}\qquad \frac{1}{T^{1\frac{1}{2}}a^{k}}\sum_{t<j}\Big[\frac{u_t\ x_t}{\bar{\sigma}_t^3\ \hat{g}(x_t)}\ (\hat{u}_j^2-\sigma_j^2)\ K_{t,j}I_t +$$

$$\frac{u_j\ x_j}{\bar{\sigma}_j^3\ \hat{g}(x_j)}\ (\hat{u}_t^2-\sigma_t^2)\ K_{j,t}I_j\Big].$$

Here, what has been done is to make symmetric the double sum and thus, to be able to employ U-statistic theory for absolutely regular stochastic processes. Also, we have discarded the terms t=j, since as far as the asymptotic result is concerned it does not make any difference. Expression (A.1) can be written as:

$$\frac{1}{T^{1\frac{1}{2}}a^{k}}\sum_{t<j}\Big[\frac{u_t\ x_t}{\bar{\sigma}_t^3\ \hat{g}(x_t)}\ (u_j^2-\sigma_j^2)\ K_{t,j}I_t +$$

$$\frac{u_j\ x_j}{\bar{\sigma}_j^3\ \hat{g}(x_j)}\ (u_t^2-\sigma_t^2)\ K_{j,t}I_j\Big]$$

$$+\ \frac{(\beta-\beta^{o.l.s.})'}{T^{1\frac{1}{2}}a^{k}}\sum_{t<j}\Big[\frac{u_t\ x_t}{\bar{\sigma}_t^3\ \hat{g}(x_t)}\ x_j\ u_j\ K_{t,j}I_t +$$

$$\frac{u_j\ x_j}{\bar{\sigma}_j^3\ \hat{g}(x_j)}\ x_t\ u_t\ K_{j,t}I_j\Big]$$

$$+\ \frac{(\beta-\beta^{o.l.s.})'}{T^{1\frac{1}{2}}a^{k}}\sum_{t<j}\Big[\frac{u_t\ x_t}{\bar{\sigma}_t^3\ \hat{g}(x_t)}\ x_jx_j'\ K_{t,j}I_t +$$

$$\frac{u_j\ x_j}{\bar{\sigma}_j^3\ \hat{g}(x_j)}\ x_tx_t'\ K_{j,t}I_j\Big](\beta-\beta^{o.l.s.}).$$

If one calls $p_T(z_j, z_t)$ the term inside the first term of the above expression where z_i stands for (u_i, x_i), it is straightforward to show that its expectation is equal to zero when z_j and z_t have been taken as independent. Therefore, taking the expectations conditional on x_j or x_t we can conclude that this term is of probability order:

$$O_p[T^{-\gamma/2}a^{-k/2}b^{-1}]$$

by lemma 2 of Yoshihara (1976).

About the other two terms it is obvious that they are also:

$$O_p[T^{-\gamma/2}a^{-k/2}b^{-1}].$$ Q.E.D.

Lemma 2

Let the conditions C1, C2 and C7 hold. Then:

$$\text{(A.2)} \qquad \frac{T^{\frac{1}{2}}}{2M} \sum_{-M+1}^{M} \tilde{f}_{\overline{vv}}^{-1} \left[\frac{\pi k}{M}\right] f_{\overline{vx(1-I)}} \left[\frac{\pi k}{M}\right] = o_p(1).$$

Proof:

The L.H.S. of equation (A.2) is equal to:

$$\frac{T^{\frac{1}{2}}}{2M} \sum_{-M+1}^{M} \left[\tilde{f}_{\overline{vv}}^{-1} \left[\frac{\pi k}{M}\right] - f_{\overline{vv}}^{-1} \left[\frac{\pi k}{M}\right] \right] f_{\overline{vx(1-I)}} \left[\frac{\pi k}{M}\right]$$

$$+ \frac{T^{\frac{1}{2}}}{2M} \sum_{-M+1}^{M} f_{\overline{vv}}^{-1} \left[\frac{\pi k}{M}\right] f_{\overline{vx(1-I)}} \left[\frac{\pi k}{M}\right].$$

By noting that the difference between $\tilde{f}_{\overline{vv}}(\lambda)$ and $\hat{f}_{\overline{vv}}(\lambda)$ is that the former uses the O.L.S. residuals and by Hannan (1961) the latter is greater than or equal to d>0 except in a set S which measured tends to zero as T approches to infinity, we can say that the squared norm of the first term is dominated by a constant times

$$\frac{1}{2M}\sum_{-M+1}^{M} \left\| \tilde{f}_{\overline{vv}}\left[\frac{\pi k}{M}\right]-f_{\overline{vv}}\left[\frac{\pi k}{M}\right] \right\|^2 \frac{1}{2M}\sum_{-M+1}^{M} \left\| T^{\frac{1}{2}} f_{\overline{vx}(1-I)}\left[\frac{\pi k}{M}\right]\right\|^2.$$

The first term of the product is $O_p(T^{-1}M+M^{-2q})$ (see Hannan (1970) or section 2.2.1 of chapter 2). As regards the second term of the product, one can realize, following standard steps, that it is equal to:

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right]^2 \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-|s|} \frac{u_t\ x_{t+s}}{\sigma_t\ \sigma_{t+s}}(1-I_tI_{t+s})\right]^2.$$

This term is very easy to show that it is $o_p(M)$. Because of $u_t=v_t\sigma_t$ and the v_t are absolutely regular, the above term without the term $(1-I_tI_{t+s})$ is $O_p(1)$. Consequently, using lemma 6 and proposition 4 of Robinson (1988c) we can conclude by assumption C2 that the $E[1-I_tI_{t+s}]$ is $o(1)$, and therefore the above term is in fact $o_p(M)$. Hence, we can say that the first term is $o_p(M^2/T+M^{1-2q})$, which by assumption C7 and $q>\frac{1}{2}$ is $o_p(1)$. Therefore, it only remains to show that the second term is also $o_p(1)$. But this term is equal to:

$$\sum_{-M+1}^{M} \Delta(s) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} \frac{u_t \; x_{t+s}}{\sigma_t \; \sigma_{t+s}} (1-I_t I_{t+s})\right].$$

For the same argument as above, the term in brackets is $o_p(1)$ for all s. Moreover it is known that $\Delta(s)$ belongs to the L_1-space. Therefore we claim the Lebesgue dominated convergence theorem to conclude that this second term is also $o_p(1)$.

Q.E.D.

<u>Lemma 3</u>

Given the conditions C1, C2, C4, C7, C8 and C10 then:

$$\text{(A.3)} \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} v_t \; x_{t+s} (\sigma_{t+s}^{-1} - \hat{\sigma}_{t+s}^{-1}) I_t I_{t+s} = O_p(T^{-\frac{1}{2}} a^{-2k} b^{-2} + T^{-\frac{1}{2}} \gamma a^{-k/2} b^{-1} + a^2 b^{-1}).$$

<u>Proof</u>:

The L.H.S. of (A.3) is equal to:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} v_t \; x_{t+s} \quad \left[\frac{1}{\hat{\sigma}_{t+s}} - \frac{1}{\bar{\sigma}_{t+s}}\right] I_t I_{t+s}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} v_t \; x_{t+s} \quad \left[\frac{1}{\bar{\sigma}_{t+s}} - \frac{1}{\sigma_{t+s}}\right] I_t I_{t+s}.$$

The second term is a straightforward consequence of lemma 2 of chapter 3, and so it will be

$$O_p(a^2 b^{-1} + T^{-\frac{1}{2}} a^{-k} b^{-1}),$$

since, although we have the square roots of the variances and $\bar{\sigma}_{t+s}$, it does not make any difference by noting the usual trick of multiplying and dividing by, $\bar{\sigma}_{t+s}+\sigma_{t+s}$, and observing that $\bar{\sigma}_{t+s}>c>0$ by assumption.

Thus, it only remains to find the probability order of the first term of the above expression. Using the equality (see Robinson, 1987a):

$$x^{-1}=\sum_{q=0}^{r}(y-x)^{q}y^{-q-1}+(y-x)^{r+1}x^{-1}y^{-r-1}$$

we can rewrite this first term of the above term as:

$$\text{(A.4)}\qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{T-s} v_t\, x_{t+s}\left[\frac{(\bar{\sigma}_{t+s}-\hat{\sigma}_{t+s})}{\bar{\sigma}^2_{t+s}}\right]I_t I_{t+s}$$

$$\text{(A.5)}\qquad +\frac{1}{\sqrt{T}}\sum_{t=1}^{T-s} v_t\, x_{t+s}\left[\frac{(\bar{\sigma}_{t+s}-\hat{\sigma}_{t+s})^2}{\bar{\sigma}^2_{t+s}\,\hat{\sigma}_{t+s}}\right]I_t I_{t+s}.$$

- As far as (A.5) is concerned:

Appealing to lemma 6 of chapter 3 and using the argument of the above paragraph, we can say that this term is of probability order:

$$O_p(T^{-\frac{1}{2}}a^{-2k}b^{-1}).$$

- For (A.4), we have that it is equal to (multiplying and dividing by $\hat{\sigma}_{t+s}+\bar{\sigma}_{t+s}$):

$$\text{(A.4.1)} \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} \frac{v_t \, x_{t+s}}{\bar{\sigma}^2_{t+s}(\hat{\sigma}_{t+s}+\bar{\sigma}_{t+s})\hat{g}(x_t)} \Big[\frac{1}{Ta^k} \sum_{j=1}^{T} (\hat{u}_j^2-\sigma_j^2) \; K_{t+s,j}\Big] I_t I_{t+s}.$$

Although we have solved the problem of the square roots, we still have the problem of the denominator $(\hat{\sigma}_{t+s})$, which depends on all the observations and in particular on $u_1,\ldots,u_T$. Thus, in order to overcome this problem we will use the following device, since

$$\left[\frac{1}{\bar{\sigma}_{t+s}+\hat{\sigma}_{t+s}} - \frac{1}{2\bar{\sigma}_{t+s}}\right] - \left[\frac{(\hat{\sigma}_{t+s}-\bar{\sigma}_{t+s})}{2\bar{\sigma}_{t+s}(\bar{\sigma}_{t+s}+\hat{\sigma}_{t+s})}\right]$$

$$+ \left[\frac{(\hat{\sigma}^2_{t+s}-\bar{\sigma}^2_{t+s})}{2\bar{\sigma}_{t+s}(\bar{\sigma}_{t+s}+\hat{\sigma}_{t+s})^2}\right]$$

the above term (A.4.1) is equal to:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} \frac{v_t \, x_{t+s}}{\hat{g}(x_t) \; 2\bar{\sigma}^3_{t+s}} \Big[\frac{1}{Ta^k} \sum_{j=1}^{T} (\hat{u}_j^2-\sigma_j^2) \; K_{t+s,j}\Big] I_t I_{t+s} +$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-s} \frac{v_t \, x_{t+s}}{\hat{g}(x_t) \; \bar{\sigma}^2_{t+s}} \Big[\frac{1}{Ta^k} \sum_{j=1}^{T} (\hat{u}_j^2-\sigma_j^2) \; K_{t+s,j}\Big] \times$$

$$\left[\frac{1}{(\bar{\sigma}_{t+s}+\hat{\sigma}_{t+s})} - \frac{1}{2\bar{\sigma}_{t+s}}\right] I_t I_{t+s}.$$

By lemma 1, the first term is of probability order:

$$O_p(T^{-\gamma/2} a^{-k/2} b^{-1}),$$

and arguing as we did with (A.5), the second term can be shown to be of probability order:

$O_p(T^{-\frac{1}{2}}a^{-2k}b^{-1})$.

Q.E.D.

Lemma 4

Let v_t and x_t be two stationary stochastic processes as defined in lemma 2. Then:

$$\text{(A.6)} \quad \sum_{t=1}^{T-s} \sum_{j=t+1}^{t+s} \frac{(v_t v_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3} K_{j,t}\, I_t I_{t+s} = O_p(Ts^{\frac{1}{2}}a^k b^{-\frac{1}{2}}).$$

Proof:

The absolute value of the L.H.S. of equation (A.6) is less than or equal to:

$$\left| \sum_{t=1}^{T-s} \sum_{j=t+1}^{t+s} \frac{v_t v_{t+s}\ (v_j^2-1)\sigma_j^2}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3} K_{j,t}\, I_t I_{t+s} \right|$$

$$+ \left| \sum_{t=1}^{T-s} \sum_{j=t+1}^{t+s} \frac{\gamma(s)(v_j^2-1)\sigma_j^2}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3} K_{j,t}\, I_t I_{t+s} \right|.$$

About the second term it is very easy to check that it is the R.H.S. of (A.6) times $\gamma(s)$, since:

$$[\sum_j (v_j^2-1)]^2 = O_p(s) \quad \text{and} \quad E|K_{j,t}K_{i,t}| = O_p(a^{2k}).$$

Insofar as the first term is concerned:

$$\left|\sum_{t=1}^{T-s} v_t\left[\sum_{j=t+1}^{t+s}(v_j^2-1)\right] v_{t+s}\right| \leq$$

$$\sum_{t=1}^{T-s}(E|v_t|^3)^{1/3}(E|\sum_{j=t+1}^{t+s}(v_j^2-1)|^3)^{1/3}(E|v_{t+s}|^3)^{1/3}=O(Ts^{\frac{1}{2}})$$

and as long as $E|K_{j,t}K_{i,t}|=O(a^{2k})$ we conclude that the first term is also just the R.H.S. of equation (A.6).

Q.E.D.

Lemma 5

Under the same conditions given in lemma 3.

$$\text{(A.7)}\quad \sum_{t=1}^{T-|s|}\sum_{j=t+s+1}^{T}\frac{(v_tv_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3} K_{j,t}\ I_tI_{t+s}=O_p(Ts^{\frac{1}{2}}a^kb^{-\frac{1}{2}}).$$

Proof:

$$E[(A.7)^2]\leq$$

$$\text{(A.8)}\quad E\sum_{1\leq t+s\leq r+s<j\leq i\leq T}\frac{(v_tv_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2(v_rv_{r+s}-\gamma(s))(v_i^2-1)\sigma_i^2}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3\ \hat{g}^{\frac{1}{2}}(x_r)\bar{\sigma}_r^3}$$

$$\times\ [K_{j,t}\ K_{i,r}\ I_tI_{t+s}\ I_rI_{r+s}]+$$

$$\text{(A.9)}\quad E\sum_{1\leq t+s<j<r+s\leq T}\frac{(v_tv_{t+s}-\gamma(s))(v_j-1)\sigma_j(v_rv_{r+s}-\gamma(s))(v_i-1)\sigma_i}{\hat{g}^{\frac{1}{2}}(x_t)\bar{\sigma}_t^3\ \hat{g}^{\frac{1}{2}}(x_r)\bar{\sigma}_r^3}$$

$$\times\ [K_{j,t}\ K_{i,r}\ I_tI_{t+s}\ I_rI_{r+s}].$$

As regards to (A.8), it can be split up into two further terms:

$$\sum_{r-t>i-j} \quad + \sum_{r-t<i-j} .$$

The second term is the simplest one because the proof of its statistical order is the same as certain steps of the proof of Yoshihara's (1976) lemma 2, and it is of order:

$$T^2 a^{2k} b^{-1} \sum_k \zeta(k)^{\delta/\delta+2}(k+1) = T^{3-\gamma} a^{2k} b^{-1}$$

where $\zeta(k)$ is the absolutely regular coefficient.

As far as the first term is concerned, we can distinguish three different possibilities, i.e.:

a) $r-t<s$,

b) $s+1<r-t\leq 2s$

and

c) $2s<r-s$.

- About a), it is easy to show that it is

$$T^2 a^{2k} b^{-1} \sum_k^s \zeta(k)^{\delta/\delta+2}(k+1) = O(T^2 s^{1-\gamma} a^{2k} b^{-1})$$

since we consider u_t on the one hand and the remaining of the u's on the other.

- About b), we will have:

$$T^2 a^{2k} b^{-1} \sum_{k=s}^{2s} \zeta(s)^{\delta/\delta+2}(k+1) = O(T^2 s^2 \zeta(s)^{\delta/2+\delta} a^{2k} b^{-1}).$$

- About c), this last term will be of order:

$$T^2 a^{2k} b^{-1} \sum_{2s+1} \zeta(s+k)^{\delta/\delta+2}(2s+k) = O[T^{3-\gamma}+T^2 s] a^{2k} b^{-1}).$$

Thus, we have that the term (A.8) is:

$$[T^{3-\gamma}+T^2 s] a^{2k} b^{-1}.$$

Therefore it only remains to check (A.9) to conclude the lemma. But arguing in the same way as in (A.8) we can conclude that it is of the same probability order.

Q.E.D.

<u>Lemma 6</u>

Under the same conditions of lemma 2 and C5, C6, we have that:

$$\text{(A.10)} \qquad \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} (v_t v_{t+s} - \gamma(s)) \quad (\bar{\sigma}_t^{-1} - \hat{\sigma}_t^{-1}) I_t I_{t+s} = o_p(1).$$

<u>Proof</u>:

The L.H.S. of the above equation can be split up into three terms (following the same steps as in lemma 2):

$$\text{(A.11)} \quad \frac{1}{2}\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} \frac{(v_t v_{t+s} - \gamma(s))}{\bar{\sigma}_t^3 \, \hat{g}(x_t)} I_t I_{t+s}$$

$$\times \left[\frac{1}{Ta^k}\sum_{j=1}^{T}(\hat{u}_j^2-\sigma_j^2)\ K_{j,t}\right]$$

$$\text{(A.12)} + \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|s|}\frac{(v_t v_{t+s}-\gamma(s))}{\bar{\sigma}_t^3\ \hat{g}(x_t)[\hat{\sigma}_t-\bar{\sigma}_t]} I_t I_{t+s}$$

$$\times \left[\frac{1}{Ta^k}\sum_{j=1}^{T}(\hat{u}_j^2-\sigma_j^2)\ K_{j,t}\right]^2$$

$$\text{(A.13)} + \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|s|}\frac{(v_t v_{t+s}-\gamma(s))}{\bar{\sigma}_t^3\ \hat{\sigma}_t\ \hat{g}(x_t)} I_t I_{t+s}$$

$$\times \left[\left[\frac{1}{Ta^k}\sum_{j=1}^{T}\hat{u}_j^2\ K_{j,t}\right]^{\frac{1}{2}} - \left[\frac{1}{Ta^k}\sum_{j=1}^{T}\sigma_j^2\ K_{j,t}\right]^{\frac{1}{2}}\right]^2.$$

(A.12) and (A.13) are both $o_p(1)$, by arguing as we did with (A.5) for instance, in fact they will be:

$$O_p(T^{-1}a^{-2k}M\ b^{-1}).$$

The cumbersome term will be the first one, that is (A.11), but this term is equal to:

$$\text{(A.11.1)} \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\ \frac{1}{T}\sum_{t=1}^{T-|s|}\frac{(v_t v_{t+s}-\gamma(s))}{\bar{\sigma}_t^3\ \hat{g}(x_t)} I_t I_{t+s} +$$

$$\times \left[\frac{1}{Ta^k}\sum_{j=1}^{T}(u_j^2-\sigma_j^2)\ K_{j,t}\right] +$$

$$(A.11.2)\quad (\beta-\beta^{o.l.s.})' \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \sum_{t=1}^{T-|s|} \frac{(v_t v_{t+s}-\gamma(s))}{\bar{\sigma}_t^3\, \hat{g}(x_t)} I_t I_{t+s}$$

$$\times\, [\frac{1}{Ta^k} \sum_{j=1}^{T} v_j \sigma_j x_j\; K_{j,t}] +$$

$$(A.11.3)\quad (\beta-\beta^{o.l.s.})' \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \sum_{t=1}^{T-|s|} \frac{(v_t v_{t+s}-\gamma(s))}{\bar{\sigma}_t^3\, \hat{g}(x_t)} I_t I_{t+s}$$

$$\times\, [\frac{1}{Ta^k} \sum_{j=1}^{T} x_j\; x_j'\; K_{j,t}]\; (\beta-\beta^{o.l.s.}).$$

But (A.11.2) and (A.11.3) are both $O_p(T^{-1}Mb^{-\frac{1}{2}})$, and about (A.11.1), this is equal to

$$\frac{1}{2T^2a^k} \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \sum_{t=1}^{T-|s|} \sum_{j=1}^{t} \frac{(v_t v_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2}{\bar{\sigma}_t^3\, \hat{g}(x_t)^{\frac{1}{2}}} K_{j,t}\, I_t I_{t+s} +$$

$$\frac{1}{2T^2a^k} \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \sum_{t=1}^{T-|s|} \sum_{j=t+s+1}^{T} \frac{(v_t v_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2}{\bar{\sigma}_t^3\, \hat{g}^{\frac{1}{2}}(x_t)} K_{j,t}\, I_t I_{t+s} +$$

$$\frac{1}{2T^2a^k} \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \sum_{t=1}^{T-|s|} \sum_{j=t+1}^{t+|s|} \frac{(v_t v_{t+s}-\gamma(s))(v_j^2-1)\sigma_j^2}{\bar{\sigma}_t^3\, \hat{g}^{\frac{1}{2}}(x_t)} K_{j,t}\, I_t I_{t+s}.$$

Each term is by lemmas 4 or 5 of probability order:

$$O_p(T^{-1}M^{1\frac{1}{2}}b^{-\frac{1}{2}}).$$

Q.E.D.

Lemma 7

Under the assumptions given in lemma 5 and C6, C7, we have that:

$$(A.14)\ \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\ \frac{1}{T} \sum_{t=1}^{T-|s|} (v_t v_{t+s} - \gamma(s))\sigma_t(\bar{\sigma}_t^{-1} - \sigma_t^{-1})\ I_t I_{t+s} = o_p(1).$$

Proof:

By Parzen (1957) theorem 5A, conditional on x_t, (A.14) is $O_p(T^{-1}M)$.

Also we have that:

$$E[(\bar{\sigma}_t^{-1} - \sigma_t^{-1})(\bar{\sigma}_r^{-1} - \sigma_r^{-1}) I_t I_r] \leq$$

$$\left[E[\bar{\sigma}_t^{-1} - \sigma_t^{-1}]^2 I_t\right]^{\frac{1}{2}} \left[E[\bar{\sigma}_r^{-1} - \sigma_r^{-1}]^2 I_r\right]^{\frac{1}{2}}.$$

Recalling that:

$$[(\bar{\sigma}_t^{-1} - \sigma_t^{-1})] I_t = \bar{\sigma}_t^{-1} \sigma_t^{-1} (\bar{\sigma}_t + \sigma_t)^{-1} (\sigma_t^2 - \bar{\sigma}_t^2) I_t$$

we have, by lemma 6 of the previous chapter, that the supremum of the above equation is of probability order:

$$O_p(a^4 b^{-2} + T^{-1} a^{-2k} b^{-2})$$

and thus we conclude the lemma.

Q.E.D.

Lemma 8

Under the same conditions given in the previous lemma we have that:

$$\text{(A.15)} \qquad \| \tilde{f}_{\hat{v}\hat{v}} - \tilde{f}_{\overline{vv}I} \|^2 = O_p(T^{-2}M^3a^{-k}b^{-2} + T^{-2}a^{-4k}M^2b^{-2} + a^4b^{-2}).$$

Proof:

$$\tilde{f}_{\hat{v}\hat{v}} - \tilde{f}_{\overline{vv}I} = \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} \hat{u}_t \hat{u}_{t+s}\left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t} - \frac{1}{\sigma_{t+s}\sigma_t}\right] I_t I_{t+s}.$$

On the other hand, the difference between this and the expression with u_t is of order less than $T^{-\frac{1}{2}}$. This can be seen very easily. Hence we should work out:

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} u_t u_{t+s}\left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t} - \frac{1}{\sigma_{t+s}\sigma_t}\right] I_t I_{t+s}$$

which is equal to the sum of the following three terms:

$$\text{(A.16)} \qquad \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} v_t v_{t+s} \sigma_{t+s}\left[\frac{1}{\hat{\sigma}_{t+s}} - \frac{1}{\sigma_{t+s}}\right] I_t I_{t+s} +$$

$$\text{(A.17)} \qquad \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} v_t v_{t+s} \sigma_t\left[\frac{1}{\hat{\sigma}_t} - \frac{1}{\sigma_t}\right] I_t I_{t+s} +$$

$$\text{(A.18)} \qquad \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} u_t u_{t+s} \left[\frac{1}{\hat{\sigma}_t} - \frac{1}{\sigma_t}\right]\left[\frac{1}{\hat{\sigma}_{t+s}} - \frac{1}{\sigma_{t+s}}\right] I_t I_{t+s}.$$

Therefore, the expectation of (A.15) is less than or equal to:

$O_p(T^{-1})+4\{\|A.16\|^2+\|A.17\|^2+\|A.18\|^2\}$.

- About (A.17). This term is equal to:

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right]e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|}(v_t\, v_{t+s}-\gamma(s))\,\sigma_t\left[\frac{1}{\hat{\sigma}_t}-\frac{1}{\sigma_t}\right]I_tI_{t+s}$$

$$+\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right]e^{is\lambda}\gamma(s)\frac{1}{T}\sum_{t=1}^{T-|s|}\sigma_t\left[\frac{1}{\hat{\sigma}_t}-\frac{1}{\sigma_t}\right]I_tI_{t+s}.$$

But by lemmas 6 and 7, the first term of $\|A.17\|^2$ is of probability order:

$$O_p(T^{-1}M\ a^4b^{-2}+T^{-2}M^2a^{-4k}b^{-4}+T^{-2}M^3a^{-k}b^{-2})$$

while the second term is of probability order:

$$O_p(T^{-1}a^{-2k}b^{-2}+a^4b^{-2})$$

since the function K_1 is uniformly bounded and $\gamma(s)$ is absolutely summable. The term $\|A.16\|^2$ is identical, and $\|A.18\|^2$ works in the same fashion.

Q.E.D.

The next lemma is very important since it allows us to say that:

$$\left\|\tilde{f}^{-1}_{\hat{v}\hat{v}}(\lambda)-\hat{f}^{-1}_{\bar{v}\bar{v}}(\lambda)\right\|^2< C\ \|\tilde{f}_{\hat{v}\hat{v}}(\lambda)-\hat{f}_{\bar{v}\bar{v}}(\lambda)\|^2$$

as T goes to infinity, where C is a generic constant.

It is very well known, see Hannan (1961) (since $|\alpha_j|=O(a^{-q-1})$ and

$q>\frac{1}{2}$), that

$$\max_j | \hat{f}_{\overline{vv}}(\lambda_j) - f(\lambda_j) | \text{ goes to zero in probability.}$$

Thus, we only need to establish that:

$$\max_j | \tilde{f}_{\hat{v}\hat{v}}(\lambda_j) - \hat{f}_{\overline{vv}}(\lambda_j) | \text{ goes to zero in probability.}$$

<u>Lemma 9</u>

Under the same conditions of lemma 6, we have that:

$$\text{(A.19)} \qquad \max_\lambda | \tilde{f}_{\hat{v}\hat{v}}(\lambda) - \hat{f}_{\overline{vv}}(\lambda) | = o_p(1).$$

<u>Proof</u>:

We have that the L.H.S. of equation (A.19) is equal to:

$$\text{(A.20)} \qquad \max_\lambda \left| \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} \left[\frac{\hat{u}_t \hat{u}_{t+s}}{\hat{\sigma}_{t+s}\hat{\sigma}_t} I_t I_{t+s} - \frac{u_t u_{t+s}}{\sigma_{t+s}\sigma_t} \right] \right|.$$

But the term inside the square brackets is equal to:

$$\hat{u}_t \hat{u}_{t+s} \left[\frac{I_t I_{t+s}}{\hat{\sigma}_t \hat{\sigma}_{t+s}} - \frac{I_t I_{t+s}}{\sigma_t \sigma_{t+s}} \right] - \left[\frac{\hat{u}_t \hat{u}_{t+s} - u_t u_{t+s}}{\sigma_t \, \sigma_{t+s}} \right] I_t I_{t+s} +$$

$$\{I_t I_{t+s}\} \frac{u_t u_{t+s}}{\sigma_t \sigma_{t+s}}$$

and hence, the term inside the vertical bars of (A.20) is equal to:

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda} \frac{1}{T} \sum_{t=1}^{T-|s|} u_t u_{t+s} \left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t} - \frac{1}{\sigma_{t+s}\sigma_t} \right] I_t I_{t+s} +$$

$$(\beta-\beta^{o.l.s.})'\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} u_t x_{t+s}\left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t}-\frac{1}{\sigma_{t+s}\sigma_t}\right] I_t I_{t+s}+$$

$$(\beta-\beta^{o.l.s.})'\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} u_{t+s} x_t\left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t}-\frac{1}{\sigma_{t+s}\sigma_t}\right] I_t I_{t+s}+$$

$$(\beta-\beta^{o.l.s.})'\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} x_t x'_{t+s}\left[\frac{1}{\hat{\sigma}_{t+s}\hat{\sigma}_t}-\frac{1}{\sigma_{t+s}\sigma_t}\right] I_t I_{t+s}(\beta-\beta^{o.l.s.})$$

$$-\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} u_t x'_{t+s}\left[\frac{1}{\sigma_{t+s}\sigma_t}\right](\beta-\beta^{o.l.s.})-$$

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} u_{t+s} x'_t\left[\frac{1}{\sigma_{t+s}\sigma_t}\right](\beta-\beta^{o.l.s.})-$$

$$(\beta-\beta^{o.l.s.})'\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|} x_t x'_{t+s}\left[\frac{1}{\sigma_{t+s}\sigma_t}\right](\beta-\beta^{o.l.s.})+$$

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|}\left[\frac{u_t u_{t+s}}{\sigma_{t+s}\sigma_t}\right](1-I_t I_{t+s}).$$

Since by assumption the function K_1 is uniformly bounded, then $\sum|K_1(^s/_M)|=O(M)$. Also since the O.L.S. estimator of the parameters β is root-T consistent, it is very easy to show that the supremum in λ of the second four terms is $o_p(1)$.

As regards the last term, it is also very easy to show that it is $o_p(1)$ since adding and substracting $\gamma(s)$ into the term in brackets, we have that as long as

$$T^{-1}\sum\{1-I_tI_{t+s}\}=o_p(1)$$

then also this term is $o_p(1)$. Therefore, in order to prove the lemma, it only remains to show that the first term of the above expression is also $o_p(1)$. Note that this term is equal to:

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|}(v_tv_{t+s}\sigma_t)(\hat{\sigma}_{t+s}^{-1}-\sigma_{t+s}^{-1})I_tI_{t+s}+$$

$$\text{(A.21)}\quad \sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|}(v_tv_{t+s}\sigma_{t+s})(\hat{\sigma}_{t}^{-1}-\sigma_{t}^{-1})I_tI_{t+s}+$$

$$\sum_{-M+1}^{M} K_1\left[\frac{s}{M}\right] e^{is\lambda}\frac{1}{T}\sum_{t=1}^{T-|s|}(v_tv_{t+s}\sigma_t\sigma_{t+s})(\hat{\sigma}_{t+s}^{-1}-\sigma_{t+s}^{-1})(\hat{\sigma}_{t}^{-1}-\sigma_{t}^{-1})I_tI_{t+s}$$

and taking into account that:

$$\frac{1}{T}\sum_{t=1}^{T-|s|}(v_tv_{t+s}\sigma_t)(\hat{\sigma}_{t+s}^{-1}-\sigma_{t+s}^{-1})I_tI_{t+s}$$

$$=\frac{1}{T}\sum_{t=1}^{T-|s|}(v_tv_{t+s}-\gamma(s))\sigma_t(\hat{\sigma}_{t+s}^{-1}-\sigma_{t+s}^{-1})I_tI_{t+s}$$

$$+\frac{1}{T}\sum_{t=1}^{T-|s|}\gamma(s)\sigma_t(\hat{\sigma}_{t+s}^{-1}-\sigma_{t+s}^{-1})I_tI_{t+s}$$

it is very easy to show that it is $o_p(T^{-\frac{1}{2}})+o_p(1)\gamma(s)$. Hence, when the supremum in λ is taken, the first term of (A.21) will be $o_p(1)$. As regards the other two terms of (A.21), they are also $o_p(1)$ arguing in the same way.

Q.E.D.

CHAPTER 6

A FINITE SAMPLE STUDY OF THE ADAPTIVE ESTIMATOR OF CHAPTER 5

The aim of this chapter is to give some evidence about the performance in finite samples of the semiparametric estimator which we have studied in the previous chapter, and also to compare it with the G.L.S. estimator when the conditional variances are known. Also it would be interesting to observe how inaccurate is the G.L.S. estimator when we mis-specify(MIS) the heteroscedasticity.

In this study the semiparametric estimator involves two bandwidth parameters, i.e. M and **a**, which, as we have already been seen in section 5.3 of the previous chapter, are directly related via the assumption C5.

As it was pointed out in section 5.3 of chapter 5, Robinson (1988a) allowed the M parameter to be data-dependent, in an homoscedastic framework, without losing any of the good statistical properties of the estimated regression parameters. He pointed out that in a heteroscedastic regression model setting, the bandwidth parameter **a** can be chosen to maximize a pseudo-Gaussian likelihood allowing a to be data-dependent and establishing asymptotic normality and efficiency. However, in this case it was decided to choose a set of representatives M and a's.

Four different models, given by equation (1.1) of chapter 5, were simulated with $k=1$, where x_t and v_t are both Gaussian zero mean process $x_t=\omega_t+\tau\omega_{t-1}$, $v_t=\xi_t+\theta\xi_{t-1}$, ξ_t and ω_t are white noise having unit variances. The value of τ was taken as 0.5. The four models can be distinguished for both the parameter θ and the heteroscedasticity function. We have considered:

<u>Model</u> 1: $\theta=0.5$ $\sigma(x_t)=\exp(0.125+0.4x_t)$,

<u>Model</u> 2: $\theta=0.5$ $\sigma(x_t)=\exp(0.25+0.8x_t)$,

<u>Model</u> 3: $\theta=0.9$ $\sigma(x_t)=\exp(0.125+0.4x_t)$,

<u>Model</u> 4: $\theta=0.9$ $\sigma(x_t)=\exp(0.25+0.8x_t)$

although we are going to present only the results for the first two models since the findings for models 3 and 4 are in the same spirit.

A decision was made to choose $\theta=0.5$ or 0.9, since they were used by other investigators, e.g. Robinson (1988a). Also, two heteroscedasticity models were used. The severity of the heteroscedasticity will be measured by the coefficient of variation of $\sigma^2(x_t)$, defined as the ratio of its standard deviation to its mean. The coefficient of variation for the model $\sigma(x_t)=\exp(.125+0.4x_t)$ is equal to 1.05, while for the second one it is 2.84.

The simulation study was performed using the two different methods for the estimation of the parameter β as described in section 2.2.2 of chapter 2. By means of method 1, one could understand the procedure described in chapter 5, i.e. via smoothing the spectral density estimates. A second method of estimation (method 2) makes use of the periodograms, as in Robinson (1973,1976,1988a) and Hannan (1971). For a complete description of this two methods see for instance Robinson (1988a) pp. 8-9 and section 2.2.2 of chapter 2. The only difference is that before getting the discrete Fourier transform (DTF) of y_t and x_t, these random variables are divided by $\sigma(x_t)^*$, where $\sigma(x_t)^*$ will be an estimate of the conditional variance, e.g. $\sigma(x_t)^*$ will be the true ones, the nonparametric estimate defined by equation (2.1) of chapter 5, or a parametric estimate of the conditional variance when we believe that it follows a polynomial function.

Although, only the theorem of chapter 5 was shown for method 1, i.e. via smoothing the spectral densities, we see no reason why this theorem will not hold under method 2, since for the parametric case the difference between both methods are negligible in large samples, in fact they converge to the same asymptotic distribution function. However, it was decided to use this method in this simulation study since, the results of Robinson's (1988a) study could be used to guess the possible grid of M's for this particular setting. As a result, a comparison could be made between both methods of estimation for finite samples.

In order to see what could happen when the heteroscedasticity function was mis-specified, a pseudo G.L.S. estimate of β was computed when the $\sigma(x_t)$ are estimated from the regression equation:

$$\hat{u}_t^2=\gamma_1+\gamma_2 x_t+\gamma_3 x_t^2$$

in the usual fashion.

Two more points have to be mentioned. The first one is that no attempt was made to get the G.L.S. estimate of β based on parametric modelling of the disturbances, because the real aim of this Monte-Carlo experiment is to know how good the adaptive estimator is, especially against the heteroscedasticity, and because several Monte-Carlo studies of Hannan's estimate have been reported, e.g. Duncan and Jones (1966), Engle and Gardner (1976), Robinson (1976). The second point is that different windows were used for these methods. In method 1 a Hanning-Tukey type of kernel given by equation (2.5) of chapter 2 was used, while in the weighted periodogram the Parzen window was used, given by equation (2.6) of chapter 2 due to its computational advantage, and also because the Hanning-Tukey type kernel function has a spectral window which is not positive.

All computations were carried out in double precision FORTRAN on the University of London's Amdahl computer, using a random number generator and fast Fourier transform (FFT) algorithm from the NAG Library. For each of the four models, 1000 bivariate time series u_t and x_t of lengths $T=2^7$ and 2^8 were generated and 500 for $T=2^9$.

In what follows, method 1 stands for the estimate of β via smoothing the spectrum, while method 2 stands for the weighted periodogram. In tables 6.1-6.6 only the M.S.E. for the parametric, mis-specified and semiparametric was reported, for different M's and a's. Also the M.S.E. for the O.L.S. estimate was reported. The reason, why only the M.S.E. was shown, is that the findings of this study about the bias are exactly the same as those of Robinson (1988a), i.e. the bias decreases as the sample size increases owing perhaps to the "imperfection" in the random generator whereby the "normal" deviates were not quite symmetric.

As it is impossible to give a closed formula for the asymptotic covariance matrix of the G.L.S. estimator of β, in terms of τ and θ, as it will be the case in the absence of heteroscedasticity, it was decided to give an estimate of this covariance matrix by the M.S.E. of the G.L.S. estimate of β. The values that we have found were: for model 1 and $T=128$ $\zeta=.00474$, for $T=256$ $\zeta=0.00195$ and for $T=512$ $\zeta=0.00098$. For the second model and $T=128$ $\zeta=.00115$, for $T=256$ $\zeta=.00039$ and for $T=512$ $\zeta=.000172$.

TABLE 6.1

M.S.E. for model 1 with sample size T=128, and 1000 replicates.

O.L.S. 0.0360

Method 1.-

	M=1	M=3	M=5	M=7	M=9
G.L.S.	0.0060	0.0059	0.0051	0.0050	0.0050
a=.4	0.0083	0.0082	0.0075	0.0075	0.0076
a=.7	0.0084	0.0083	0.0076	0.0076	0.0077
a=1.0	0.0097	0.0097	0.0088	0.0088	0.0090
MIS	0.0189	0.0189	0.0175	0.0170	0.0168

Method 2.-

	M=3.69	M=4.75	M=5.80	M=6.86	M=7.91
G.L.S.	0.0050	0.0048	0.0048	0.0048	0.0048
a=.4	0.0070	0.0069	0.0068	0.0069	0.0069
a=.7	0.0070	0.0068	0.0068	0.0068	0.0069
a=1.0	0.0081	0.0079	0.0079	0.0080	0.0080
MIS	0.0174	0.0172	0.0171	0.0171	0.0171

TABLE 6.2

M.S.E. for model 1 with sample size T=256, and 1000 replicates.

O.L.S.	0.0175

Method 1.-

	M=3	M=6	M=9	M=12	M=15
G.L.S.	0.0026	0.0021	0.0020	0.0020	0.0020
a=.3	0.0034	0.0030	0.0029	0.0029	0.0030
a=.5	0.0032	0.0027	0.0027	0.0027	0.0028
a=.7	0.0033	0.0028	0.0028	0.0028	0.0029
MIS	0.0087	0.0080	0.0079	0.0078	0.0077

Method 2.-

	M=4.24	M=5.45	M=6.67	M=7.88	M=9.09
G.L.S.	0.0020	0.0020	0.0020	0.0020	0.0020
a=.3	0.0028	0.0027	0.0027	0.0027	0.0027
a=.5	0.0026	0.0025	0.0025	0.0025	0.0025
a=.7	0.0027	0.0026	0.0026	0.0026	0.0026
MIS	0.0080	0.0079	0.0079	0.0079	0.0079

TABLE 6.3

M.S.E. for model 1 with sample size T=512, and 500 replicates.

O.L.S. 0.00874

Method 1.-

	M=5	M=8	M=11	M=14	M=17
G.L.S.	0.00102	0.00102	0.00102	0.00104	0.00109
a=.2	0.00151	0.00150	0.00150	0.00151	0.00154
a=.4	0.00125	0.00124	0.00125	0.00125	0.00129
a=.6	0.00125	0.00124	0.00124	0.00126	0.00129
MIS	0.00335	0.00336	0.00336	0.00338	0.00348

Method 2.-

	M=4.87	M=6.27	M=7.66	M=9.05	M=10.45
G.L.S.	0.00101	0.00100	0.00099	0.00099	0.00099
a=.2	0.00146	0.00144	0.00143	0.00143	0.00143
a=.4	0.00120	0.00119	0.00118	0.00118	0.00112
a=.6	0.00120	0.00119	0.00119	0.00119	0.00119
MIS	0.00344	0.00342	0.00342	0.00342	0.00343

TABLE 6.4

M.S.E. for model 2 with sample size T=128, and 1000 replicates.

O.L.S. 0.3105

Method 1.-

	M=1	M=3	M=5	M=7	M=9
G.L.S.	0.0016	0.0016	0.0013	0.0013	0.0013
a=.4	0.0073	0.0079	0.0080	0.0086	0.0092
a=.7	0.0070	0.0078	0.0079	0.0084	0.0090
a=1.0	0.0092	0.0102	0.0103	0.0108	0.0116
MIS	0.0196	0.0200	0.0188	0.0185	0.0182

Method 2.-

	M=3.69	M=4.75	M=5.80	M=6.86	M=7.91
G.L.S.	0.0013	0.0012	0.0012	0.0012	0.0012
a=.4	0.0059	0.0058	0.0058	0.0059	0.0060
a=.7	0.0056	0.0055	0.0056	0.0056	0.0057
a=1.0	0.0074	0.0073	0.0073	0.0073	0.0074
MIS	0.0186	0.0186	0.0185	0.0185	0.0186

TABLE 6.5

M.S.E. for model 2 with sample size T=256, and 1000 replicates.

O.L.S. 0.1596

Method 1.-

	M=3	M=6	M=9	M=12	M=15
G.L.S.	0.0005	0.0005	0.0004	0.0004	0.0005
a=.3	0.0023	0.0023	0.0025	0.0027	0.0030
a=.5	0.0020	0.0020	0.0022	0.0024	0.0027
a=.7	0.0021	0.0021	0.0022	0.0024	0.0027
MIS	0.0089	0.0080	0.0078	0.00768	0.0076

Method 2.-

	M=4.24	M=5.45	M=6.67	M=7.88	M=9.09
G.L.S.	0.0004	0.0004	0.0004	0.0004	0.0004
a=.3	0.0018	0.0018	0.0018	0.0019	0.0019
a=.5	0.0015	0.0015	0.0016	0.0016	0.0016
a=.7	0.0015	0.0016	0.0016	0.0016	0.0016
MIS	0.0079	0.0078	0.0078	0.0078	0.0078

TABLE 6.6

M.S.E. for model 2 with sample size T=512, and 500 replicates.

O.L.S. 0.0789

Method 1.-

	M=5	M=8	M=11	M=14	M=17
G.L.S.	0.00019	0.00019	0.00019	0.00020	0.00021
a=.2	0.00089	0.00085	0.00081	0.00077	0.00074
a=.4	0.00077	0.00073	0.00069	0.00066	0.00062
a=.6	0.00075	0.00071	0.00067	0.00063	0.00059
MIS	0.00366	0.00367	0.00371	0.00377	0.00392

Method 2.-

	M=4.87	M=6.27	M=7.66	M=9.05	M=10.45
G.L.S.	0.00018	0.00018	0.00018	0.00018	0.00018
a=.2	0.00064	0.00065	0.00065	0.00066	0.00067
a=.4	0.00054	0.00054	0.00055	0.00056	0.00057
a=.6	0.00050	0.00051	0.00052	0.00053	0.00054
MIS	0.00372	0.00372	0.00373	0.00374	0.00375

From the above tables it can be seen that both methods of estimating the parameter β perform very similar, although method 2 seems to perform slightly better. However, it seems that the performance of the weighted periodogram is better across both models and both sample sizes when the $\sigma^2(x_t)$ is estimated nonparametrically. Two additional features are the "stability" of the M.S.E. across the a's and also the M's, and the similarity between the G.L.S., MIS and the semiparametric estimator. The second characteristic is that the "best" M for the G.L.S. and the semiparametric estimator are the same, as it would be expected from the conditions of the theorem since in this example k=1. Also O.L.S. performs very badly in the presence of both serial correlation and heteroscedasticity, the latter perhaps being the major cause of its poor performance.

As expected, it was observed that as the severity of the heteroscedasticity increases, the performance of the mis-specified estimator under both methods of estimation is worse compared with the parametric and semiparametric estimators.

Also, it turned out that the "best" a in terms of the M.S.E. of the estimator of β is the same for both methods of estimation, and the smooth parameter M seems to be in the same "region" for both methods of estimation.

Also, as expected the performance of the semiparametric estimator of β, for both methods 1 and 2, worsens as the severity of the heteroscedasticity increases. It was observed that in this case a greater amount of data was needed to adapt to the unknown heteroscedasticity, owing to the fact that the estimation of the residuals via the O.L.S. estimator is poorer, and the estimator (2.1) used in this study may be more sensitive to this O.L.S. residuals than for the estimation of the spectral densities or periodograms.

REFERENCES

- AMEMIYA, T. (1973a): "Regression Analysis when the Variance of the Dependent Variable is Proportional to the Square of its Expectation". Journal of the American Statistical Association, 68, 928-934.

- AMEMIYA, T. (1973b): "Generalized Least Squares with an Estimated Autocovariance Matrix". Econometrica, 41, 723-732.

- AMEMIYA, T. (1973c): "Regression Analysis when the Dependent Variable is Truncated Normal". Econometrica, 41, 997-1016.

- BARLETT, M.S. (1963): "Statistical Estimation of Density Functions". Sankhya, Ser. A, 25, 145-154.

- BEACH, C.M. AND MACKINNON, J.G. (1978a): "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors". Econometrica, 46, 51-58.

- BEACH, C.M. AND MACKINNON, J.G. (1978b): "Full Maximum Likelihood Estimation of Second Order Autoregressive

Error Models". Journal of Econometrics, 7, 187-198.

- BEGUN, J.M., HALL, W.J., HUANG, W.M. AND WELLNER, J.A. (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models". Annals of Statistics, 11, 432-452.

- BELTRAO, K.I. AND BLOOMFIELD, P. (1987): "Determining the Bandwidth of a Kernel Spectrum Estimate". Journal of Time Series Analysis, 8, 21-38.

- BICKEL, P. (1982): "On Adaptive Estimation". Annals of Statistics, 10, 647-671.

- BOX, G.E.P. AND JENKINS, G.M. (1976): Time Series Analysis: Forecasting and Control, revised version. Holden-Day, San Francisco.

- BROWN, B.M. (1971): "Martingale Central Limit Theorems". Annals of Mathematical Statistics, 42, 59-66.

- CARROLL, R.D. (1982): "Adapting for Heteroscedasticity in Linear Models". Annals of Statistics, 10, 1224-1233.

- CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions". Journal of Econometrics, 34, 305-324.

- CHAN, K.S. AND TONG, H. (1986): "On Estimating Thresholds in

Autoregressive Models". Journal of Time Series Analysis, 7, 179-190.

- CHOW, Y.S. (1971): "On the L_p-Convergence for $n^{-\frac{1}{2}} S_n$, $0<p<2$ " The Annals of Mathematical Statistics, 42 , 393-394.

- COCHRANE, D. AND ORCUTT, G.H. (1949): "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms". Journal of the American Statistical Association, 44, 32-61.

- COLLOMB, G. (1985): "Nonparametric Time Series Analysis and Predictions: Uniform Almost Sure Convergence of the Window and K-NN Autoregression Estimates". Mathematische Operationsforschung und Statistik. Statistics, 16, 297-307.

- COGBURN, R. AND DAVIS, H.T. (1974): "Periodic Splines and Spectral Estimation". Annals of Statistics, 2, 1108-1126.

- DAVIES, N. PETRUCCELLI, J.D. AND WATSON, S.M. (1987): "Non-linearity, Non-Normality and Outliers in Time Series Analysis". Preprint, University of Trent.

- DAVYDOV, Y.A. (1968): "Convergence of Distributions Generated by Stationary Stochastic Processes". Theory of Probability and its Applications, 13, 691-696.

- DELGADO, M.A. (1989): "Semiparametric Estimation in Cross-sectional Heteroscedasticity Econometric Models". Ph.D. Thesis. London School of Economics.

- DENKER, M. AND KELLER, G. (1983): "On U-statistics and V. Mises' Statistics for Weakly Dependent Processes". Zeitschrift für Wakrscheinlichkeitstheorie und Verwandte Gebiete, 64, 505-522.

- DOUKHAN, P. AND GHINDES, M. (1980): "Estimations dans le Processus $\langle X_{n+1}=f(X_n)+\epsilon_n \rangle$". Comptes Rendus de L'Academie des Sciences. Paris, 291, 61-64.

- DUNCAN, R.B. AND JONES, R. (1966): "Multiple Regression with Stationary Errors". Journal of the American Statistical Association, 61, 917-923.

- DURBIN, J. (1960): "Estimation of Parameters in Time Series Regression Models". Journal of the Royal Statistical Society., Ser. B, 22, 139-153.

- EICKER, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions". Annals of Mathematical Statistics, 34, 447-456.

- EICKER, F. (1967): "Limit Theorems for Regression with Unequal and Dependent Errors". Procedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1 (Berkeley: Univerisity of California Press), 59-82.

- ENGLE, R.F. (1974): "Specification of the Disturbance for Efficient Estimation". Econometrica, 42, 135-147.

- ENGLE, R.F. AND GARDNER, R. (1976): "Some finite Sample Properties of Spectral Estimators of a Linear Regression". Econometrica, 44, 149-165.

- EPANECHNIKOV, V.A. (1969): "Non-Parametric Estimation of a Multivariate Probability Density". Theory of Probability and its Applications, 14, 153-158.

- FAIR, R.C. (1973): "A Comparison of Alternative Estimators of Macroeconomics Models". International Economic Review, 4, 261-277.

- FEIGIN, P.D. AND TWEEDIE, R.L. (1985): "Random Coefficient Autoregressive Process. A Markov Chain Analysis of Stationary and Finiteness of Moments". Journal of Time Series Analysis, 6, 1-14.

- FULLER, W.A. (1976): Introduction to Statistical Time Series. New York: Wiley.

- GASTWIRTH, J.L. AND RUBIN, H. (1975): "The Asymptotic Distribution Theory of the Empiric CDF for Mixing Stochastic Processes." Annals of Statistics, 3, 809-824.

- GRANGER, C.W.J. AND ANDERSEN, A.P. (1978): "An Introduction to Bilinear Time Series Analysis". Vandenhoek and Ruprecht. Gottingen.

- GRENANDER, U. (1954): "On the Estimation of Regression Coefficients in the case of Autocorrelated Disturbance". <u>Annals of Mathematical Statistics</u>, <u>25</u>, 252-272.

- HANNAN, E.J. (1961): "Testing for a Jump in the Spectral Function". <u>Journal of the Royal Statistical Society</u>, Ser.B , <u>23</u>, 394-404.

- HANNAN, E.J. (1963): "Regression for Time Series". In <u>Time Series Analysis</u> ed. by M. Rosenblatt. New York: Wiley, 17-37.

- HANNAN, E.J. (1970): <u>Multiple Time Series Analysis</u>. New York: John Wiley.

- HANNAN, E.J. (1971): "Non-linear Time Series Regression". <u>Journal of Applied Probability</u>, <u>8</u>, 767-780.

- HANNAN, E.J. AND ROBINSON, P.M. (1973): "Lagged Regression with Unknown Lags". <u>Journal of the Royal Statistical Society</u>, Series B, <u>35</u>, 252-267.

- HANNAN, E.J. AND TERRELL, R.D. (1973): "Multiple Equation Systems with Stationary Errors". <u>Econometrica</u>, <u>41</u>, 299-320.

- HANSEN, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators". Econometrica, 50, 1029-1054.

- HARDLE, W. AND STOKER, T.M. (1987): "Investigating Smooth Multiple Regression by the Method of Average Derivatives". Preprint. Universität of Bonn.

- HARRISON, M.J. AND McCABE, B.P.M. (1975): "Autocorelation with Heteroscedasticity: A Note on the Robustness of Durbin-Watson, Geary and Henshaw Tests". Biometrika, 62, 214-216.

- HARVEY, A.C. (1981): The Econometric Analysis of Time Series. Phillip Allan.

- HARVEY, A.C. AND PHILLIPS, G.P.A. (1979): "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbance". Biometrika, 66, 49-58.

- HARVEY, A.C. AND ROBINSON, P.M. (1988): "Efficient Estimation of Nonstationary Time Series Regression". Journal of Time Series Analysis, 9, 201-214.

- HATANAKA, M. (1974): "An Efficient Two-step Estimator for the Dynamic Adjustment Model with Autoregressive Errors". Journal of Econometrics, 2, 199-220.

- HECKMAN, N.E. (1986): "Spline Smoothing in a Partly Linear Model". Journal of the Royal Statistical Society. Ser. B, 48, 244-248.

- HOADLEY, B. (1971): "Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case". Annals of Mathematical Statistics, 42, 1977-1991.

- HUBER, P.J. (1980): Robust Statistics. New York: John Wiley.

- JONES, D.A. (1978): "Non-linear Autoregressive Processes". Proceedings to the Royal Society. London. A360, 71-95.

- JUDGE, G.G., GRIFFITHS, W.E., CARTER HILL, R., LUTKEPHOL, H. AND LEE, T.C. (1985): The Theory and Practice of Econometrics. New York: John Wiley.

- KREISS, J.P. (1987): "On Adaptive Estimation in Stationary ARMA Processes". Annals of Statistics, 15, 337-342.

- LeCAM, L. (1960): "Locally Asymptotically Normal Families of Distributions". University of California. Publications in Statistics, 3, 37-98.

- LEVINE, D. (1983): A Remark on Serial Correlation in Maximum Likelihood". Journal of Econometrics, 23, 145-194.

- MANSKI, C.F. (1984): "Adaptive Estimation of Non-linear Regression Models, (with Comment)". Econometric Reviews, 3, 145-194.

- MOKKADEM, A. (1987): "Sur une Modèle Autorégressif non Linéaire, Ergodicité et Ergodicité Géométrique". Journal of Time Series Analysis, 8, 195-204.

- NADARAYA, E.A. (1964): "On estimating Regression". Theory of Probability and its Applications, 9, 141-142.

- NAIMARK, M.A. (1960): Normed Rings. Groningen, Noordhoff.

- NELSON, H.L. AND GRANGER, C.W.J. (1979): "Experience with using the Box-Cox Transformation when Forecasting Economic Time Series". Journal of Econometrics, 10, 57-69.

- NEWEY, W.K. (1986): "Efficient Estimation of Models with Conditional Moment Restrictions". Manuscript.

- NEWEY, W.K. AND WEST, K.D. (1987): "A Simple, Positive, Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix". Econometrica, 55, 703-708.

- NUMMELIN, E. AND TUOMINEN, P. (1982): "Geometric Ergodicity of Harris Recurrent Markov Chains with Applications to Renewal Theory. Stochastic Processes and their Applications, 12, 187-202.

- OZAKI, T. AND HAGGAN, V. (1981): "Modelling Non-linear Random Vibrations using an Amplitude-dependent Autoregressive Time Series Model". Biometrika, 68, 189-196.

- PARZEN, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series". Annals of Mathematical Statistics, 28, 329-348.

- PARZEN, E. (1962): "On Estimation of a Probability Density Function and Mode". The Annals of Mathematical Statistics, 33, 1065-1076.

- PEMBERTON, J. AND TONG, H. (1981): "A note on the Distribution of Nonlinear Autoregression Stochastic Models". Journal of Time Series Analysis, 2, 49-52.

- PHAM, D.T. AND TRAN, L.T. (1986): "Some Mixing Properties of Time Series Models". Stochastic Processes and their Applications, 19, 297-303.

- PIERCE, D.A. (1971): "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors". Biometrika, 58, 299-312.

- PRAIS, S.J. AND WINSTEN, C.B. (1954): "Trend Estimator and Serial Correlation". Cowles Commission Discussion paper No. 383, Chicago.

- PRAKASA-RAO, B.L.S. (1983): Nonparametric Functional Estimation. New York. Academic Press.

- PRESS, W.H., FLANNERY, B.P., TEVKOLSKY, S.A. AND VETTERLING, W.T. (1986). Numerical Recepies. Cambridge University Press.

- ROBINSON, P.M. (1972): "Non-linear Regression for Multiple Time-series". Journal of Applied Probability, 9, 758-768.

- ROBINSON, P.M. (1973): "Generalized Canonical Analysis for Time Series". Journal of Multivariate Analysis, 3, 141-160.

- ROBINSON, P.M. (1976): "Instrumental Variable Estimation of Differential Equations". Econometrica, 44, 765-776.

- ROBINSON, P.M. (1977): "The Estimation of a Nonlinear Moving Average Model". Stochastic Processes and their Applications, 5, 81-90.

- ROBINSON, P.M. (1983): "Nonparametric Estimation for Time Series". Journal of Time Series Analysis, 4, 185-207.

- ROBINSON, P.M. (1984): "Kernel Estimation and Interpolation for Time Series Containing Missing Observations". Annals of the Institute of Statistical Mathematics, 36, 403-417.

- ROBINSON, P.M. (1987a): "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form". Econometrica, 55, 875-891.

- ROBINSON, P.M. (1987b): "Best Nonlinear Three Stage Least Squares Estimation of Certain Nonlinear Econometric Models". Manuscript. London School of Economics.

- ROBINSON, P.M. (1987c): "Adaptive Estimation of Heteroskedastic Econometric Models". R. de Econometria, V. VII nº2,5-28.

- ROBINSON, P.M. (1988a): "Automatic Generalized Least Squares". Manuscript. London School of Economics.

- ROBINSON, P.M. (1988b): "Semi-parametric Econometrics: A Survey". Journal of Applied Econometrics, 3, 35-51.

- ROBINSON, P.M. (1988c): "Root-N-Consistent Semiparametric Regressors". Econometrica, 56, 931-954.

- ROSENBLATT, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition". Proceedings of the National Academie of Science U.S.A., 42, 43-47.

- ROUSSAS, G.G. (1972): Contiguity of Probability Measures. Cambridge University Press.

- STONE, C.J. (1975): "Adaptive Maximum Likelihood Estimation of a Location Parameter". Annals of Statistics, 3, 267-284.

- STONE, C.J. (1977): "Consistent Nonparametric Regression (with Discussion)". Annals of Statistics, 5, 595-645.

- THEIL, H. (1971): Principles of Econometrics. New York: John Wiley.

- TIAO, G.C. AND ALI, M.M. (1971): "Analysis of Correlated Random Effects: Linear Model with two Random Components". Biometrika, 58, 37-51.

- TONG, H. AND LIM, K.S. (1980): "Threshold Autoregression, Limit Cycles and Cyclical Data". Journal of Royal Statistical Society. Ser. B, 245-292.

- VOLKONSKII, V.A. AND ROZANOV, Y.A. (1961): "Some Limit Theorems for Random Functions II". Theory of Probability and its Applications, 6, 186-198.

- WATSON, G.S. (1955): "Serial Correlation in Regression Analysis I". Biometrika, 42, 327-341.

- WATSON, G.S. (1964): "Smooth Regression Analysis". Sankhya, Ser. A, 26, 359-372.

- WATSON, G.S. AND HANNAN, E.J. (1956): "Serial Correlation in Regression Analysis II". Biometrika, 43, 436-445.

- WHISTLER, D. (1988): "Semiparametric ARCH Estimation of Intradaily Exchange Rates Volatility". Manuscrit.

London School of Economics.

- YAKOWITZ, S. (1987): "Nearest-Neighbour Methods for Time Series Analysis". Journal of Time Series Analysis, 8, 235-247.

- YOSHIHARA, K. (1976): "Limiting Behaviour of U-statistics for Stationary Absolutely Regular Processes". Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 35, 237-252.