USC

MRG WP # 8630

MODELLING RESEARCH GROUP

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

db

USC

MRG WP # 8630

MODELLING RESEARCH GROUP

ESTIMATING NONLINEAR DYNAMIC MODELS USING LEAST ABSOLUTE ERROR ESTIMATION

ANDREW A. WEISS

MRG WORKING PAPER #M8630

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

db

ESTIMATING NONLINEAR DYNAMIC MODELS USING LEAST ABSOLUTE ERROR ESTIMATION

ANDREW A. WEISS

MRG WORKING PAPER #M8630

July 1986

ABSTRACT

We consider least absolute error estimation in a nonlinear dynamic model with neither independent nor identically distributed errors. Under the null hypothesis and local alternatives, the estimator is shown to be consistent and asymptotically normal, with asymptotic covariance matrix depending upon the heights of the density functions of the errors at their median (zero). A consistent estimator of the asymptotic covariance matrix of the estimator is given, and the Wald, Lagrange multiplier and Likelihood ratio tests for linear restrictions on the parameters in the regression equation are discussed. The Wald and Lagrange multiplier tests are distributed as central χ^2 under the null and non-central χ^2 under local alternatives. The Likelihood ratio test on the other hand is not always equivalent to the other two tests and may not be asymptotically distributed as χ^2. A simple artificial regression form of the Lagrange multiplier test is available in omitted variables problems, making this test attractive in many testing situations. Since the form of the covariance matrix and tests depend on the presence or absence of heteroscedasticity, a Lagrange multiplier test based on the absolute residuals is analysed.

1. INTRODUCTION

Many common estimation and testing procedures in econometrics are derived to be optimal in the case of Gaussian errors. It is well known however that departures from normality can greatly affect the performance of the procedures. Partly in response to this, robust methods have been developed (see, for example, Koenker (1982)). One such method is least absolute error (LAE) estimation.

In the static, linear model, Bassett and Koenker (1978) have proved that LAE is consistent and asymptotically normal, and is also more efficient than ordinary least squares (OLS) whenever the median is superior to the mean as an estimator of location. This holds for a large number of distributions which have peaked density at the median and/or long tails. Similar results have been obtained by Amemiya (1982) in the simultaneous equations model and Powell (1984) who considered the censored regression model. In the latter, more can be said. Departures from normality imply the usual maximum likelihood methods are inconsistent whereas LAE gives consistent estimates for a wide range of error distributions.

In this paper, we consider the use of LAE in a nonlinear, dynamic model with neither independent nor identically distributed errors. This extends the material in Bassett and Koenker (1978), Oberhofer (1982) who proved the consistency of LAE in the nonlinear model with fixed regressors, and Koenker and Bassett (1982) who analysed the Wald, Lagrange multiplier (LM) and Likelihood ratio (LR) tests in the linear model with fixed regressors estimated by LAE. As we shall see, the results in our model are similar to those in the linear model. Hence LAE will also guard against the same type of error distributions. We show that the LAE estimator is consistent and asymptotically normal and prove the consistency of an estimator of the asymptotic covariance matrix. With these, the Wald, LM and LR tests can be analysed. Two features of these tests are that in a simple exclusion hypothesis in the absence of heteroscedasticity, the LM test statistic has a convenient artificial regression form; and analogously to estimation by OLS, the LR test is not distributed as χ^2 in the presence of heteroscedasticity. To facilitate discussion of the asymptotic power of the various tests, the analysis is done under a sequence of local alternatives.

Since the estimation and testing of the model is simplified when the errors are homoscedastic, we also discuss several tests for heteroscedasticity. The focus is on the natural test based on the absolute values of the residuals. That is, in the spirit of Breusch and Pagan (1979), we show that an LM test based on an artificial regression along the lines of the regressions in Glejser (1969) may be defined. The test statistic is simply TR^2 from the artificial regression.

One important feature of the estimation is that in general, since the model is dynamic, various moments of the errors must be finite. For example, the asymptotic covariance matrix of the estimates involves a second moment matrix which in the linear model matrix is simply "$X'X$". In the dynamic model, the X's include lags of the dependent variable. We also note that in many time series situations, there seems little justification for assuming finite moments for independent variables, but not for the errors.

We therefore consider the results to be applicable in situations where the true, unknown distribution of the errors departs from the normal, but the higher moments still exist. In this case, the OLS estimator also has the usual properties, and in discussing LAE, an important consideration is efficiency relative to OLS. In this paper however, we concentrate on deriving the results described above. As we shall see, given the general model considered, only in the simplest cases is it straightforward to compare the estimators and tests. However a small numerical example shows how OLS can easily be inefficient with heteroscedasticity or non-normal errors.

Another feature resulting from the dynamic nature of the model is that when the distribution is fat-tailed, any outliers will feed back into the design, creating influential observations. However for the linear autoregressive model at least, Koenker (1982) has noted that the outlier and influential observation effects tend to cancel each other and bounded influence estimation (*e.g.*, Belsley, Kuh and Welsch (1979)) is not necessary. We do not consider this in the nonlinear case.

In section 2, we define the model and estimator and discuss the assumptions imposed on the model. Section 3 gives the consistency and asymptotic normality results while in section 4 we discuss the estimation of the asymptotic covariance matrix of the LAE estimator. In section 5 we consider the hypothesis testing and in section 6 we analyse the tests for heteroscedasticity. Section 7 contains some discussion and concluding comments.

2. THE MODEL, ESTIMATES AND ASSUMPTIONS

We consider the following nonlinear dynamic regression model

$$y_t = f(x_t, \beta_T) + \epsilon_t \quad (t = 1, \ldots, T)$$

where x_t is a $(p \times 1)$ vector of inputs, possibly including lags of y_t, β_T is a $(k \times 1)$ vector of unknown parameters and ϵ_t is the tth error.[1] The data generation process is allowed to vary with T, *i.e.,* β_T represents a sequence of true values. Additional T subscripts on y_t, x_t and ϵ_t have been dropped. To obtain useful distributions for the test statistics, we assume that β_T converges to some value β_0. In particular,

$$\beta_T = \beta_0 + \gamma/\sqrt{T} \tag{1}$$

for a (possibly) non-zero $(k \times 1)$ vector γ.

Given T observations on y_t and x_t, the LAE estimator $\hat{\beta}$ is obtained by minimizing

$$Q_T(\beta) = \sum_{t=1}^{T} |y_t - f(x_t, \beta)| \tag{2}$$

with respect to β. In future, we will drop the indices from the summation. As is well-known, $\hat{\beta}$ is not unique. However the different values will be close (since any data point must lie on the same side of every fit line), and we assume that there exists some rule for choosing a unique $\hat{\beta}$. In equation (2), we are also implicitly conditioning on any starting values if x_t contains lagged dependent variables. Neither this nor the non-uniqueness of LAE will not affect the asymptotic results.

For ease of exposition, we adopt the approach of initially defining a number of terms and stating a set of assumptions about these which are sufficient for the majority of the results. Hence, not all the assumptions are used for each result, and in many cases, they could be weakened. Other assumptions are more conveniently introduced later. A discussion follows the statement of the assumptions. Assume:

(A1) (Parameter space) β_0 is an interior point of a compact set $B \in R^k$.

(A2) (Regression function)

i) $f(x_t, \beta)$ is a known, measurable function for each β in B.[2]

ii) $f(x_t, \beta)$ is twice differentiable in β uniformly in t a.s., and the second derivative is continuous in β a.s.

(A3) (Unconditional Error Distribution) Let $u_{t,T}(y)$ be the unconditional density of y_t. Assume that for almost all y, $u_{t,T}(y) \to u_t(y)$ as $T \to \infty$.

Next, let P_T and E_T denote probabilities and expectations taken with respect to $u_{t,T}$, and let P and E denote those taken with respect to u_t. Assume:

(A4) (Conditional Error Distribution) Let I_{t-1} be the information set (σ-algebra) generated by x_{t-i}, $i \geq 0$, and ϵ_{t-i}, $i > 0$, and let $h_{t,T}(\lambda)$ be the density function of ϵ_t conditional on I_{t-1}. Assume that conditional on I_{t-1}, ϵ_t has median zero and that $h_{t,T}(\lambda)$ is continuous, bounded from above and there exist $p_1, h > 0$ such that $P_T(h_{t,T}(0) \geq h) > p_1$, all uniformly in t, T. Further, assume that each $h_{t,T}$ is Lipschitz continuous, *i.e.*, for all λ_1, λ_2 and some finite constant $L_0 > 0$,

$$|h_{t,T}(\lambda_2) - h_{t,T}(\lambda_1)| \leq L_0|\lambda_2 - \lambda_1|$$

uniformly in t, T. Finally, assume that for all λ in a compact set including $\lambda = 0$, $h_{t,T}(\lambda) \to h_t(\lambda)$ a.s. as $T \to \infty$.

(A5) (Dependence Conditions) The random vectors $\{\epsilon_t, x_t\}$ are either ϕ-mixing of size $2r/(r-1)$, $r > 1$, or α-mixing of size $2r/(r-2)$, $r > 2$, where the mixing coefficients are defined with respect to the σ-algebras generated by $\{\epsilon_t, x_t\}$. (See White and Domowitz (1984) for details.)

(A6) (Dominance Conditions) Define the operators $\nabla \equiv \partial/\partial\beta$, $\nabla_i \equiv \partial/\partial\beta_i$, $\nabla^2 \equiv \partial^2/\partial\beta\partial\beta'$ and $\nabla_{ij}^2 \equiv \partial^2/\partial\beta_i\partial\beta_j$, where β_i is the ith element of β. Assume that there exist measurable functions $a_1(x_t, \epsilon_{t-1})$, $a_2(x_t, \epsilon_{t-1})$ and $a_3(y_t)$ such that for all $\beta \in B$ and t, T,

$$|\nabla_i f_t| \leq a_1(x_t, \epsilon_{t-1}), \quad |\nabla_{ij}^2 f_t| \leq a_2(x_t, \epsilon_{t-1}), \quad \text{and} \quad |u_{t,T}(y_t)| \leq a_3(y_t)$$

With v a σ-finite measure, assume that either

i) If $\{x_t, \epsilon_t\}$ are ϕ-mixing, then

$$\int a_1(x_t, \epsilon_{t-1})^3 a_3(y_t) dv < \infty$$

$$\int a_2(x_t, \epsilon_{t-1})^2 a_3(y_t) dv < \infty$$

$$\text{and} \quad \int \epsilon_t^2 a_3(y_t) dv < \infty \quad \text{or}$$

ii) If $\{x_t, \epsilon_t\}$ are α-mixing, then assume that there exists $\mu > 0$ such that for all t, T,

$$\int a_1(x_t, \epsilon_{t-1})^{3+\mu} a_3(y_t) dv < \infty$$

$$\int a_2(x_t, \epsilon_{t-1})^{2+\mu} a_3(y_t) dv < \infty$$

$$\text{and} \quad \int \epsilon_t^2 a_3(y_t) dv < \infty$$

(A7) (Covariance Matrices) Define

$$A_T = T^{-1} \sum E(\nabla f_t \nabla' f_t)$$

$$\bar{A}_T = T^{-1} \sum E(\nabla f_t^0 \nabla' f_t^0)$$

$$D_T = 2T^{-1} \sum E(h_t(0) \nabla f_t \nabla' f_t)$$

and

$$\bar{D}_T = 2T^{-1} \sum E(h_t(0) \nabla f_t^0 \nabla' f_t^0)$$

where ∇f_t^0 is ∇f_t evaluated at β_0. Assume that β_0 is a regular point of $\bar{A}_T$ and $\bar{D}_T$ and that there exists a matrix A such that $\rho' \bar{A}_T \rho - \rho' A \rho \to 0$ as $T \to \infty$ for any real, non-zero $k \times 1$ vector ρ.

(A8) (Identification) Define the OLS criterion function

$$\sigma_T^2(\beta) = T^{-1} \sum [y_t - f(x_t, \beta)]^2$$

and its expectation

$$\bar{\sigma}_T^2(\beta) = T^{-1} \sum E[y_t - f(x_t, \beta)]^2$$

Assume that β_0 is identifiably unique, in the sense of White and Domowitz (1984).[3]

Assumptions (A1) and (A2) are typically of those required on the parameter space and regression function. (A3) and (A4) give the conditions on the errors and their distributions. Note, however that in the dynamic model these are mainly imposed on the conditional, rather than unconditional distributions of the errors. If the conditional median is equal to zero, then

$$E_T(\psi(\epsilon_t) | I_{t-1}) = 0$$

for all t, T, where $\psi(x) = sign(x)$. This can be viewed as the orthogonality condition equivalent to $E(\epsilon_t | I_{t-1})$ or $E(\epsilon_t x_t) = 0$ in least squares. It implies that at β_T, the appropriately defined gradient has zero expectation and assumes that for all β_T, the regression equation is "correctly specified" other than the change in parameters from β_0. Therefore, correlation between the x_t and ϵ_t is ruled out, for example, since then x_t may affect the sign of ϵ_t. We comment on this further in section 7. Also, this part of (A4) is not restrictive provided there is an unknown constant term in the regression function.

Next, we assume that the conditional density $h_{t,T}$ depends on a set of parameters α_T, where $\alpha_T \to \alpha_0$ as $T \to \infty$, and is continuous in α with probability one. This implies that $h_{t,T} \to h_t$ a.s. as $T \to \infty$. For example, consider the ARCH model introduced by Engle (1982). Suppose that

$$\epsilon_t | I_{t-1} \sim N(0, \sigma_{t,T}^2)$$

where $\sigma_{t,T}^2 = E_T(\epsilon_t^2|I_{t-1}) = \alpha_{1,T} + \alpha_{2,T}\epsilon_{t-1}^2$, $\alpha_{1,T} > 0$ and $0 \le \alpha_{2,T} < 1$ for all T. As $T \to \infty$, $\alpha_{1,T} \to \alpha_1$ and $\alpha_{2,T} \to \alpha_2$. The unconditional density $u_{t,T}$ also depends on the parameters $\alpha_{1,T}$ and $\alpha_{2,T}$. Engle (1982) notes that the unconditional fourth moment of ϵ_t is greater than that implied by unconditional normality. In this model,

$$h_{t,T}(0) = \frac{1}{\sqrt{2\pi}\sigma_{t,T}}$$

which is bounded from above, but not below. However, $P_T(h_{t,T}(0) > h) > p_1$ is equivalent to $P_T(\sigma_{t,T} < M) > p_1$, or $P_T(\epsilon_{t-1}^2 < M_1) > p_1$, for some $M, M_1 < \infty$. That is, ϵ_{t-1}^2 is $O_p(1)$. Note also that since $h_{t,T}(0) - h_t(0) \to 0$ as $T \to \infty$, there exists $p > 0$ such that $P(h_t(0) > h) > p$.

Assumption (A6) gives the moment conditions. These conditions, together with assumption (A3) on the unconditional density of the errors, are used for the Law of Large Numbers (LLN) and Central Limit Theorem (CLT). Under the dependence assumption, (A5), these theorems are provided by the LLN and CLT for mixing random variables given in Newey (1985).[4] Essentially, in assumption (A6) we require at least third moments of $\nabla_i f_t$ and second moments of $\nabla_{ij}^2 f_t$ for the ϕ-mixing case. α-mixing is a weaker form of dependence and hence the moment conditions are stronger. The dominance conditions and the degree of dependence in the mixing conditions are also stronger than those required for the usual applications of the LLN and CLT. They are used in section 3 to circumvent the difficulties caused by the discontinuity of the gradient of the absolute value function.

Depending on the form of $f(\cdot)$, the moment conditions will probably impose similar requirements on the moments of ϵ_t. This may also impose particular restrictions on the parameter space B.

Assumption (A7) is also required for the CLT. In particular, as we shall see, $\bar{A}_T$ is the covariance matrix of the gradient of Q_T, evaluated at β_0, and hence must be positive definite. But following White and Domowitz (1984), with assumption (A7), the identification assumption, (A8), implies that A_T and $\bar{A}_T$ are positive definite. Assumption (A8) is also, of course, one of the conditions required for the OLS estimator of β_0 to have the the usual properties. $\bar{D}_T$ turns out to be the equivalent of the matrix of second derivatives of Q_T and hence corresponds to "$X'X$" in the usual linear model. Since $P(h_t(0) > h) > p$, $\bar{D}_T$ and D_T are also positive definite.

These definitions of A_T and D_T allow for the heteroscedasticity in the errors. As noted in White and Domowitz (1984) however, the (technical) requirement that $A_T \to A$ will place some restrictions on the allowable heterogeneity.

Assumption (A8) is used in a different way in the consistency theorem. Define

$$\begin{aligned} V_T(\beta, \beta_T) &= T^{-1}\sum|y_t - f(x_t, \beta)| - T^{-1}\sum|\epsilon_t| \\ &= T^{-1}Q_T(\beta) - T^{-1}Q_T(\beta_T) \end{aligned}$$

The estimator $\hat{\beta}$ also minimizes $V_T(\beta, \beta_T)$ since $Q_T(\beta_T)$ does not depend on β. This form of the criterion function also facilitates the proof of consistency without the (explicit) assumption that the first moment of ϵ_t is finite. Under the moment conditions in assumption (A6), $E_T[V_T(\beta, \beta_T)]$ exists, and $E_T[V_T(\beta, \beta_T)] \to E[V_T(\beta, \beta_0)]$. Under assumptions (A6) and (A8), $E[V_T(\beta, \beta_0)]$ is uniquely minimized at β_0, in the sense of White and Domowitz (1984). As shown in Theorem 1 below, this then implies the consistency of $\hat{\beta}$. We now turn to the formal discussion of the asymptotic properties of $\hat{\beta}$.

3. CONSISTENCY AND ASYMPTOTIC NORMALITY

The existence of the LAE estimator follows from Lemma 2 of Jennrich (1969). Next, we have

THEOREM 1: *(consistency) The LAE estimator $\hat{\beta}$ is weakly consistent for β_0.*

Detailed proofs are given in the appendix. As noted in section 2, Theorem 1 does not explicitly require that any moments of ϵ_t exist, although $\{f(x_t, \beta) - f(x_t, \beta_0)\}$ or equivalently $\nabla' f_t$, must be dominated by uniformly $(r_1 + \theta_1)$ integrable functions, $r_1 > 1$, $\theta_1 > 0$.

Consider next the asymptotic normality of the LAE estimator. Of course, the distribution theory for $\hat{\beta}$ is greatly complicated by the problems with the differentiability of $Q_T(\beta)$. Therefore the standard proofs based on the mean value theorem are not applicable. Instead, following Powell (1984), we base the asymptotic normality result on an extension of the results in Huber (1967). Huber gave conditions which ensure the asymptotic normality of maximum likelihood type estimators in i.i.d. samples when the usual regularity conditions do not apply.

From equation (2), define the gradient vector

$$T^{-\frac{1}{2}}\nabla Q_T(\beta) = T^{-\frac{1}{2}}\sum q_t(\beta)$$

where $q_t(\beta) = \nabla f_t(\beta)\psi\big(y_t - f(x_t, \beta)\big)$ and (β) means the various terms are evaluated at β. Theorem 2 gives conditions under which the sequence $\{\hat{\beta}\}$ which satisfies the "first-order" conditions

$$T^{-\frac{1}{2}}\nabla Q_T(\hat{\beta}) \xrightarrow{p} 0$$

is asymptotically normal. In particular, we show that

$$T^{-\frac{1}{2}}\sum E_{\hat{\beta}}[q_T(\beta)] = -T^{-\frac{1}{2}}\sum q_t(\beta_T) + o_p(1) \tag{3}$$

where $E_{\hat{\beta}}[q_t(\beta)]$ means $E[q_t(\beta)]$ evaluated at $\hat{\beta}$. The asymptotic distribution of $\bar{A}_T^{-\frac{1}{2}}T^{-\frac{1}{2}}$ $\sum q_t(\beta_T)$ is $N(0, I)$ and

$$T^{-\frac{1}{2}}\sum E_{\hat{\beta}}[q_t(\beta)] = D_T(\hat{\beta})T^{1/2}(\hat{\beta} - \beta_T)$$

where $D_T(\hat{\beta})$ is D_T evaluated at $\hat{\beta}$. The asymptotic distribution of $\hat{\beta}_T$ follows immediately.

THEOREM 2: *(asymptotic normality).*

$$\sqrt{T}\bar{A}_T^{-\frac{1}{2}}\bar{D}_T(\hat{\beta} - \beta_0) - \bar{A}_T^{-\frac{1}{2}}\bar{D}_T\gamma \xrightarrow{d} N(0, I)$$

Notice also the curious reversal of the roles of the "hessian" and "information matrix" relative to OLS with heteroscedasticity. That is, in OLS, any heteroscedasticity affects the variance of the gradient vector, rather than the expected value of the matrix of second derivatives of the objective function ($\bar{D}_T$).

In the homoscedastic model, $\bar{D}_T$ reduces to $2h(0)T^{-1}\sum E(\nabla f_t^0 \nabla' f_t^0) = 2h(0)\bar{A}_t$, where $h(0)$ is the height of the common error density function at zero. The asymptotic covariance matrix, $\bar{D}_T^{-1}\bar{A}_T\bar{D}_T^{-1}$, becomes $\omega^2\bar{A}_T^{-1}$ where $\omega = [2h(0)]^{-1}$. This is equivalent to the covariance matrix in the static model. Comparing this to the usual covariance matrix for OLS, *i.e.*, $\sigma_\epsilon^2\bar{A}_T^{-1}$, where $\sigma_\epsilon^2 = E(\epsilon_t^2)$, implies that LAE is more efficient than OLS if

$$\omega^2 < \sigma_\epsilon^2$$

This occurs whenever the sample median is a more efficient estimator of location than the mean, since ω^2 is the asymptotic variance of the sample median of the errors in this case.

We comment further on the comparison of LAE and OLS in section 7 where we briefly compare the (appropriately defined) asymptotic relative efficiency of the Wald, LM and LR test statistics based on LAE and OLS. In the simplest case, this comparison depends only on ω^2 and σ_ϵ^2 as above. The final preliminary to deriving the tests is the estimation of the asymptotic covariance matrix of $\hat{\beta}$.

4. ESTIMATION OF THE ASYMPTOTIC COVARIANCE MATRIX

In order for the asymptotic normality result, Theorem 2, to be useful in hypothesis testing, a consistent estimate of the covariance matrix $\bar{D}_T^{-1}\bar{A}_T\bar{D}_T^{-1}$ is needed. This obviously entails the estimation of $\bar{A}_T$ and $\bar{D}_T$ separately. The estimate of $\bar{A}_T$ is the usual estimate given by replacing the expectations by the corresponding sample quantities evaluated at the LAE estimate, *i.e.*,

$$\hat{A}_T = T^{-1}\sum \nabla\hat{f}_t \nabla'\hat{f}_t$$

where $\nabla\hat{f}_t$ is ∇f_t evaluated at $\hat{\beta}$. The proof that $|\hat{A}_T - \bar{A}_T| \xrightarrow{p} 0$ is standard and is omitted.

The estimation of $\bar{D}_T$ is more difficult since $\bar{D}_T$ involves the density functions of the errors. As noted in Powell (1984) there are no "natural" sample counterparts to the

$h_t(\cdot)$. Estimators usually involve the "smoothing" of the empirical distribution function of the residuals. In the homoscedastic case, $\bar{D}_T$ is $2h(0)\bar{A}_T$ and the problem is then the estimation of $h(0)$. A possible estimator is

$$\hat{h}(0) = [\hat{H}(\hat{c}_T) - \hat{H}(-\hat{c}_T)]/2\hat{c}_T \tag{4}$$

where $\hat{H}(\cdot)$ is the empirical distribution function of the residuals, *i.e.*,

$$\hat{H}(x) = T^{-1}\sum 1(\hat{\epsilon}_t < x)$$

$1(x)$ is the indicator function of the event x, $\hat{\epsilon}_t = y_t - f(x_t, \hat{\beta})$ is the tth residual, and $\hat{c}_T$ is an appropriate constant. Hopefully if $\hat{c}_T \to 0$ as $T \to \infty$, then $\hat{h}(0) \to h(0)$. Powell (1984) considered this problem in the censored regression model.

Notice that $\hat{h}(0)$ may be written as

$$\hat{h}(0) = (2\hat{c}_T T)^{-1}\sum 1(-\hat{c}_T < \hat{\epsilon}_t < \hat{c}_T)$$

This suggests the estimator of $\bar{D}_T$,

$$\hat{D}_T = (\hat{c}_T T)^{-1}\sum 1(-\hat{c}_T < \hat{\epsilon}_t < \hat{c}_T)\nabla \hat{f}_t \nabla' \hat{f}_t$$

This estimator is an extension of one noted in Powell (1984). For $\hat{c}_T$, we assume

(A9) (c_T sequence) For some $\tau_1 < \frac{1}{2}$,

$$T^{\tau_1}\left|\frac{\hat{c}_T}{c_T} - 1\right| \xrightarrow{p} 0$$

where the non-stochastic sequence c_T satisfies $c_T = o(1)$ and $c_T^{-1} = o(T^{\frac{1}{2}-\tau_2})$ for some $0 < \tau_2 < \frac{1}{2}$.

Thus one possible $\hat{c}_T$ sequence is

$$\hat{c}_T = c_0 T^{-\gamma}\hat{\sigma}_\epsilon^2$$

where $c_0 > 0$, $\gamma = .25$ and $\hat{\sigma}_\epsilon^2 = T^{-1}\sum \hat{\epsilon}_t^2$. $\hat{\sigma}_\epsilon^2$ takes into account the scale of the data. The corresponding c_T sequence is $c_T = c_0 T^{-\gamma}\sigma_\epsilon^2$ and it is straightforward to show that $T^{\tau_2}|\hat{c}_T/c_T - 1| \xrightarrow{p} 0$ provided the relevant moments exist. The specification of c_0 and γ is discussed in Powell (1984).

The final requirement is a strengthening of moment conditions.

(A10) (Dominance Condition) In the notation in assumption (A6), assume that

$$\int a_1(x_t, \epsilon_{t-1})^4 a_3(y_t)dv < \infty.$$

THEOREM 3: *(Estimation of D_T).*

$$|\hat{D}_T - \bar{D}_T| \xrightarrow{p} 0$$

Assumption (A10) is used to show that $(c_T T)^{-1}\sum 1(-c_T < \epsilon_t < c_T)\nabla f_t^0 \nabla' f_t^0 - \bar{D}_T \xrightarrow{p} 0$. The proof involves an application of Chebyshev's Inequality and hence requires the fourth moments of ∇f_t^0. In the homoscedasticity case, the equivalent result to Theorem 3 is that $\hat{h}(0) \xrightarrow{p} h(0)$ and setting $\nabla f_t^0 \nabla' f_t^0 = 1$ gives the equivalent proof. The fourth moment assumption is no longer needed.

We now turn to hypothesis testing.

5. HYPOTHESIS TESTING

With the asymptotic normality result and a consistent estimate of the asymptotic covariance matrix, we may analyse the asymptotic distributions of the Wald, LM and LR tests for hypotheses about the parameters in the model. We consider q linear null hypotheses of the form

$$H_0 : R\beta = r$$

in the usual notation. As in White (1984, pp. 76-78) the extension to nonlinear hypotheses is straightforward.[5]

To simplify the distributions of the test statistics under the sequence of local alternatives described by equation (1), we make an additional assumption.

(A11) (Covariance Matrix) There exists a matrix $D > 0$ such that $\bar{D}_T \to D$ as $T \to \infty$.

To analyse the behavior under the null, *i.e.*, with $\gamma = 0$, this assumption is irrelevant.

The Wald test is based on the estimated coefficients when the restrictions are not applied. That is,

$$\xi_W = T(R\hat{\beta} - r)'[R\hat{D}_T^{-1}\hat{A}_T\hat{D}_T^{-1}R']^{-1}(R\hat{\beta} - r)$$

The asymptotic distribution of ξ_W follows directly from Theorem 2, and under the sequence of local alternatives,

$$\xi_W \xrightarrow{d} \chi_q^2(m)$$

where $\chi_q^2(m)$ denotes a non-central χ_q^2 with non-centrality parameter

$$m = \gamma' R'[RD^{-1}AD^{-1}R']^{-1}R\gamma$$

Under H_0, $\xi_W \xrightarrow{d} \chi_q^2$, a central χ_q^2.

Koenker and Bassett (1982) base the LR test on the static

$$\xi_{LR} = 2\omega^{-1}[Q_T(\hat{\beta}_R) - Q_T(\hat{\beta})] \tag{5}$$

where $\hat{\beta}_R$ is the estimate obtained by minimizing $Q_T(\beta)$ subject to the restrictions $R\beta = r$. In practice, ω could be replaced by $\hat{\omega} = [2\hat{h}(0)]^{-1}$.

Following Koenker and Bassett, ξ_{LR} is analysed by approximating the criterion function $Q_T(\beta)$ by a well-behaved quadratic function whose minimizer is asymptotically equivalent to the LAE estimator. In particular, write

$$Q_T(\beta) = \sum|\epsilon_t - f_t(x_t,\beta) + f_t(x_t,\beta_T)|$$
$$= \sum|\epsilon_t - \nabla'\tilde{f}_t(\beta - \beta_T)|$$

where $\nabla' \equiv \partial/\partial\beta'$ and $\nabla'\tilde{f}_t$ is $\nabla' f_t$ evaluated at $\tilde{\beta}$, which lies between β and β_T. With $\delta = \sqrt{T}(\beta - \beta_T)$, also define

$$\tilde{Q}_T(\delta) = \sum |\epsilon_t - \nabla'\tilde{f}_t\delta/\sqrt{T}|$$

where $\tilde{\beta}$ is viewed as a function of δ. A T subscript on δ has been suppressed and in terms of δ, the restrictions $R\beta = r$ are written $R\delta = -R\gamma$. (See, for example, Dijkstra (1984)). The normalized "gradient" vector of $Q_T(\beta)$ is

$$\frac{1}{\sqrt{T}}\sum \nabla f_t(\beta)\psi(\epsilon_t - \nabla'\tilde{f}_t(\beta - \beta_T))$$

and in terms of δ, this becomes

$$g_T(\delta) \equiv T^{-\frac{1}{2}}\sum \nabla f_t(\beta)\psi(\epsilon_t - \nabla'\tilde{f}_t\delta/\sqrt{T}) \tag{6}$$

where β is also viewed as a function of δ, *i.e.*, $\beta = \beta_T + \delta/\sqrt{T}$. The approximating quadratic objective function is given by

$$S_T(\delta) = \frac{1}{2}\delta'\bar{D}_T\delta - \delta' g_T(0) + \tilde{Q}_T(0)$$

To derive the distribution of ξ_{LR}, we begin by showing that $S_T(\delta)$ and $Q_T(\delta)$ are asymptotically equivalent. We then write the LR test statistic as the sum of a number of terms and use this equivalence to show that all but one of these terms converges to zero. This leaves the statistic in a convenient form.

Notice that $S_T(0) = \tilde{Q}_T(0)$. Therefore, if the gradients of $S_T(\delta)$ and $\tilde{Q}_T(\delta)$ are also close, then it follows that $S_T(\delta)$ and $\tilde{Q}_T(\delta)$ will be close. The gradient of $\tilde{Q}_T(\delta)$ is $-g_T(\delta)$, since $Q_T(\beta) = \tilde{Q}_T(\delta)$ and $\partial\delta/\partial\beta = \sqrt{T}$, while that of $S_T(\delta)$ is $\bar{D}_T\delta - g_T(0)$. We have

LEMMA 1: *(asymptotic linearity of $g_T(\delta)$).*

$$\sup_{\|\delta\|<M} \| g_T(\delta) - g_T(0) + \bar{D}_T\delta \| = o_p(1) \quad \text{for all } M > 0$$

Lemma 1 is equivalent to Lemma 4.1 of Bickel (1975), and implies that the gradient vector is asymptotically linear in δ. It will therefore also prove useful for considering the properties of the LM test. Lemma 1 implies

LEMMA 2: *(asymptotic equivalence of $S_T(\delta)$ and $\tilde{Q}_T(\delta)$).*

$$\sup_{\|\delta\|<M} |S_T(\delta) - \tilde{Q}_T(\delta)| = o_p(1) \quad \text{for all} \quad M > 0$$

The proof is immediate from Jaeckel (1972), Lemma 1, and is omitted. Note that the unique minimizer of $S_T(\delta)$ is

$$\hat{\delta}_S \equiv \bar{D}_T^{-1} g_T(0) \tag{7}$$

With Lemma 2, the asymptotic equivalence of $\hat{\delta}$ and $\hat{\delta}_S$ then follows, using the arguments in Jaeckel (1972), Theorem 3.

Returning to equation (5), we see that $Q_T(\hat{\beta}) = \tilde{Q}_T(\hat{\delta})$, and the equivalent of $Q_T(\hat{\beta}_R)$ is

$$\tilde{Q}_T(\hat{\delta}_R) = \sum |\epsilon_t - \nabla \dot{f}_t \hat{\delta}_R / \sqrt{T}|$$

where $\nabla \dot{f}_t$ is ∇f_t evaluated at $\dot{\beta}$ which lies between $\hat{\beta}_R$ and β_T and $\hat{\delta}_R = \sqrt{T}(\hat{\beta}_R - \beta_T)$. To obtain the distribution of ξ_{LR}, write

$$\begin{aligned} Q_T(\hat{\beta}_R) - Q_T(\hat{\beta}) &= \tilde{Q}_T(\hat{\delta}_R) - \tilde{Q}_T(\hat{\delta}) \\ &= [\tilde{Q}_T(\hat{\delta}_R) - S_T(\hat{\delta}_R)] + [S_T(\hat{\delta}_R) - S_T(\hat{\delta}_{S,R})] \\ &\quad + [S_T(\hat{\delta}_{S,R}) - S_T(\hat{\delta}_S)] + [S_T(\hat{\delta}_S) - S_T(\hat{\delta})] \\ &\quad + [S_T(\hat{\delta}) - \tilde{Q}_T(\hat{\delta})] \end{aligned} \tag{8}$$

where $\hat{\delta}_{S,R}$ is the value of δ from minimizing $S_T(\delta)$ subject to $R\delta = -R\gamma$. That is,

$$\begin{aligned} \hat{\delta}_{S,R} &= (I - \bar{D}_T^{-1} R'(R\bar{D}_T^{-1} R')^{-1} R)\bar{D}_T^{-1} g_T(0) \\ &\quad - \bar{D}_T^{-1} R'(R\bar{D}_T^{-1} R')^{-1} R\gamma \end{aligned} \tag{9}$$

Next, using Lemma 2 and the arguments in Amemiya (1982), equations (3.15)-(3.16), equation (8) implies that

$$\begin{aligned} 2\omega^{-1}[\tilde{Q}_T(\hat{\delta}_R) - \tilde{Q}_T(\hat{\delta})] &= 2\omega^{-1}[S_T(\hat{\delta}_{S,R}) - S_T(\hat{\delta}_S)] + o_p(1) \\ &= \omega^{-1}(\hat{\delta}_{S,R} - \hat{\delta}_S)'\bar{D}_T(\hat{\delta}_{S,R} - \hat{\delta}_S) + o_p(1) \end{aligned}$$

where the second equality follows from a Taylor Series expansion. But from equation (9),

$$\begin{aligned}\hat{\delta}_S - \hat{\delta}_{S,R} &= T^{\frac{1}{2}}\bar{D}_T^{-1}R(R\bar{D}_T^{-1}R')^{-1}(R\hat{\beta}_S - r) \\ &= T^{\frac{1}{2}}\bar{D}_T^{-1}R(R\bar{D}_T^{-1}R^1)^{-1}(R\hat{\beta} - r) + o_p(1)\end{aligned} \tag{10}$$

where $\hat{\delta}_S = T^{\frac{1}{2}}(\hat{\beta}_S - \beta_T)$ and $\hat{\beta}_S$ is asymptotically equivalent to $\hat{\beta}$. Therefore,

$$\xi_{LR} = T\omega^{-1}(R\hat{\beta} - r)'(R\bar{D}_T^{-1}R')^{-1}(R\hat{\beta} - r) + o_p(1)$$

Thus ξ_{LR} is equivalent to a statistic in the form of a Wald test statistic. Comparing this to ξ_W, however, we see that ξ_{LR} and ξ_W are not equivalent unless $\bar{D}_T = \omega^{-1}\bar{A}_T = 2h(0)\bar{A}_T$ *i.e.,* there is no heteroscedasticity in the errors. Therefore, in general ξ_{LR} is not asymptotically χ^2, and using χ^2 gives the test the wrong size asymptotically.

The LM test is based on the behavior of the gradient of the sum of absolute residuals, *i.e.,* $g_T(\delta)$, under the restrictions. In particular, since $\hat{\delta}_R = O_p(1)$, it follows from Lemma 1 that

$$g_T(\hat{\delta}_R) - g_T(0) + \bar{D}_T\hat{\delta}_R = o_p(1)$$

where $g_T(\hat{\delta}_R) = T^{-\frac{1}{2}}\sum\nabla f_{tR}\psi(\tilde{\epsilon}_t)$, $\tilde{\epsilon}_t = y_t - f(x_t, \hat{\beta}_R)$ and ∇f_{tR} is ∇f_t evaluated at $\hat{\beta}_R$. Therefore, since $R\hat{\delta}_R = -R\gamma$ and $g_T(0) = -T^{-\frac{1}{2}}\sum\nabla f_t(\beta_T)\psi(\epsilon_t)$,

$$\begin{aligned}R\bar{D}_T^{-1}g_T(\hat{\delta}_R) &= R\bar{D}_T^{-1}g_T(0) + R\gamma + o_p(1) \\ &\xrightarrow{d} N(R\gamma,\ RD^{-1}AD^{-1}R')\end{aligned} \tag{11}$$

The LM test statistic is

$$\begin{aligned}\xi_{LM} &\equiv g_T(\hat{\delta}_R)'\tilde{D}_T^{-1}R'(R\tilde{D}_T^{-1}\tilde{A}_T\tilde{D}_T^{-1}R')^{-1}R\tilde{D}_T^{-1}g_T(\hat{\delta}_R) \\ &\xrightarrow{d} \chi_q^2(m)\end{aligned}$$

where $\tilde{A}_T$ and $\tilde{D}_T$ are A_T and D_T estimated at $\hat{\beta}_R$, respectively. Combining equations (9)-(11) also shows that $\xi_{LM} - \xi_W \xrightarrow{p} 0$.

In the homoscedastic model, ξ_{LM} reduces to

$$g_T(\hat{\delta}_R)'\tilde{A}_T^{-1}R'(R\tilde{A}_T^{-1}R')^{-1}R\tilde{A}_T^{-1}g_T(\hat{\delta}_R)$$

which is attractive since it does not require the estimation of $\bar{D}_T$ or the height of the density of the errors. Clearly, with heteroscedasticity, this advantage of the LM test over the Wald test is lost.

The LM test can be further simplified if the null involves simple exclusion restrictions. Suppose that $\beta' = (\beta_1' : \beta_2')$ and the simple null hypothesis is

$$H_0 : \beta_2 = R\beta = 0$$

where $R = (0 : I)$. With homoscedasticity, the LM test is

$$\xi_{LM} = g_2(\hat{\delta}_R)'\tilde{A}_T^{22}g_2(\hat{\delta}_R)$$

where $g_2(\hat{\delta}_R) = Rg_T(\hat{\delta}_R)$ and $\tilde{A}_T^{22}$ is the block of the partitioned inverse of $\tilde{A}_T$ associated with β_2. Further, define $X = (X_1 : X_2)$ where X_1 and X_2 are the matrices with tth rows $\partial f_t/\partial\beta_1'$ and $\partial f_t/\partial\beta_2'$ respectively, evaluated at $\hat{\beta}_R$, and let $\Psi' = (\psi(\tilde{\epsilon}_1)\dots\psi(\tilde{\epsilon}_T))$. Then under H_o, the gradient of $Q_T(\beta_1)$ is $g_1(\delta) = (I : 0)g_T(\delta)$, and as in the proof of Theorem 2, the first-order conditions for $\hat{\delta}_R$ imply that

$$g_1(\hat{\delta}_R) = T^{-\frac{1}{2}}X_1'\Psi = o_p(1)$$

Hence, using the rules for partitioned inverses, we have

$$\xi_{LM} = T\Psi'X(X'X)^{-1}X_1'\Psi/T + o_p(1)$$

The first term is simply TR^2 from the regression of Ψ on X, since $\Psi'\Psi = T$.

Engle (1982) notes that a wide variety of tests can be viewed as omitted variables problems. Since the LM test statistic can be estimated from an auxiliary regression in these cases, the test is particularly convenient.

6. TESTING FOR HETEROSCEDASTICITY

The absence of heteroscedasticity in the errors leads to simpler expressions for the asymptotic covariance matrix of the LAE estimator and the LM test for exclusion restrictions. The results in sections 3-5 also imply that if any heteroscedasticity is not taken into account, then the covariance matrix of $\hat{\beta}$ will be incorrectly estimated and the size of any test based on this will not be equal to the nominal size asymptotically. Therefore it is important to test for heteroscedasticity.

The easiest tests are again the LM tests and the usual LM tests based on squared errors equivalent to those in Breusch and Pagan (1979) and Engle (1982), for example, may be derived. As noted above however, LAE will be useful when the distribution of the errors is more peaked or has fatter tails than the normal distribution. This suggests that a more powerful test for heteroscedasticity in these situations will be based on a distribution such as the double exponential. In this section, we consider such a test.

The comparison of these tests under different distributions is, however, beyond the scope of this paper. Here, we merely study the tests in isolation.

In any case, to derive the LM test based on the double exponential, we follow White (1982) and Weiss (1986a), for example, who have derived the form of LM tests when the likelihood function (LF) is not necessarily correctly specified. Hence we base the LF on the double exponential without actually assuming that the errors follow this distribution. Define

$$\ell_t(\epsilon_t) = (2\omega_{t,T})^{-1} \exp\{-|\epsilon_t|/\omega_{t,T}\} \tag{12}$$

It is also convenient to define the heteroscedasticity, and hence $\omega_{t,T}$, in terms of $|\epsilon_t|$. In particular, consider heteroscedasticity of the form

$$\begin{aligned} E\left(|\epsilon_t| \;\middle|\; \mathcal{F}_{t-1}\right) &= \omega_{t,T} \\ &= \omega(z_t'\alpha_T) \end{aligned} \tag{13}$$

where $\mathcal{F}_t$ is the Information set generated by $z_{t-i},\ i \geq 0,\ x_{t-i},\ i \geq 0$ and $\epsilon_{t-i},\ i > 0,\ \alpha_T$ is a v×1 vector of parameters, not functionally related to β, and z_t is a v×1 vector with

ith element z_{it}. The function $\omega(\cdot)$ is continuously differentiable in α and has its minimum at $\alpha_{i,T} = 0$, $i = 2, \ldots$, v, where $\alpha_{i,T}$ is the ith element of α_T. The first element of z_t is unity *i.e.*, $z_t' \equiv (1 : z_t^{*'})$, and z_t may contain lagged squared errors, for example. To allow for cases such as this, we define $z_t(\beta)$ as z_t evaluated at β and will use z_{it}^* and $\hat{z}_{it}^*$ to refer to the true elements of z_t^*, *i.e.*, $z_{it}(\beta_T)$, and the elements evaluated at $\hat{\beta}$, *i.e.*, $z_{it}(\hat{\beta})$, respectively. We assume that $z_t(\beta)$ is continuous in β. Further, let

$$\alpha_T = \alpha + \gamma_1/\sqrt{T}$$

where $\alpha' = (\alpha_1, 0, \ldots, 0)$ and γ_1 is a (possibly) non-zero (v $\times$ 1) vector. The null hypothesis sets $\alpha_{2,T} = \ldots = \alpha_{\text{v},T} = 0$. Finally, the mixing conditions in assumption (A5) are assumed to apply to the σ-algebras generated by $\{\epsilon_t, x_t, z_t\}$.

Note that in many distributions, equation (13) also implies conditions on $E(\epsilon_t^2|\mathcal{F}_{t-1})$. For example, in the double exponential, $E(\epsilon_t^2|\mathcal{F}_{t-1}) = 2\omega_{t,T}^2$, while in the normal, $E(\epsilon_t^2|\mathcal{F}_{t-1}) = \pi\omega_{t,T}^2/2$. In both these cases, the heteroscedasticity could be defined on $E(\epsilon_t^2|\mathcal{F}_{t-1})$, and this would then implicitly specify $\omega(\cdot)$. For example, in ARCH model given in section 2, $\omega_{t,T} = (2/\pi)^{\frac{1}{2}}(\alpha_{1,T} + \alpha_{2,T}\epsilon_{t-1}^2)^{\frac{1}{2}}$. In the above notation, $\omega_{t,T} = \omega(z_t'\alpha_T)$, where $\omega(x) = (2/\pi)^{\frac{1}{2}}x^{\frac{1}{2}}$, $z_t = (1, \epsilon_{t-1}^2)$ and $\alpha_T' = (\alpha_{1,T}, \alpha_{2,T})$. Allowing for dynamic behavior such as ARCH may also imply restrictions on the α_i and the form of $\omega(\cdot)$. We do not consider these here (but see, for example, Engle (1982) and Weiss (1986b) for conditions on the ARCH model.)

The log LF is given by

$$L = -T^{-1}\sum \log\ \omega_{t,T} - T^{-1}\sum|\epsilon_t|\omega_{t,T}^{-1}$$

and the score vector for α is

$$\frac{\partial L}{\partial \alpha} = -T^{-1}\sum\omega_{t,T}^{-1}\partial\omega_{t,T}/\partial\alpha + T^{-1}\sum|\epsilon_t|\omega_{t,T}^{-2}\partial\omega_{t,T}/\partial\alpha$$

where $\partial\omega_{t,T}/\partial\alpha = \omega'(z_t'\alpha)z_t$ and $\omega'(x) = \partial\omega/\partial x$. Under the null, $\partial L/\partial\alpha$ is proportional to

$$T^{-1}\sum z_t(|\epsilon_t|\omega_{0,T}^{-1} - 1)$$

where $\omega_{0,T}$ is $\omega_{t,T}$ evaluated at $\alpha_{2,T} = \ldots = \alpha_{\mathrm{v},T} = 0$, *i.e.*, $\omega_{0,T} = \omega(\alpha_{1,T})$. Hence, following Breusch and Pagan (1979) and Engle (1982), we may expect that under some conditions, the LM test will be based on the regression of $|\hat{\epsilon}_t|$ on $\hat{z}_t$, where $\hat{z}_t \equiv z_t(\hat{\beta}) = (1 : z_t^*(\hat{\beta})')'$. In particular, we consider the distribution of

$$d_t = T^{-\frac{1}{2}} \sum (\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t| \hat{\omega}_0^{-1} - 1) \tag{14}$$

where $\hat{z}_t^* = z_t^*(\hat{\beta})$, $\bar{Z}^* = T^{-1} \sum \hat{z}_t^*$ and $\hat{\omega}_0 = T^{-1} \sum |\hat{\epsilon}_t|$. To do this we also introduce several new assumptions concerning α, $\omega(\cdot)$ and z_t. Again, these are discussed after the statement of the assumptions. Let $\epsilon_t(\beta) = y_t - f(x_t, \beta)$ for $\beta \in B$. Assume:

(A13) (Parameter space of α) For all T, $\alpha_T \in$ interior B_α where B_α is a compact subset of R^{v}.

(A14) (Conditions on $z_t(\beta)$) Assume that there exist measurable functions $a_i(z_t, x_t, \epsilon_{t-1})$, $i = 4, 5, 6$, and $a_7(x_t, \epsilon_t)$, a.s. continuously differentiable functions $\tilde{z}_{it}(\beta) \equiv \tilde{z}_{it}(z_t, x_t, \epsilon_{t-1}, \beta)$, $i = 2, \ldots, \mathrm{v}$, and constants $\eta > 0$ and $s > 1$ such that for all t, and $\beta \in B$,

i)
$$|z_{it}(\beta)| \leq a_4(z_t, x_t, \epsilon_{t-1}) \quad i = 2, \ldots, \mathrm{v}$$

$$|\partial z_{it}/\partial \beta_j| \leq a_5(z_t, x_t, \epsilon_{t-1}) \quad i = 2, \ldots, \mathrm{v},\ j = 1, \ldots, k$$

$$|\omega'(z_t'\alpha)| \leq a_6(z_t, x_t, \epsilon_{t-1}) \quad \text{and} \quad |\epsilon_t(\beta)| \leq a_7(x_t, \epsilon_t)$$

ii)
$$\int [a_7(x_t, \epsilon_t) a_4(z_t, x_t, \epsilon_{t-1})]^{2s} a_3(y_t) dv < \infty$$

iii)
$$\int [a_5(z_t, x_t, \epsilon_{t-1}) a_1(x_t, \epsilon_{t-1})]^{1+\eta} a_3(y_t) dv < \infty$$

iv)
$$\int [a_7(z_t, \epsilon_t) a_4(z_t, x_t, \epsilon_{t-1})^2 a_6(z_t, x_t, \epsilon_{t-1})]^{1+\eta} a_3(y_t) dv < \infty$$

v)
$$\sum |z_{it}^*(\beta) - \tilde{z}_{it}(\beta)| \xrightarrow{p} 0 \quad i = 2, \ldots, \mathrm{v}$$

(A15) (Covariance Matrix and Non-Centrality Parameter) Define

$$\Lambda_T = T^{-1}\sum E_T\left[(|\epsilon_t|\omega_{t,T}^{-1} - 1)^2(z_t^* - Z_T^*)(z_t^* - Z_T^*)'\right]$$

and

$$K_T = \omega_0^{-1}\omega'(\alpha_1)T^{-1}\sum E_T\left[(z_t^* - Z_T^*)z_t'\right]$$

where $Z_T^* = T^{-1}\sum E_T(z_t^*)$. Assume that there exists a positive definite matrix Λ and a matrix K such that $\Lambda_T \to \Lambda$ and $K_T \to K$ as $T \to \infty$.

When the limiting matrices in assumption (A15) exist, they are given by

$$\Lambda = \lim_{T\to\infty} T^{-1}\sum E(|\epsilon_t|\omega_0^{-1} - 1)^2(z_t^* - Z^*)(z_t^* - Z^*)'$$

and

$$K = \omega'(\alpha_1)\omega_0^{-1}\lim_{T\to\infty} T^{-1}\sum E[(z_t^* - Z^*)z_t']$$

where $Z^* = \lim T^{-1}\sum E(z_t^*)$ as $T \to \infty$ and $\omega_0 = \omega(\alpha_1) = \lim \omega_{0,T}$ as $T \to \infty$. These matrices are related to the asymptotic distribution of d_t in equation (14). In particular, in the proof of Theorem 4 below, we show that d_t has the same asymptotic distribution as

$$\tilde{d}_t = T^{-\frac{1}{2}}\sum(z_t^* - Z_T^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1)$$

which in turn is asymptotically normal with covariance matrix Λ_T. Therefore Λ_T must be invertible for T large enough. Also, since the test sets $\alpha_{i,T} = 0$ $i = 2,\ldots,$v and under the local alternative, $E_T[(|\epsilon_t|\omega_{0,T}^{-1} - 1)|\mathcal{F}_{t-1}] = \omega_{t,T}\omega_{0,T}^{-1} - 1 \neq 0$, the asymptotic distribution has a non-zero mean which is related to K_T (through a mean value expansion of $\omega_{t,T} = \omega(z_t'\alpha_T)$ about $(\alpha_{1,T}, 0\ldots, 0)$.)

We discuss assumption (A14) by way of three examples. Most of assumption (A14) is concerned with the conditions for the random variables d_t and $\tilde{d}_t$ to have the same asymptotic distribution, and in particular, with replacing $\hat{z}_t^*$ by z_t^*. For this, the $\tilde{z}_{it}$ are viewed as continuously differentiable functions which approximate the z_{it}^*, *i.e.*, $\hat{z}_{it} - z_{it}^* \equiv$

$z_{it}(\hat{\beta}) - z_{it}(\beta_T) = z_{it}(\hat{\beta}) - \tilde{z}_{it}(\hat{\beta}) + \tilde{z}_{it}(\hat{\beta}) - \tilde{z}_{it}(\beta_T) + \tilde{z}_{it}(\beta_T) - z_{it}(\beta_T)$. More details on this are given in the third example where we focus on part v) of assumption (A14).

The first example is when z_t is not a function of β, *e.g.*, z_t contains elements in x_t. Then $\hat{z}_{it} = z^*_{it}$ by definition and the approximation is unnecessary, as are parts iii) and v) of assumption (A14). The LM test in Breusch and Pagan (1979) uses variables of this type. There, the dependent variable in the artificial regression is $\hat{\epsilon}_t^2$ rather than $|\hat{\epsilon}_t|$. Since $\epsilon_t^2(\beta)$ is continuously differentiable in β, a mean value expansion may be used to replace $\hat{\epsilon}_t^2$ by ϵ_t^2, simplifying the analysis.

Next, consider the case where $z_t^* = \epsilon_{t-1}^2$, *i.e.*, a simple ARCH-type model. In this case, $z_t^*(\beta) = \epsilon_{t-1}^2(\beta)$ is already a continuously differentiable function of β, and hence we set $\hat{z}_{it} = \tilde{z}_{it}$. Part v) of assumption (A14) is satisfied automatically, and using a mean values expansion of $\epsilon_{t-1}^2(\hat{\beta})$, $\hat{z}_t - z_t^*$ becomes

$$\epsilon_{t-1}^2(\hat{\beta}) - \epsilon_{t-1}^2(\beta_T) = 2\epsilon_{t-1}\partial\epsilon_{t-1}/\partial\beta(\hat{\beta} - \beta_T)$$

where $\epsilon_{t-1}\partial\epsilon_{t-1}/\partial\beta$ is evaluated at β^*, which lies between $\hat{\beta}$ and β_T. This makes the analysis relatively straightforward. The LM test for ARCH in Engle (1982) uses lagged squared errors as regressors. Again, $\hat{\epsilon}_t^2$ is the dependent variable.

The third example we consider is $z_t^* = |\epsilon_{t-1}|$. Then z_t^* is not continuously differentiable and we cannot take either of the above approaches. A possible form for the approximating function is

$$\tilde{z}_t(\beta) = 2k_T^{-1} \cdot \log(1 + \exp[-k_T\epsilon_{t-1}(\beta)]) + \epsilon_{t-1}(\beta)$$

for suitable constants k_T. This function was used by Amemiya (1982) to analyse LAE in a static model. $\sum\tilde{z}_t(\beta)$ (with ϵ_t replacing ϵ_{t-1}) was used as a twice continuously differentiable criterion function which is asymptotically equivalent to $\sum|\epsilon_t(\beta)|$, and whose minimum is asymptotically normal and asymptotically equivalent to the LAE estimator. With $k_T = T^d$, $d > \frac{2}{3}$, this choice for $\tilde{z}_t$ satisfies the conditions of assumption (A14).

THEOREM 4: *(LM Test for Heteroscedasticity) Under the sequence of local alternatives* α_T,

$$\xi_H \equiv T^{-1}\sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)(\hat{z}_t^* - \bar{Z}^*)'\hat{\Lambda}^{-1}\sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)(\hat{z}_t^* - \bar{Z}^*) \xrightarrow{d} \chi^2_{V-1}(m)$$

where

$$\hat{\Lambda} = T^{-1}\sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)^2(\hat{z}_t^* - \bar{Z}^*)(\hat{z}_t^* - \bar{Z}^*)' \xrightarrow{p} \Lambda$$

and

$$m = \gamma_1' K' \Lambda^{-1} K \gamma_1.$$

Under some additional conditions, the LM test has a simple TR^2 form. In particular, if $E[(|\epsilon_t|\omega_0^{-1} - 1)^2|\mathcal{F}_{t-1}]$ does not depend on t, then Λ reduces to

$$E(|\epsilon_t|\omega_0^{-1} - 1)^2 \lim_{T\to\infty} T^{-1}\sum E(z_t^* - Z^*)(z_t^* - Z^*)'$$

and ξ_H becomes

$$\hat{\xi}_H = T\sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)(z_t^* - \bar{Z}^*)'\left[\sum(z_t^* - \bar{Z}^*)(z_t^* - \bar{Z}^*)'\right]^{-1} \times \sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)(\hat{z}_t^* - \bar{Z}^*) \,/\, \sum(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1)^2$$

which is TR^2 from the regression of $|\hat{\epsilon}_t|$ on $\hat{z}_t$.

Of course, in the case v= 2, *i.e.*, $z_t' = (1, z_t^*)$, this regression is equivalent to that suggested by Glejser (1969). However, instead of considering the t-statistics on the estimated coefficients in the regression as Glejser did, the LM test is based on the overall R^2. Presumably, of course, in the linear heteroscedasticity model, with $\omega(x) = x$, an equivalent Wald test could be based on these t-statistics. A similar test was also considered by Bickel (1978).

7. CONCLUDING COMMENTS

In a dynamic nonlinear model, we have analysed estimation and hypothesis testing based on the least absolute error criterion. The results represent extensions of those given in Koenker and Bassett (1982) and Oberhofer (1982).

The next step is to compare the results with those based on squared errors. First, consider the Wald and LM tests for restrictions on β_0. As is well known,

$$\xi_{LS} = (R\hat{\beta}_{LS} - r)'[RD_{LS}^{-1}A_{LS}D_{LS}^{-1}R']^{-1}(R\hat{\beta}_{LS} - r) \xrightarrow{d} \chi_q^2(m_{LS})$$

where $\hat{\beta}_{LS}$ is the OLS estimator of β_0,

$$m_{LS} = \gamma' R'[RD_{LS}^{-1}A_{LS}D_{LS}^{-1}R']^{-1}R\gamma$$

$$A_{LS} = E[T^{-1}\sum \epsilon_t^2 \nabla f_t^0 \nabla' f_t^0]$$

and

$$D_{LS} = E[T^{-1}\sum \nabla f_t^0 \nabla' f_t^0]$$

The asymptotic relative efficiency (ARE) of ξ_W to ξ_{LS} is the ratio of the noncentrality parameters and "may be interpreted as the ratio of sample sizes required to achieve a specified power for both tests for a specified level and alternative" (Koenker and Bassett (1982)). In the simplest case of no heteroscedasticity and a simple exclusion hypothesis, the ARE becomes $\sigma_\epsilon^2/\omega^2$. At the normal distribution, $ARE \approx .64$, so under this interpretation, ξ_W needs 36% more observations asymptotically than ξ_{LS}. Koenker and Bassett (1982) comment on this example further.

In more complex models, it may be difficult to compare the noncentrality parameters analytically. As the distribution becomes long-tailed, however, the advantage of the tests based on OLS is lost. In table 1 we present the results from a specific example. Here,

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$

Table 1. Ratio of Noncentrality Parameters[a]

a) Normal

ϕ	α_2 = 0	.1	.2	.3	.4	.5
0	.64	.65	.70	.79	.95	1.23
.4	.64	.66	.70	.80	.93	1.31
.8	.64	.65	.70	.78	.94	1.24

b) Student-t_{11}

ϕ	α_2 = 0	.1	.2	.3	.4	.5
0	.74	.79	.88	1.03	1.27	1.74
.4	.74	.79	.89	1.06	1.38	1.84
.8	.74	.77	.85	1.02	1.21	1.93

[a]Average ratio over 200 replications, each with 1000 observations.

$$E(\epsilon_t \mid I_{t-1}) = 0$$

$$E(\epsilon_t^2 \mid I_{t-1}) = 1 + \alpha_2 \epsilon_{t-1}^2$$

$R = 1$ and the conditional distribution is either normal or a student-t_{11} transformed to have conditional variance $(1 + \alpha_2 \epsilon_{t-1}^2)$. In each category in table 1, 1000 observations were generated and used to estimate the ratio of the non-centrality parameters. The entries are the average ratios over 200 replications. A more extreme case is the double exponential distribution where for $\alpha_2 = 0$, the ratio is already 2. Clearly, it is easy to find models in which the tests based on LAE are more powerful than those based on OLS.[6]

Similarly, we may compare the LM test for heteroscedasticity given in section 6 with the usual LM tests. In the latter, the covariance matrix in the test statistic uses $(\epsilon_t^2 \sigma_{t,T}^{-1} - 1)^2$ rather than $(|\epsilon_t| \omega_{t,T}^{-1} - 1)^2$ and presumably either $\omega(\cdot)$ would take on a different functional form or z_t would be defined differently. The relative efficiency of the two approaches under different error distributions is a topic for future work, *e.g.*, a Monte Carlo experiment. However the numerical results in the simple case considered by Newey and Powell (1986) again demonstrate the potential degree of relative inefficiency of OLS-based methods in the presence of non-normal, heteroscedastic errors.

Next, in section 2 we essentially assumed that the model was correctly specified by assuming that the errors always have median zero. With misspecification, we cannot necessarily assume this and hence the properties of the estimators and tests will change. An analysis of misspecification similar to that in Newey (1985) would then be appropriate. Similarly, the properties of the Hausman test based on the difference between the LAE and OLS estimates would follow from this approach.

Finally, examples of the use of LAE in time series models may be found in Weiss and Andersen (1984) and Bloomfield and Steiger (1983). In the former, LAE is used to estimate ARIMA models for 80 series. This produced better forecasts relative to the same models estimated by OLS. Included in the latter is a small Monte Carlo experiment on the sampling properties of LAE in an autoregressive model.

MATHEMATICAL APPENDIX

Proof of Theorem 1. The proof follows that of Powell (1984), Theorem 1, or Oberhofer (1982), *i.e.*,

i) Show that $V_T(\beta,\beta_T) \xrightarrow{p} E[V_T(\beta,\beta_0)]$ where $E[V_T(\beta,\beta_0)]$ is continuous in β.

ii) Show that $E[V_T(\beta,\beta_0)]$ is uniquely minimized at β_0.

A convergence in probability version of Lemma 2.2 of White (1980) then gives consistency.

i) Let $z_{t,T} = z_t(\beta,\beta_T) = f(x_t,\beta) - f(x_t,\beta_T)$ and write

$$\begin{aligned} V_T(\beta,\beta_T) &= T^{-1}\sum|\epsilon_t - z_{t,T}| - |\epsilon_t|, \\ &\leq T^{-1}\sum(|\epsilon_t| + |z_{t,T}| - |\epsilon_t|) \\ &= T^{-1}\sum|z_{t,T}| \end{aligned}$$

But $z_{t,T} = \nabla f_t^*(\beta - \beta_T)$ where ∇f_t^* is ∇f_t evaluated at β^* which lies between β and β_T. Under assumptions (A3), (A5) and (A6) we may apply the LLN. Therefore

$$|V_T(\beta,\beta_T) - E[V_T(\beta,\beta_0)]| \xrightarrow{p} 0$$

where $E[V_T(\beta,\beta_0)]$ is continuous in β.

ii) Let $v_t(\beta,\beta_0) = |\epsilon_t - z_t| - |\epsilon_t|$, where $z_t = z_t(\beta,\beta_0)$. Then by considering the four regions $\epsilon_t - z_t > 0,\ \epsilon_t > 0$; $\epsilon_t - z_t > 0,\ \epsilon_t < 0$; $\epsilon_t - z_t < 0,\ \epsilon_t > 0$ and $\epsilon_t - z_t < 0,\ \epsilon_t < 0$; and hence the two regions $z_t > 0$ and $z_t < 0$, we obtain

$$\begin{aligned} E(v_t(\beta,\beta_0)|I_{t-1}) = 1(z_t > 0)\ 2\int_{-z_t}^{0} \{|z_t| + \lambda\} h_t(\lambda) d\lambda \\ + 1(z_t < 0)\ 2\int_{0}^{-z_t} \{|z_t| - \lambda\} h_t(\lambda) d\lambda \end{aligned}$$

where 1 is the indicator function. Therefore

$$\begin{aligned} E[v_t(\beta,\beta_0)] =& E\Big\{E\big([2\ 1(z_t > 0)\int_{-z_t}^{0} \{|z_t| + \lambda\} h_t(\lambda) d\lambda \\ &+ 2\ 1(z_t < 0)\int_{0}^{-z_t} \{|z_t| - \lambda\} h_t(\lambda) d\lambda\big] \,\Big|\, h_t(0) > h\big)\Big\}. \end{aligned}$$

Now, since $h_t(\lambda)$ is continuous, when $h_t(0) > h$ there exists $h_1 > 0$ such that $h_t(\lambda) > h_1$ for $|\lambda| < h_1$. Hence for any ϕ such that $0 < \phi < h_1$

$$\begin{aligned}
E[v_t(\beta,\beta_0)] &\geq E\Big\{E([2\,1(z_t > 0)\,1(z_t > \phi) \int_{-\phi}^{0} (\phi+\lambda)h_1 d\lambda \\
&\qquad + 2\,1(z_t < 0)\,1(-z_t > \phi) \int_{0}^{\phi} (\phi-\lambda)h_1 d\lambda] \Big| h_t(0) > h)\Big\} \\
&\geq E\Big\{E([1(z_t > \phi)\,\phi^2 h_1 + 1(-z_t > \phi)\,\phi^2 h_1] \Big| h_t(0) > h)\Big\} \\
&= \phi^2 h_1 E\{E[1(|z_t| > \phi)|h_t(0) > h]\} \\
&= \phi^2 h_1 P(|z_t| > \phi)
\end{aligned}$$

Therefore

$$E[V_T(\beta,\beta_0)] \geq \phi^2 h_1 T^{-1} \sum P[|z_t| > \phi]$$

But under assumptions (A6) and (A8), and following Powell (1984) Theorem 1, for $\| \beta - \beta_0 \| \geq \phi_1 > 0$, $E[V_T(\beta,\beta_0)] > 0$.

Proof of Theorem 2. We begin the proof with the extension of the results in Huber (1967). This gives several sufficient conditions, which we then verify. The application of the mixing CLT then gives asymptotic normality.

Define

$$\lambda_T(\beta) = T^{-1} \sum E_T[q_t(\beta)]$$

and for $d > 0$,

$$\mu_t(\beta, d) = \sup_{\tau}\{\| q_t(\tau) - q_t(\beta) \| : \| \tau - \beta \| \leq d\}$$

The following conditions are imposed on q_t, λ_T and μ_t.

Assumption (N-1) For each t, $q_t(\beta)$ is measurable for each $\beta \in B$ and is separable in the sense of Doob (Doob (1953), p. 51-52).

Assumption (N-2) For each T, there is some $\beta \in B$ such that $\lambda_t(\beta) = 0$.

Assumption (N-3) There are strictly positive numbers a, b, c, d_0 and T_0 such that, for all t and $T \geq T_0$

i) $\| \lambda_T(\beta) \| \geq a \| \beta - \beta_T \|$ for $\| \beta - \beta_T \| \leq d_0$.

ii) $E_T[\mu_t(\beta, d)] \leq bd$ for $\| \beta - \beta_T \| + d \leq d_0$, $d \geq 0$.

iii) There exists constant $D > 0$ and $d \geq 0$ such that either

a) If $\{x_t, \epsilon_t\}$ is ϕ-mixing then

$$E_T[\mu_t(\beta, d)]^2 \leq c.d \quad \text{for} \quad \| \beta - \beta_T \| + d \leq d_0$$

and $\{x_t, \epsilon_t\}$ is of size 2 or

b) If $\{x_t, \epsilon_t\}$ is α-mixing then

$$E_T[\mu_t(\beta, d)]^q \leq c.d \quad \text{for} \quad \| \beta - \beta_T \| + d \leq d_0$$

and $\{x_t, \epsilon_t\}$ is of size $2q/(q-2)$, $q > 2$.

Assumption (N-4) $E_T \| q_t(\beta_T) \|^2 \leq K$ for all t, T and some $K > 0$.

Remark: The role of assumption (N-1) is to ensure measurability of the suprema occuring below. Under the assumptions in the text, this assumption is satisfied.

LEMMA A.1: *Define*

$$Z_T(\tau, \beta) = \frac{|\sum(q_t(\tau) - q_t(\beta) - \lambda_T(\tau) + \lambda_T(\beta))|}{\sqrt{T} + T \| \lambda_T(\tau) \|}$$

Then assumptions (N-1)-(N-4) imply

$$\sup_{\|\tau - \beta_T\| \leq d_0} Z_T(\tau, \beta_T) \xrightarrow{p} 0$$

Further, if $T^{-\frac{1}{2}} \sum q_t(\hat{\beta}) \xrightarrow{p} 0$ *and* $\hat{\beta} - \beta_T \xrightarrow{p} 0$, *then*

$$T^{-\frac{1}{2}} \sum q_t(\beta_T) + T^{\frac{1}{2}} \lambda_T(\hat{\beta}) \xrightarrow{p} 0$$

PROOF: The proof is identical to that in section 4 of Huber (1967) except that

i) terms of the form

$$T\lambda(\theta), \quad TE\mu(z, \theta, d), \text{ etc.}$$

(in Huber's notation) can be replaced by their "averaged" counterparts

$$T\lambda_T(\theta), \quad \sum E\mu_t(z_t, \theta, d), \text{ etc.}$$

without affecting the validity of the argument if T is sufficiently large (as was noted by Powell (1984)).

ii) In the applications of Chebychev's Inequality, we use the mixing conditions and Lemma 2.2 of White and Domowitz (1984) to ensure finite variances.

Proof of the Theorem. The method is now to show that $T^{-\frac{1}{2}}\sum q_t(\hat{\beta}) \xrightarrow{p} 0$ and that assumptions (N-2)-(N-4) are satisfied. The second conclusion in Lemma A.1 will then imply asymptotic normality.

Now

$$T^{-\frac{1}{2}}\sum q_t(\hat{\beta}) = T^{-\frac{1}{2}}\sum \nabla f_t(\hat{\beta})\psi(y_t - f(x_t, \hat{\beta}))$$

Following Ruppert and Carroll (1980), for constant $a > 0$ define

$$G_j(a) = T^{-\frac{1}{2}}\sum |y_t - f(x_t, \hat{\beta} + ae_j)|$$

where the summation is over terms with $\nabla_j f_t(\hat{\beta}) \neq 0$ and $e_j = (0 \ldots 0\ 1\ 0 \ldots 0)'(k \times 1)$ with the 1 in the jth position. Further, let

$$H_j(a) = T^{-\frac{1}{2}}\sum H_{tj}(a)$$

where $H_{tj}(a) = -\nabla_j f_t(\hat{\beta} + ae_j)\psi(y_t - f(x_t, \hat{\beta} + ae_j))$ and the summation is over the same terms as G_j. Now for each H_{tj}, pick a_{tj} small enough so that $\psi(\nabla_j f_t(\beta))$ is constant over $\{\hat{\beta} - a_{tj}e_j, \hat{\beta} + a_{tj}e_j\}$. Since $\nabla_j f_t$ is continuous in β, each $a_{tj} > 0$. Pick $\epsilon = \min_{t,j}\{a_{tj} : \nabla_j f_t(\hat{\beta}) \neq 0\}$. Then

$$H_j(-\epsilon) \leq H_j(0) \leq H_j(\epsilon)$$

and because $G_j(a)$ achieves its minimum at $a = 0$, $H_j(-\epsilon) \leq 0$ and $H_j(\epsilon) \geq 0$. Consequently

$$|H_j(0)| \leq H_j(\epsilon) - H_j(-\epsilon)$$

and letting $\epsilon \to 0$, $H_j(0) = T^{-\frac{1}{2}} \sum q_t(\hat{\beta})$ and

$$\begin{aligned} |H_j(0)| &\leq 2T^{-\frac{1}{2}} \sum |\nabla f_t(\hat{\beta})| 1(y_t - f(x_t \hat{\beta}) = 0) \\ &\leq 2T^{-\frac{1}{2}} a_1(x_t, \epsilon_{t-1}) \sum 1(y_t - f(x_t, \hat{\beta}) = 0). \end{aligned}$$

But $T^{-\frac{1}{2}} a_1(x_t, \epsilon_{t-1}) \xrightarrow{p} 0$ and the summation is finite with probability one for all T large enough since the distribution of ϵ_t is continuous.

Next, assumptions (N-2)-(N-4) must be verified.

a) (N-2). Simply note that $E_T(q_t(\beta_T)|I_{t-1}) = 0$.

b) (N-3) i). Write

$$\lambda_T(\beta) = T^{-1} \sum E_T \left\{ \nabla f_t(\beta) E_T[\psi(y_t - f(x_t, \beta)) \;\Big|\; I_{t-1}] \right\}$$

Evaluating the conditional expectation gives

$$\begin{aligned} 2\Big[P_T(\epsilon_t < 0 \mid I_{t-1}) - P_T\big(\epsilon_t < \nabla' f_t(\beta^*)(\beta - \beta_T) \mid I_{t-1}\big)\Big] &= 2 \int_{\nabla' f_t^*(\beta - \beta_T)}^{0} h_{t,T}(\lambda) d\lambda \\ &= -2\nabla' f_t(\beta^*)(\beta - \beta_T) h_{t,T}(\lambda_t^*) \end{aligned}$$

where $\nabla' f_t^* = \nabla' f_t(\beta^*)$, β^* lies between β and β_T, and λ_t^* lies between $\nabla' f_t^*(\beta - \beta_T)$ and 0. Hence

$$\lambda_T(\beta) = -2T^{-1} \sum E_T[h_{t,T}(\lambda_t^*) \nabla f_t(\beta) \nabla' f_t(\beta^*)](\beta - \beta_T). \tag{A.1}$$

But using the Lipschitz continuity of $h_{t,T}$ and a mean value expansion for ∇f_t,

$$\begin{aligned} T^{-1} &\sum E_T[h_{t,T}(\lambda_t^*) \nabla f_t(\beta) \nabla' f_t(\beta^*)] - D_T \\ &= T^{-1} \sum E_T\Big[[h_{t,T}(\lambda_t^*) - h_{t,T}(0)] \nabla f_t(\beta) \nabla' f_t(\beta^*)\Big] \\ &\quad + T^{-1} \sum E_T\Big[h_{t,T}(0)[\nabla f_t(\beta) \nabla' f_t(\beta^*) - \nabla f_t(\beta) \nabla' f_t(\beta)]\Big] \\ &\leq L_0 T^{-1} \sum E_T\Big[\nabla f_t(\beta) \nabla' f_t(\beta^*) |\nabla f_t'(\beta^*)(\beta - \beta_T)|\Big] \\ &\quad + T^{-1} \sum E_T\Big[h_{t,T}(0) \nabla f_t(\beta) \nabla^2 f_t(\beta^{**})(\beta^* - \beta)\Big] \end{aligned}$$

where β^{**} lies between β^* and β. Hence

$$\lambda_T(\beta) = -D_T(\beta - \beta_T) + O(\| \beta - \beta_T \|^2)$$

and

$$\| \lambda_T(\beta) \| \geq a \| \beta - \beta_T \|$$

for $\| \beta - \beta_T \|$ sufficiently small.

c) (N-3) ii). Write

$$\mu_t(\beta, d) \leq \sup_{\|\tau-\beta\|<d} \| \nabla f_t(\beta)\psi_t(\beta) - \nabla f_t(\beta)\psi_t(\tau) \|$$

$$+ \sup_{\|\tau-\beta\|<d} \| \nabla f_t(\beta)\psi_t(\tau) - \nabla f_t(\tau)\psi_t(\tau) \|$$

where $\psi_t(\beta) = \psi\big(y_t - f(x_t, \beta)\big)$. Now, by a mean value expansion of $\nabla f_t(\beta)$ and since $|\psi_t(\tau)| = 1$,

$$\sup_{\|\tau-\beta\|<d} \| \big(\nabla f_t(\beta) - \nabla f_t(\tau)\big)\psi_t(\tau) \| \leq a_2(x_t, \epsilon_{t-1})d \tag{A.2}$$

Further,

$$\sup_{\|\tau-\beta\|<d} \| \nabla f_t(\beta)\big(\psi_t(\beta) - \psi_t(\tau)\big) \|$$

$$\leq \sup_{\|\tau-\beta\|<d} 2 \| \nabla f_t(\beta) 1\Big(|\epsilon_t - \big(f(x_t, \beta) - f(x_t, \beta_T)\big)| < |f(x_t, \beta) - f(x_t, \tau)|\Big) \| \tag{A.3}$$

$$\leq 2\, a_1(x_t, \epsilon_{t-1}) 1\Big(|\epsilon_t - \big(f(x_t, \beta) - f(x_t, \beta_T)\big)| < a_1(x_t, \epsilon_{t-1})d\Big)$$

Also,

$$\begin{aligned} E_T\Big[1\Big(|\epsilon_t - \big(f(x_t, \beta) - f(x_t, \beta_T)\big)| < a_1(x_t, \epsilon_{t-1})d\Big) \;\Big|\; I_{t-1}\Big] \\ = a_1(x_t, \epsilon_{t-1}) h_{t,T}(\lambda_t^{**}) d \\ \leq H\, a_1(x_t, \epsilon_{t-1}) d \end{aligned} \tag{A.4}$$

where H is the upper bound on $h_{t,T}$ and λ_t^{**} lies between $\big(f(x_t, \beta) - f(x_t, \beta_T) - a_1(x_t, \epsilon_{t-1})d\big)$ and $\big(f(x_t, \beta) - f(x_t, \beta_T) + a_1(x_t, \epsilon_{t-1})d\big)$. Combining equations (A.2) and (A.4) shows that $E_T[\mu_t(\beta, d)] \leq bd$ for some $b > 0$.

d) (N-3) iii). This follows from the "c_r inequality" (given as Theorem 1.2.6 of Lukacs (1968)), equations (A.2) and (A.3), assumption (A6) and noting that $d \leq d_0$.

e) (N-4). Simply note that $q_t(\beta) \leq a_1(x_t, \epsilon_{t-1})$.

Hence, the conditions of Lemma A.1 are satisfied and so, by Lemma A.1, equation (A.1), the consistency of $\hat{\beta}$ and the mixing LLN applied to D_T,

$$T^{-\frac{1}{2}}\sum q_t(\beta_T) - \bar{D}_T T^{\frac{1}{2}}(\hat{\beta} - \beta_T) \xrightarrow{p} 0$$

Further, $E_T[q_t(\beta_T)] = 0$,

$$\begin{aligned} E_T[T^{-\frac{1}{2}}\sum q_t(\beta_T)]^2 &= T^{-1}\sum E_T[q_t(\beta_T)]^2 \\ &= T^{-1}\sum E_T[\nabla f_t(\beta_T)\nabla' f_t(\beta_T)] \\ &\longrightarrow A \end{aligned}$$

and $E_T|q_t(\beta_T)|^{2r} < \infty$, for some $r > 1$ and all t, T. Hence by the mixing CLT,

$$T^{-\frac{1}{2}}\sum q_t(\beta_T) \xrightarrow{d} N(0, A).$$

That is,

$$T^{\frac{1}{2}}\bar{A}_T^{-\frac{1}{2}}\bar{D}_T(\hat{\beta} - \beta_0) - \bar{A}_T^{-\frac{1}{2}}\bar{D}_T\gamma \xrightarrow{d} N(0, I)$$

Proof of Theorem 3. Firstly,

$$\begin{aligned} \hat{D}_T = \frac{1}{2}\Big[&2(\hat{c}_T T)^{-1}\sum 1(0 < \hat{\epsilon}_t < \hat{c}_T)\nabla f_t \nabla f_t' \\ &+ 2(\hat{c}_T T)^{-1}\sum 1(-\hat{c}_T < \hat{\epsilon}_t < 0)\nabla f_t \nabla f_t'\Big] \end{aligned}$$

so by symmetry, we need only consider

$$2(\hat{c}_T T)^{-1}\sum 1(0 < \hat{\epsilon}_t < \hat{c}_T)\nabla f_t \nabla f_t'$$

This is asymptotically equivalent to

$$\tilde{D}_T = 2(c_T T)^{-1}\sum 1(0 < \hat{\epsilon}_t < \hat{c}_T)\ \nabla f_t \nabla' f_t$$

since $\hat{c}_T/c_T \xrightarrow{p} 1$. The remainder of the proof entails two steps.

i) Show that $\Delta_T = |\tilde{D}_T - 2(c_T T)^{-1}\sum 1(0 < \epsilon_t < c_T)\nabla f_t \nabla' f_t| \xrightarrow{p} 0$ where the derivative is evaluated at any $\beta \in B$, uniformly in β.

ii) Show that $|2(c_T T)^{-1}\sum 1(0 < \epsilon_t < c_T)\nabla f_t \nabla' f_t - \bar{D}_T| \xrightarrow{p} 0$, where the derivative is evaluated at $\hat{\beta}$.

i). Define

$$\Delta_{1T} = 2(c_T T)^{-1}|\sum[1(0 < \epsilon_t < \hat{c}_T) - 1(0 < \epsilon_t < c_T)]\nabla f_t \nabla' f_t|$$

and

$$\Delta_{2T} = 2(c_T T)^{-1}|\sum[1(0 < \hat{\epsilon}_t < \hat{c}_T) - 1(0 < \epsilon_t < \hat{c}_T)\nabla f_t \nabla' f_t|$$

Then

$$\Delta_T \leq \Delta_{1T} + \Delta_{2T}$$

and we consider Δ_{1T} and Δ_{2T} separately. Firstly,

$$\Delta_{1T} \leq 2(c_T T)^{-1}\sum[1(c_T < \epsilon_t < \hat{c}_T) + 1(\hat{c}_T < \epsilon_t < c_T)]\nabla f_t \nabla' f_t$$

Note that the sum of indicator functions is one when ϵ_t lies between c_T and $\hat{c}_T$. Now

$$1(c_T < \epsilon_t < \hat{c}_T) + 1(\hat{c}_T < \epsilon_t < c_T) \leq 1(c_T - \mu_T < \epsilon_t < c_T + \mu_T) + 1(|\hat{c}_T - c_T| > \mu_T)$$

where $\mu_T > 0$ and $c_T \mu_T^{-1} = o(T^{\tau_1})$. Therefore,

$$\Delta_{1T} \leq 2(c_T T)^{-1}\sum 1(c_T - \mu_T < \epsilon_t < c_T + \mu_T)\nabla f_t \nabla' f_t$$

$$+ 2(c_T T)^{-1} 1(|\hat{c}_T - c_T| > \mu_T)\sum \nabla f_t \nabla' f_t$$

The second term is $o_p(1)$ since for any $\eta > 0$,

$$P(c_T^{-1}1(|\hat{c}_T - c_T| > \mu_T) > \eta) \longrightarrow 0$$

with $c_T\mu_T^{-1} = O(T^{\tau_1})$ and since $T^{-1}\sum \nabla f_t \nabla' f_t$ is $O_p(1)$ uniformly in β. Also, for any $\eta > 0$,

$$P_T[(c_T T)^{-1}\sum 1(c_T - \mu_T < \epsilon_t < c_T + \mu_T)\nabla f_t \nabla' f_t > \eta]$$
$$\leq (\eta c_T T)^{-1}\sum E_T\{P_T(c_T - \mu_T < \epsilon_t < c_T + \mu_T | I_{t-1})\nabla f_t \nabla' f_t\} \tag{A.5}$$

But $P_T(c_T - \mu_T < \epsilon_t < c_T + \mu_T | I_{t-1}) \leq 2H\mu_T$ by assumption (A4), where H is the upper bound on $h_{t,T}(\cdot)$. Thus the right hand side of equation (A.5) is arbitrarily small for T large enough. Therefore $\Delta_{1T} \xrightarrow{p} 0$, uniformly in β.

Next, consider Δ_{2T}. Now

$$\Delta_{2T} \leq 2(c_T T)^{-1}\sum 1(|\epsilon_t - \hat{c}_T| < |\hat{\epsilon}_t - \epsilon_t|)\nabla f_t \nabla' f_t$$

where the indicator function is one if $\hat{c}_T$ lies between ϵ_t and $\hat{\epsilon}_t$. Also $|\epsilon_t - \hat{c}_T| < |\hat{\epsilon}_t - \epsilon_t|$ implies $|\epsilon_t - c_T| < |\hat{\epsilon}_t - \epsilon_t| + |c_T - \hat{c}_T|$. Therefore,

$$\Delta_{2T} \leq 2(c_T T)^{-1}\sum 1(|\epsilon_t - c_T| < |\hat{\epsilon}_t - \epsilon_t| + |c_T - \hat{c}_T|)\nabla f_t \nabla' f_t$$

But for $0 < \tau < min[2\tau_1, \tau_2/(1/2 - \tau_2)]$

$$
\begin{aligned}
&P_T\Big\{(c_TT)^{-1}\sum 1(|\epsilon_t - c_T| < |\hat{\epsilon}_t - \epsilon_t| + |c_T - \hat{c}_T|)\nabla f_t \nabla' f_t > \eta\Big\}\\
&\le P_T\Big\{(c_TT)^{-1}\sum 1(|\epsilon_t - c_T| < 2c_T^{1+\tau})\nabla f_t \nabla' f_t > \eta\Big\}\\
&\qquad + P\Big(|\hat{\epsilon}_t - \epsilon_t| > c_T^{1+\tau}\Big)\\
&\qquad + P\Big(|\hat{c}_T - c_T| > c_T^{1+\tau}\Big)\\
&\le (\eta c_T T)^{-1}\sum E_T[P_T(|\epsilon_t - c_T| < 2c_T^{1+\tau}|I_{t-1})\nabla f_t \nabla' f_t]\\
&\qquad + P_T(\| \nabla \epsilon_t \| \, c_T^{-1-\tau} \, \| \hat{\beta} - \beta_0 \| > 1)\\
&\qquad + P_T(c_T^{-\tau}|\hat{c}_T/c_T - 1| > 1)\\
&\le 4(\eta T)^{-1} H c_T^{\tau} \sum E_T(\nabla f_t \nabla' f_t) + o(1) \longrightarrow 0
\end{aligned}
$$

uniformly in β, since $P_T(|\epsilon_t - c_T| < 2c_T^{1+\tau}|I_{t-1}) \le 4Hc_T^{1+\tau}$, uniformly in t, T, $\| \nabla \epsilon_t \|$ is $O_p(1)$, $c_T^{-1-\tau}$ is $o(T^{(\frac{1}{2}-\tau_2)(1+\tau)}) = o(T^{\frac{1}{2}})$ and $c_T^{-\tau} = o(T^{\tau_1})$. Finally,

$$
\Delta_T \le \Delta_{1T} + \Delta_{2T} \xrightarrow{p} 0
$$

uniformly in β.

ii). By a mean value expansion,

$$
\begin{aligned}
2(c_TT)^{-1}\sum 1(0 < \epsilon_t < c_T)\nabla_i f_t \nabla_j f_t|_{\hat{\beta}} &= 2(c_TT)^{-1}\sum 1(0 < \epsilon_t < c_T)\nabla_i f_t^0 \nabla_j f_t^0\\
&\quad + 2(c_TT)^{-1}T^{-\frac{1}{2}}\sum 1(0 < \epsilon_t < c_T)\\
&\quad \times (\nabla^2_{ik} f_t \nabla_j f_t + \nabla_i f_t \nabla^2_{jk} f_t)|_{\dot{\beta}} T^{\frac{1}{2}}(\hat{\beta}_k - \beta_{0k})
\end{aligned}
$$

where $\dot{\beta}$ lies between $\hat{\beta}$ and β_0, and $\hat{\beta}_k$ and β_{0k} are the kth elements of $\hat{\beta}$ and β_0 respectively. Since $c_T^{-1}T^{-\frac{1}{2}} \to 0$, we need only consider the matrix which has the first term on the right hand side as its ijth element. Subtracting $\bar{D}_T$ from this matrix and taking absolute value gives a matrix which is less than or equal to $|\Delta_{3T}| + |\Delta_{4T}|$, where

$$
\Delta_{3T} = 2(c_TT)^{-1}\sum[1(0 < \epsilon_t < c_T) - E_T(1(0 < \epsilon_t < c_T)|I_{t-1})]\nabla_i f_t^0 \nabla_j f_t^0
$$

and

$$\Delta_{4T} = 2(c_T T)^{-1}\sum E_T(1(0 < \epsilon_t < c_T)|I_{t-1})\nabla f_t^0 \nabla' f_t^0 - \bar{D}_T$$

Clearly $E_T(\Delta_{3T}) = 0$ and further,

$$\begin{aligned}\mathrm{var}(\Delta_{3T}) &= 4(c_T T)^{-2}\sum E_T\{[1(0 < \epsilon_t < c_T) - E_T(1(0 < \epsilon_t < c_t)|I_{t-1})]\nabla_i f_t^0 \nabla_j f_t^0\}^2 \\ &\leq 4(c_T T)^{-2}\sum E_T[P_T(0 < \epsilon_t < c_T|I_{t-1})(\nabla_i f_t^0)^2(\nabla_j f_t^0)^2]\end{aligned}$$

expanding the squared term and first taking expectations conditional on I_{t-1}. Next,

$$\begin{aligned}c_T^{-1}P_T(0 < \epsilon_t < c_T|I_{t-1}) &= c_T^{-1}\int_0^{c_T} h_{t,T}(\lambda)d\lambda \\ &= \int_0^1 h_{t,T}(\lambda c_T)d\lambda \\ &= h_t(0) + o(1)\end{aligned}$$

for all t by the bounded convergence theorem and since $h_{t,T}(\lambda) \rightarrow h_t(\lambda)$. Hence

$$\begin{aligned}var(\Delta_{3T}) &\leq 4(c_T T)^{-1}T^{-1}\sum E_T[h_t(0)(\nabla_i f_t^0)^2(\nabla_j f_t^0)^2] + o(1) \\ &\longrightarrow 0\end{aligned}$$

Hence $\Delta_{3T} \xrightarrow{p} 0$. Similarly,

$$\begin{aligned}\Delta_{4T} &= 2T^{-1}\sum h_t(0)\nabla f_t^0 \nabla' f_t^0 - \bar{D}_T + o(1) \\ &\xrightarrow{p} 0\end{aligned}$$

by the LLN. Combining Δ_{3T} and Δ_{4T} we have

$$|2(c_T T)^{-1}\sum 1(0 < \epsilon_t < c_T)\nabla f_t^0 \nabla' f_t^0 - \bar{D}_T| \leq |\Delta_{3T}| + |\Delta_{4T}| \xrightarrow{p} 0$$

which completes the proof.

Proof of Lemma 1. From Lemma A.1,

$$Z_T(\beta,\beta_T) = \{|g_T(\delta) - g_T(0) + D_T\delta + O(T^{-\frac{1}{2}} \parallel \delta \parallel^2)|\}/\{1 + \parallel D_T\delta + O(T^{\frac{1}{2}} \parallel \delta \parallel^2) \parallel\}$$

Notice that for $\parallel \delta \parallel < M$, the denominator is bounded. Hence, since for any M, $M/\sqrt{T} < d_0$ for T large enough, Lemma A.1 implies that

$$\sup_{\|\delta\|<M} \parallel g_T(\delta) - g_T(0) + D_T\delta \parallel = o_p(1)$$

as required.

Proof of Theorem 4. The proof has two main parts.

i) Show that

$$\Delta_{5T} = T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1) - T^{-\frac{1}{2}}\sum(z_t^* - Z_T^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1) - K_T\gamma_1 \xrightarrow{p} 0$$

ii) Show that

$$T^{-\frac{1}{2}}\sum(z_t^* - Z_T^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1) \xrightarrow{d} N(0,\Lambda)$$

The form of the LM test then follows from the consistency of $\hat{\Lambda}_T$ for Λ, which is implied by $\hat{\beta} \xrightarrow{p} \beta_0$, the LLN, and Lemma 2.6 of White (1980).[7]

i) By the triangle inequality,

$$|\Delta_{5T}| \leq |\Delta_{6T}| + |\Delta_{7T}| + |\Delta_{8T}|$$

where

$$\begin{aligned}
\Delta_{6T} &= T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - 1) \\
&\quad - T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t|\omega_{t,T}^{-1} - 1) - K_T\gamma_1 \\
\Delta_{7T} &= T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t|\omega_{t,T}^{-1} - 1) \\
&\quad - T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1)
\end{aligned}$$

and

$$\Delta_{8T} = T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1)$$
$$- T^{-\frac{1}{2}}\sum(z_t^* - Z_T^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1)$$

Consider Δ_{6T}. Now

$$\begin{aligned}\Delta_{6T} &= T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)\left(|\hat{\epsilon}_t|\hat{\omega}_0^{-1} - |\hat{\epsilon}_t|\omega_{t,T}^{-1}\right) - K_T\gamma_1 \\ &= T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)|\hat{\epsilon}_t|(\hat{\omega}_0^{-1} - \omega_{0,T}^{-1}) \\ &\quad + T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)|\hat{\epsilon}_t|(\omega_{0,T}^{-1} - \omega_{t,T}^{-1}) - K_T\gamma_1 \\ &= \Delta_{9T} + \Delta_{10T} - K_T\gamma_1 \quad \text{(say)}\end{aligned}$$

But

$$\Delta_{9T} = -\hat{\omega}_0^{-1}\omega_{0,T}^{-1}T^{\frac{1}{2}}(\hat{\omega}_0 - \omega_{0,T})T^{-1}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t| - \omega_{0,T})$$

Further, $\omega_{0,T} = \omega(\alpha_{1,T}) = \omega(\alpha_1) + \omega'(\dot{\alpha}_1)(\alpha_{1,T} - \alpha_1) \rightarrow \omega(\alpha_1)$, where $\dot{\alpha}_1$ lies between $\alpha_{1,T}$ and α_1. Similarly, as $T \rightarrow \infty$ and $\hat{\beta} \rightarrow \beta_0$, $\hat{\omega}_0 \xrightarrow{p} \omega_0$ using Lemma 2.6 of White (1980). Next,

$$T^{\frac{1}{2}}(\hat{\omega}_0 - \omega_{0,T}) = T^{-\frac{1}{2}}\sum(|\hat{\epsilon}_t| - |\epsilon_t|) + T^{-\frac{1}{2}}\sum(|\epsilon_t| - \omega_{t,T}) + T^{-\frac{1}{2}}\sum(\omega_{t,T} - \omega_{0,T})$$

Writing the first term as a function of δ and using the method leading to Lemma 2, we have, for $L > 0$,

$$\sup_{\|\delta\|<L} |T^{-\frac{1}{2}}\sum(|\epsilon_t - \nabla'\tilde{f}_t\ \delta/\sqrt{T}| - |\epsilon_t|)| \xrightarrow{p} 0$$

and hence $T^{-\frac{1}{2}}\sum(|\hat{\epsilon}_t| - |\epsilon_t|) \xrightarrow{p} 0$, since $P(\|\ \hat{\delta}\ \|< L) > 0$. Further, $T^{-\frac{1}{2}}\sum(|\epsilon_t| - \omega_{t,T})$ is $O_p(1)$ since $var[T^{-\frac{1}{2}}\sum(|\epsilon_t| - \omega_{t,T}] < \infty$ under assumption (A14) part iii), and

$$T^{-\frac{1}{2}}\sum(\omega_{t,T} - \omega_{0,T}) = T^{-1}\sum\omega'(z_t'\alpha_T^*)z_t'\gamma_1$$
$$\longrightarrow \omega'(\alpha_1) \lim_{T\rightarrow\infty} T^{-1}\sum E[z_t']\gamma_1 < \infty$$

where α_T^* lies between α_T and α. Hence $T^{\frac{1}{2}}(\hat{\omega}_0 - \omega_{0,T}) = O_p(1)$. Finally,

$$T^{-1}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t| - \omega_{0,T}) = T^{-1}\sum\hat{z}_t^*(|\hat{\epsilon}_t| - \omega_{0,T}) - \bar{Z}^* T^{-1}\sum(|\hat{\epsilon}_t| - \omega_{0,T}) \xrightarrow{p} 0$$

using Lemma 2.6 of White (1980). Hence, $\Delta_{9T} \xrightarrow{p} 0$. Similarly, applying a mean value expansion to $\omega_{t,T}^{-1}$ in Δ_{10T} implies that

$$\Delta_{10T} = T^{-1}\sum(\hat{z}_t^* - \bar{Z}^*)|\hat{\epsilon}_t|\omega(z_t'\alpha_T^*)^{-2}\omega'(z_t'\alpha_T^*)z_t'\gamma_1$$

But by Lemma 2.6 of White (1980) and the continuity of $\omega(\cdot)$ and $\omega'(\cdot)$ in α, $\Delta_{10T} - K_T\gamma_1 \xrightarrow{p} 0$.

Next, consider Δ_{7T}. Now

$$\begin{aligned}\Delta_{7T} &= T^{-\frac{1}{2}}\sum(\hat{z}_t^* - \bar{Z}^*)(|\hat{\epsilon}_t| - |\epsilon_t|)\omega_{t,T}^{-1} \\ &= T^{-\frac{1}{2}}\sum\hat{z}_t^*(|\hat{\epsilon}_t| - |\epsilon_t|)\omega_{t,T}^{-1} - \bar{Z}^* T^{-\frac{1}{2}}\sum(|\hat{\epsilon}_t| - |\epsilon_t|)\omega_{t,T}^{-1}\end{aligned}$$

Consider the first term in Δ_{7T}, or equivalently,

$$\Delta_{11T}(\delta) \equiv T^{-\frac{1}{2}}\sum z_t^*(\delta)(|\epsilon_t - \nabla'\tilde{f}_t\ \delta/\sqrt{T}| - |\epsilon_t|)\omega_{t,T}^{-1}$$

The gradient of this with respect to δ is $\Delta_{12T} + \Delta_{13T}$, where

$$\Delta_{12T} = T^{-1}\sum z_t^*(\delta)\nabla' f_t\psi(\epsilon_t - \nabla'\tilde{f}_t\ \delta/\sqrt{T})\omega_{t,T}^{-1}$$

and

$$\Delta_{13T} = T^{-\frac{1}{2}}\sum\partial z_t^*/\partial\delta(|\epsilon_t - \nabla'\tilde{f}_t\ \delta/\sqrt{T}| - |\epsilon_t|)\omega_{t,T}^{-1}$$

In Δ_{12T}, $T^{-\frac{1}{2}}z_t^*(\delta)$ is a.s. $o(1)$ and from Lemma 1, the rest of the expression is $O_p(1)$. Hence, following Lemma 2, $\Delta_{12T} \xrightarrow{p} 0$. Also,

$$(\Delta_{13T})_{ij} \leq T^{-1}\sum|\partial z_{it}/\partial\delta_j|\ |\nabla'\tilde{f}_t\delta| \xrightarrow{p} 0$$

under assumption (A14), iv), since $\partial z_{it}/\partial\delta_j = T^{-\frac{1}{2}}\partial z_{it}/\partial\beta_j$. Similarly, the second term in Δ_{7T} is $o_p(1)$.

Next, consider Δ_{8T}. Now

$$\Delta_{8T} = T^{-\frac{1}{2}}\sum(\hat{z}_t^* - z_t^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1) - (\bar{Z}^* - Z_T^*)T^{-\frac{1}{2}}\sum(|\epsilon_t|\omega_{t,T}^{-1} - 1)$$

Clearly, the second term on the right hand side is $o_p(1)$. Write the first term as

$$T^{-\frac{1}{2}}\sum[z_{it}^*(\hat{\beta}) - \tilde{z}_{it}(\hat{\beta}) + \tilde{z}_{it}(\hat{\beta}) - \tilde{z}_{it}(\beta_0) + \tilde{z}_{it}(\beta_0) - z_{it}^*(\beta_0)](|\epsilon_t|\omega_{t,T}^{-1} - 1)$$

By assumption (A14) part vii), this is $o_p(1)$ since $T^{-\frac{1}{2}}(|\epsilon_t|\omega_{t,T}^{-1} - 1)$ is a.s. $o(1)$.

ii) Consider the distribution of $T^{-\frac{1}{2}}\sum r_{t,T}$, where $r_{t,T} = (z_t^* - Z_T^*)(|\epsilon_t|\omega_{t,T}^{-1} - 1)$. Clearly $E_T[r_{t,T}] = 0$ since $E_T\left[|\epsilon_t| \,\middle|\, \mathcal{F}_{t-1}\right] = \omega_{t,T}$. Also,

$$\begin{aligned} E_T[T^{-\frac{1}{2}}\sum r_{t,T}][T^{-\frac{1}{2}}\sum r_{t,T}]' &= E_T[T^{-1}\sum r_{t,T}r_{t,T}'] \\ &= \Lambda_T \\ &\longrightarrow \Lambda \end{aligned}$$

by Assumption (A15). Applying the mixing CLT completes the proof.

FOOTNOTES

1. The restriction that x_t contains only a finite number of lags excludes moving average models, for example. As White (1986) notes, however, the results in Gallant and White (1986) may be used to allow an infinite number of lags.

2. Implicitly, all events are assumed to take place on a complete probability space.

3. This basically ensures that $\bar{\sigma}_T^2(\beta)$ is not too flat in the neighborhood of β_0.

4. Inspection of the proof of the uniform Law of Large Numbers, Lemma A.7 of Newey (1985), shows that, in his notation,

$$plim \sup_B |T^{-1}\sum\omega_t(b) - T^{-1}\sum\bar{E}\omega_t(b)| = 0$$

This is implicit in the proof, but slightly more general than his conclusion. The Central Limit Theorem is essentially his Lemma A.8.

5. If the null hypothesis is of the form $S(\beta_0) = 0$ where ∇S evaluated at β_0 has rank $q \leq k$, then in the test statistics, $S(\hat{\beta})$ is used instead of $(R\hat{\beta} - r)$ and $\nabla S(\hat{\beta})$ is used instead of R.

6. To use the analogy in Koenker and Bassett (1982, p. 1583), perhaps the statistical weather need not be too inclement.

7. Lemma 2.6 of White (1980) gives conditions, satisfied by $\hat{\beta} \xrightarrow{p} \beta_0$ and the uniform mixing LLN, for using a sample average, *e.g.*, $\hat{\Lambda}_T$, to consistently estimate the corresponding expectation, Λ in this case.

REFERENCES

Amemiya, T. Two Stage Least Absolute Deviations Estimators. *Econometrica* 50 (1982): 689-711.

Bassett, G. and R. Koenker. Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Society* 73 (1978): 618-622.

Belsley, D.A., Kuh, E. and R.E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley, 1979.

Bickel, P.J. One Step Huber Estimation in the Linear Model. *Journal of the American Statistical Association* 70 (1975): 428-434.

Bickel, P.J. Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity. *Annals of Statistics* 6 (1978): 266-291.

Bloomfield, P. and W.L. Steiger. *Least Absolute Deviations.* Stuttgart: Birkhäuser, 1983.

Breusch, T.S. and A.R. Pagan. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47 (1979): 1287-1294.

Dijkstra, T.K. Misspecification and the Choice of Estimators, A Heuristic Approach. In T.K. Dijkstra (ed.) *Misspecification Analysis.* Berlin:Springer-Verlag, 1984.

Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of Inflationary Expectations. *Econometrica* 50 (1982): 987-1007.

Engle, R.F. A General Approach to Lagrange Multiplier Model Diagnostics. *Journal of Econometrics* 20 (1982): 83-104.

Gallant, A.R. and H. White. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. University of California, San Diego, Department of Economics Discussion Paper (1986).

Glejser, H. A New Test for Heteroscedasticity. *Journal of the American Statistical Association* 64 (1969): 316-323.

Huber, P.J. The Behavior of Maximum Likelihood Estimates under Nonstandard Conditons. In L.M. Le Cam and J. Neyman (eds.) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1.* Berkeley: University of California Press, 1967.

Jaeckel, L.A. Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *Annals of Mathematical Statistics* 42 (1972): 1328-1338.

Jennrich, R.I. Asymptotic Properties of Nonlinear Least Squares Estimators. *Annals of Mathematical Statistics* 40 (1969): 633-643.

Koenker, R. Robust Methods in Econometrics. *Econometric Reviews* 1 (1982): 213-255.

Koenker, R. and G. Bassett. Tests of Linear Hypotheses and ℓ_1 Estimation. *Econometrica* 50 (1982): 1577-1583.

Martin, R.D. Robust Methods for Time Series. In D.F. Findley (e.d.) *Applied Time Series Analysis II.* New York: Academic Press, 1981.

Newey, W.K. and J.L. Powell. Asymmetric Least Squares Estimation and Testing. Mimeograph, Princeton University Department of Economics (1986).

Oberhofer, W. The Consistency of Nonlinear Regression Minimizing the L_1-Norm. *Annals of Statistics* 10 (1982), 316-319.

Powell, J.L. Least Absolute Deviations Estimation for the Censored Regression Model. *Journal of Econometrics* 25 (1984): 303-325.

Weiss, A.A. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, forthcoming (1986a).

Weiss, A.A. On the Stability of a Heteroscedastic Process. *Journal of Time Series Analysis*, forthcoming (1986b).

Weiss, A.A. and A.P. Andersen. Estimation Time Series Models using the Relevant Forecast Evaluation Criterion. *Journal of the Royal Statistical Society* A147 (1984): 484-487.

White, H. Nonlinear Regression on Cross-Section Data. *Econometrica* 48 (1980): 721-746.

White, H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50 (1982): 1-25.

White, H. *Asymptotic Theory for Econometricians.* New York: Academic Press, 1984.

White H. Specification Testing in Dynamic Models. University of California, San Diego, Department of Economics Discussion Paper 85-30R (1986).

White, H. and I. Domowitz. Nonlinear Regression with Dependent Observations. *Econometrica* 52 (1984): 143-161.

REPORTS IN THIS SERIES

M8604: RICHARD H. DAY and T.-Y. LIN: Irregular Fluctuations and Comparative Policy Dynamics in the Keynesian Regime.
M8605: RICHARD M. GOODWIN: The Economy as an Evolutionary Pulsator.
M8606: Z.M. BERREBI and J. SILBER: Interquantile Differences, Income Inequality Measurement and the Gini Concentration Index.
M8607: JOSEPH G. HIRSCHBERG: A Comparison of Three Flexible Cost Functions Using Establishment Level Electricity Use Data.
M8608: IAN E. NOVOS and MICHAEL WALDMAN: Complementarity and Partial Nonexcludability: An Analysis of the Software/Computer Market.
M8609: ARNOLD ZELLNER: Further Results on Bayesian Minimum Expected Loss (MELO) Estimates and Posterior Distributions for Structural Coefficients.
M8610: ARNOLD ZELLNER: Bayesian Analysis in Econometrics.
M8611: ARNOLD ZELLNER: A Tale of Forecasting 1001 Series: The Bayesian Knight Strikes Again.
M8612: ARNOLD ZELLNER: Biased Predictors, Rationality and the Evaluation of Forecasts.
M8613: ARNOLD ZELLNER: Bayesian Estimation and Prediction Using Asymmetric Loss Functions.
M8614: RICHARD H. DAY, SUDIPTO DASGUPTA, SAMAR K. DATTA, and JEFFREY B. NUGENT: A Note on Instability in Rural-Urban Migration.
M8615: DENNIS J. AIGNER, FRIEDRICH SCHNEIDER and DAMAYANTI GHOSH: Me and My Shadow: Estimating the Size of the U.S. Underground Economy from Time Series Data.
M8616: ROBERT KALABA and LEIGH TESFATSION: Flexible Least Squares.
M8617: JOHN M. VEITCH: Wavering Virgins, the Deficit and Gramm-Rudman.
M8618: TIMUR KURAN: Chameleon Voters and Public Choice.
M8619: RICHARD H. DAY: The Evolving Economy.
M8620: JACQUES SILBER: Factor Components, Income Classes and the Computation of the Gini Index of Inequality.
M8621: MARIO D. TELLO: Manufactured Exports from Developing Countries and Government Policy Under Imperfect International Markets: Corporate Income Taxes and Exchange Rate.
M8622: DANIEL F. SPULBER: The Second Best Core.
M8623: DANIEL F. SPULBER: Variety and Efficiency: Monopolistic Competition with Nonlinear Price Schedules.
M8624: ABDELHAMID MAHBOB and JEFFREY B. NUGENT: The North's Demand for Southern Exports and Immiserization of the South.
M8625: IAN E. NOVOS and MICHAEL WALDMAN: The Emergence of Copying Technologies: What Have We Learned.
M8626: REZAUL KHANDKER: Offer Heterogeneity in a Two State Model of Sequential Search.
M8627: ROBERT KALABA, THOMAS PLUM, and LEIGH TESFATSION: Automation of Nested Matrix and Derivative Operations.
M8628: RICHARD H. DAY and KYOO-HONG KIM: A Note on Nonperiodic Demoeconomic Fluctuations with Positive Measure.
M8629: JOHN M. VEITCH: The Optimality of Loan Commitment Contracts in the Extension of Credit.
M8630: ANDREW A. WEISS: Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation.
M8631: RICHARD H. DAY and WAYNE SHAFER: Ergodic Fluctuations in Deterministic Economic Models.