USC

MRG WP #8648

MODELLING RESEARCH GROUP

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

db

MODELLING RESEARCH GROUP

A COMPARISON OF ORDINARY LEAST SQUARES AND LEAST ABSOLUTE ERROR ESTIMATION

ANDREW A. WEISS

MRG WORKING PAPER #M8648

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

A COMPARISON OF ORDINARY LEAST SQUARES AND LEAST ABSOLUTE ERROR ESTIMATION

ANDREW A. WEISS

MRG WORKING PAPER #M8648

December 1986

ABSTRACT

In a linear dynamic model with heteroscedastic errors, we compare some aspects of ordinary least squares and least absolute error estimation. After deriving the properties of the estimators and the Wald, Lagrange multiplier and Likelihood ratio tests under a local alternative, we derive the Hausman test comparing the estimators. From this, and the equivalent generalized method of moment tests, we obtain as special cases tests for specification and symmetry based on the signs of the residuals from ordinary least squares. We also show that in the presence of heteroscedasticity, asymmetry can affect the estimates of all the parameters, not just the constant term.

1. INTRODUCTION

In a recent paper, Brown and Kildea (1979) discuss estimators and tests based on linear combinations of signs of residuals. In the linear regression model, a special case of the estimator is obtained by using the matrix of regressors to form the linear combination. This is simply the least absolute error (LAE) estimator. The main test they discuss takes the signs of the residuals from ordinary least squares (OLS) and regresses these on the matrix of regressors. The test is therefore against alternatives such as asymmetry in the errors which may distort the balance of signs of these residuals.

These results suggests some relationships between OLS and LAE and in this paper we investigate these further. For example, comparing OLS and LAE in a Hausman test gives, after some simplification, the test suggested by Brown and Kildea. We consider these tests and estimators in more detail and in more general settings. For example, the model can be dynamic and there may be heteroscedasticity in the errors.

An important consideration turns out to be whether or not heteroscedasticity is in fact present in the errors. Of course, the covariance matrices of OLS and LAE are functions of the variances (the latter through the height of the error density at the zero). But heteroscedasticity also affects the power of the tests and determines, for example, whether asymmetry in the errors causes inconsistency in only the constant in the regression, or in all the regression coefficients. In the case of a distorted normal distribution, the asymmetry is equivalent to the omission of a variance term from the regression equation. Such terms are often used as proxies for the appearance of risk in economic relationships (see, for example, Pagan and Ullah (1986)).

In section 2 of the paper we introduce the model and discuss the assumptions imposed. Important amongst these is an assumption of a sequence of local alternatives subject to Pitman drift. In section 3 we review of the consistency and asymptotic normality of the OLS estimator as an introduction to the discussion of asymptotic theory for the LAE estimator. This is presented in section 4. Next, in section 5 we derive the Hausman test and the equivalent generalized method of moments (GMM) tests while in section 6 we consider some special cases of the process. These highlight the role of symmetry and heteroscedasticity. Finally, in section 7 we offer some concluding comments.

2. The Model

For simplicity, we suppose that the researcher estimates the following linear regression model

$$y_t = x_t'\beta + \epsilon_t(\beta) \qquad (t = 1, \ldots, T) \tag{2.1}$$

where x_t is a $(k \times 1)$ vector of predetermined variables, with first element 1 and possibly including lags of y_t, β is a $(k \times 1)$ vector of parameters restricted to lie in a compact set, and $\epsilon_t(\beta) = y_t - x_t'\beta$ is the tth residual. This model however may not be correctly specified.

We assume the data is generated by the process specified in Assumption 1. This is similar to the framework used by Newey (1985) and a discussion follows the statement of the assumption.

ASSUMPTION 1: (the Data Generation Process). The observed data $y_t (t = 1, \ldots, T)$ is generated by the process

$$y_t = x_t'\beta_0 + \eta_t \tag{2.2}$$

which satisfies the following conditions:

i) Define I_{t-1} as the information set (σ-algebra) generated by $x_{t-i}, i \geq 0$, and $\eta_{t-i}, i > 0$, and $h_{t,T}(\lambda)$ as the density function of η_t conditional on I_{t-1}. $h_{t,T}(\lambda)$ is indexed by an $(\ell \times 1)$ vector of parameters c_T such that for some (possibly) non-zero $(\ell \times 1)$ vector γ,

$$c_T = c_0 + \gamma/\sqrt{T} \tag{2.3}$$

The conditional density at c_0 is denoted $h_t(\lambda)$ and the parameters c_0 are such that at $c = c_0$, η_t has conditional mean and median zero. That is, at $c = c_0$,

$$E(\eta_t \mid I_{t-1}) = 0 \tag{2.4}$$

and, with $\psi(x) =$ sign x,

$$\begin{aligned} E\big(\psi(\eta_t) \mid I_{t-1}\big) &= 1 - 2P(\eta_t < 0) \\ &= 0 \end{aligned} \tag{2.5}$$

where expectations are taken with respect to h_t.

ii) Assume that with respect to c, $h_{t,T}$ is continuous for almost all x_t uniformly in t. With respect to λ, assume that $h_{t,T}(\lambda)$ is bounded from above, has only a finite number of discontinuities and that the associated conditional distribution function, $H_{t,T}$, is continuous, all uniformly in t, T.

iii) Assume that for almost all x_t, there exist $p_1, h > 0$ such that $P\big(h_t(0) \geq h\big) > p_1$, uniformly in t. Further, assume that each h_t is Lipschitz continuous, *i.e.*, for all λ_1, λ_2 and some finite constant $L_0 > 0$,

$$|h_t(\lambda_2) - h_t(\lambda_1)| \leq L_0 |\lambda_2 - \lambda_1| \tag{2.6}$$

uniformly in t.

iv) Let $u_{t,T}$ be the unconditional density of (x_t, η_t). Then $u_{t,T}$ is continuously differentiable in c uniformly in t, T. The same condition applies to the joint densities of (x_t, η_t) and (x_s, η_s), $s \neq t$.

Consider now the implications of Assumption 1. Firstly, notice that the data generation process is allowed to depend upon T through the parameters c_T and that extra T subscripts on y_t, z_t and η_t have been dropped. The c_T represent parameters which affect the correctness of the specification of the model. That is, at $c = c_T$,

$$E_T(\eta_t \mid I_{t-1}) \neq 0 \tag{2.7}$$

where E_T means the expectation is taken at $c = c_T$. Therefore at $c = c_T$, the model is misspecified, and the c_T can represent the coefficient on an omitted variable or a vector of correlations between the regressors and errors, for example. As $T \to \infty$, the sequence of locally misspecified alternatives converges to a correctly specified model, as indicated by

(2.4) and (2.5). The Pitman drift in (2.3) helps ensure that the test statistics do not have degenerate distributions. Equation (2.4) implies that OLS will be consistent and (2.5) is the orthogonality condition for LAE estimation. Under the local misspecification, the median of η_t is also non-zero, and

$$E_T[\psi(\eta_T) \mid I_{t-1}] \neq 0. \tag{2.8}$$

As we shall see, however, the effects of the misspecification on OLS and LAE will differ, implying that the difference between them can be used to test the model.

Next, note that we have left implicit a number of technical conditions which do not add to the discussion. That is, we also assume, without further comment, that all events take place on a complete probability space, that all relevant functions are both measurable and appropriately dominated, and that the process satisfies the required dependence conditions. The actual assumptions needed are equivalent to those discussed extensively in Gallant and White (1986), and a non-rigorous discussion of the type of conditions needed is given in Domowitz (1985). Briefly, however, the dominance conditions imply that the various expectations introduced below exist and the moments needed can be inferred from these.[1] Similarly, the dependence conditions can be quite general. Examples range from the $\{y_t, x_t, \eta_t\}$ being independent across time, to the variables satisfying a mixing condition which allows a substantial degree of dependence between observations close together in time.[2] In all cases, however, observations far apart in time will be essentially independent. The important feature required here is that the conditions imply the existence of a uniform Law of Large Numbers (LLN) and a Central Limit Theorem (CLT). We shall therefore assume that these results exist, and apply them when needed.[3,4]

The conditions given in Assumption 1 are those which specifically describe the process and are needed to discuss LAE estimation in particular. Other assumptions, especially concerning the invertibility of certain matrices, are introduced below. We also note that our assumptions are not the weakest set available; some are made to help simplify the analysis.

Returning to the discussion of Assumption 1, notice that no explicit restrictions have been placed on η_t in terms of heteroscedasticity. Then provided, of course, the necessary unconditional moments of η_t exist, the variance of η_t conditional on I_{t-1} can depend fairly generally on terms measurable with respect to I_{t-1}, *e.g.*, lagged squared errors or elements of x_t squared.[5]

To analyse LAE in the presence of heteroscedasticity, it is also convenient to assume that the height of the conditional density of η_t at zero, *i.e.*, $h_t(0)$, depends upon the conditional variance, but upon no other moments of the distribution. Homoscedasticity then implies that $h_t(0) = h(0)$ for all t, *i.e.*, the height of the density function at zero is constant with respect to t. Similarly, the conditions in part ii) of Assumption 1 bounding $h_{t,T}(0)$ will then be satisfied by appropriately restricting the conditional variance.

The remaining conditions in parts ii)-iv) ensure that the density of the errors is sufficiently well-behaved. For example, limiting the number of discontinuities of $h_{t,T}$ implies that the density does not change too quickly in the neighborhood of the origin. This is important otherwise it would be difficult to estimate the covariance matrix of the LAE estimator, which depends upon the $h_t(0)$. Similarly, the conditions in part iv) ensure that expectations taken with respect to $u_{t,T}$ converge to those taken with respect to u_t (the unconditional density at $c = c_0$) as $c_T \to c_0$.

3. OLS Estimation

To show the relationships between OLS and LAE estimation, we begin by quickly reviewing OLS estimation of (2.1). The OLS estimator, $\hat{\beta}_{LS}$, is obtained by minimizing

$$Q_{LS}(\beta) = \sum_{t=1}^{T} \epsilon_t(\beta)^2 \tag{3.1}$$

with respect to β. In future, we shall drop the indices from summations, and we note that any implicit conditioning on pre-sample values if x_t contains lagged variables does not affect the asymptotic results.

The first-order conditions corresponding to (3.1) are

$$\begin{aligned}\nabla Q_{LS}(\beta) &= -\Sigma x_t \epsilon_t(\hat{\beta}_{LS}) \\ &= -X'\epsilon(\hat{\beta}_{LS}) = 0\end{aligned} \tag{3.2}$$

where $\nabla \equiv \partial/\partial\beta$, $X' = (x_1 \ldots x_T)$ and $\epsilon(\beta) = \big(\epsilon_1(\beta) \ldots \epsilon_T(\beta)\big)'$. We make the usual type of assumption on the $X'X$ matrix. That is, we assume that

$$A_T = E_T\left(\frac{X'X}{T}\right) \tag{3.3}$$

converges to some positive definite matrix A. This implies that $(X'X/T)^{-1}$ exists for T large enough. It is then easy to show that $\hat{\beta}_{LS} \xrightarrow{p} \beta_0$.

Asymptotic normality is based on a linearization of the gradient $T^{-\frac{1}{2}}X'\epsilon(\hat{\beta}_{LS})$ about $\beta = \beta_0$. That is,

$$T^{-\frac{1}{2}}X'\epsilon(\hat{\beta}_{LS}) = T^{-\frac{1}{2}}X'\epsilon(\beta_0) - (T^{-1}X'X)T^{\frac{1}{2}}(\hat{\beta}_{LS} - \beta_0) \tag{3.4}$$

or

$$(T^{-1}X'X)T^{\frac{1}{2}}(\hat{\beta}_{LS} - \beta_0) = T^{-\frac{1}{2}}X'\epsilon(\beta_0) \tag{3.5}$$

But

$$T^{-\frac{1}{2}}X'\epsilon(\beta_0) = T^{-\frac{1}{2}}\Sigma\{x_t\epsilon_t(\beta_0) - E_T\big(x_t\epsilon_t(\beta_0)\big)\} + T^{-\frac{1}{2}}\Sigma E_T\big(x_t\epsilon_t(\beta_0)\big) \tag{3.6}$$

Then following Newey (1985) and Newey and Powell (1985), the CLT is applied to the first term in (3.6), and the second term converges to $K_1\gamma$, where $K_1 = \lim_{T\to\infty} K_{1T}$,

$$K_{1T} = T^{-1}\Sigma\partial E_T\big(x_t\epsilon_t(\beta_0)\big)/\partial c \tag{3.7}$$

the derivative is evaluated at $c = c_0$ and the limit is assumed to exist. Thus K_{1T} simply measures the change in the orthogonality conditions in the neighborhood of the correct specification. An alternative expression of K_{1T} is

$$K_{1T} = T^{-1}\Sigma E[x_t\eta_t\partial \ln u_t/\partial c'] + o(1) \tag{3.8}$$

where $\partial \ln u_t / \partial c$ means $\partial \ln u_{t,T} / \partial c$ evaluated at $c = c_0$. The CLT implies that

$$B_T^{-\frac{1}{2}} T^{-\frac{1}{2}} \Sigma \{ x_t \epsilon_t(\beta_0) - E_T(x_t \epsilon_t(\beta_0)) \} \xrightarrow{d} N(0, I_k) \tag{3.9}$$

where

$$\begin{aligned} B_T &= T^{-1} \Sigma E(\eta_t^2 x_t x_t') \\ &= T^{-1} \Sigma E_T(\epsilon_t(\beta_0)^2 x_t x_t') + o(1) \end{aligned} \tag{3.10}$$

B_T is also assumed to converge to some positive definite limit B. If, for example, there exists $\alpha_0 > 0$ such that $E_T(\eta_t^2 | I_{t-1}) \geq \alpha_0$ for all t, T, as is the case in say the ARCH model (Engle (1982b), Weiss (1986a)), then $B \geq \alpha_0 A$ and B is positive definite.

Combining (3.5)-(3.10), we obtain

$$B_T^{-\frac{1}{2}} A_T T^{\frac{1}{2}} (\hat{\beta}_{LS} - \beta_0) - B_T^{-\frac{1}{2}} K_{1T} \gamma \xrightarrow{d} N(0, I_k) \tag{3.11}$$

The usual consistent estimates of A_T and B_T are

$$\hat{A}_T = T^{-1} \Sigma x_t x_t' \tag{3.12}$$

and

$$\hat{B}_T = T^{-1} \Sigma \epsilon_t(\hat{\beta}_{LS})^2 x_t x_t' \tag{3.13}$$

(see, for example, White (1984, Chapter 6)). This allows χ^2 inference using $\hat{\beta}_{LS}$ although it is convenient to defer the discussion of this inference, *i.e.*, the Wald, Lagrange multiplier (LM) and Likelihood ratio (LR) tests based on (3.11), until after the corresponding tests based on LAE estimation have been introduced. We now turn to the discussion of LAE estimation.

4. LAE Estimation

The LAE estimator, $\hat{\beta}_{LA}$, is obtained by minimizing

$$Q_{LA}(\beta) = \Sigma |\epsilon_t(\beta)| \tag{4.1}$$

with respect to β. As is well known, $\hat{\beta}_{LA}$ is not unique. However the different estimates will be close and this does not affect the asymptotic results. Since the model is only locally misspecified, it is straightforward to show that $\hat{\beta}_{LA} \xrightarrow{p} \beta_0$. See, for example, Weiss (1986b, Theorem 1) or Powell (1984, Theorem 2).

Next, the appropriately normalized first-order conditions associated with (4.1) are

$$\begin{aligned} T^{-\frac{1}{2}} \nabla Q_{LA}(\hat{\beta}_{LA}) &= -T^{-\frac{1}{2}} \Sigma x_t \psi(\epsilon_t(\hat{\beta}_{LA})) \\ &= -T^{-\frac{1}{2}} X' \bar{\Psi}(\hat{\beta}_{LA}) \end{aligned} \tag{4.2}$$

where $\bar{\Psi}(\beta) = \Big(\psi(\epsilon_1(\beta)) \ldots \psi(\epsilon_T(\beta))\Big)'$. It follows from Ruppert and Carroll (1980, Lemma A.2) that

$$T^{-\frac{1}{2}} \nabla Q_{LA}(\hat{\beta}_{LA}) \xrightarrow{p} 0. \tag{4.3}$$

Because Q_{LA} is not continuously differentiable in β, a mean value expansion cannot be applied to linearize the first-order conditions as in (3.4). Instead, we follow the approach developed by Huber (1967), which approximates the discontinuous gradient by its continuously differentiable expectation. This can then be written in terms of the LAE estimator. A by-product is an asymptotic linearization of the first-order conditions which turns out to be useful for the comparison of OLS and LAE in section 5. These results are summarized in the next theorem and corollary. (See also Powell (1984) and Weiss (1986b).)

Define the equivalent of K_{1T}, *i.e.*, let

$$K_{2T} = T^{-1} \Sigma \partial E_T \big[x_t \psi(\epsilon_t(\beta_0))\big] / \partial c \tag{4.4}$$

where the derivative is evaluated at $c = c_0$. Assume that as $T \to \infty$, $K_{2T} \to K_2$, for some matrix K_2. Also define

$$\begin{aligned} D_T &= 2T^{-1} \Sigma E(h_t(0) x_t x_t') \\ &= 2T^{-1} \Sigma E_T(h_{t,T}(0) x_t x_t') + o(1) \end{aligned} \tag{4.5}$$

and as with A_T and B_T assume that D_T converges to some positive definite limit D. Then we have (with proofs in the Mathematical Appendix)

THEOREM 1: *(asymptotic normality)*

$$A_T^{-\frac{1}{2}} D_T T^{\frac{1}{2}} (\hat{\beta}_{LA} - \beta_0) - A_T^{-1} K_{2T} \gamma \xrightarrow{d} N(0, I_k) \tag{4.6}$$

An alternative expression for K_{2T} is

$$K_{2T} = 2T^{-1} \Sigma E[x_t \partial H_t / \partial c'] + o(1) \tag{4.7}$$

where $\partial H_t/\partial c$ is the derivative of $H_{t,T}(0)$ evaluated at $c = c_0$. Note also that D_T is positve definite for T large enough because A is positive definite and $P(h_t(0) \geq h) > 0$.

COROLLARY 2: *(asymptotic linearity of the gradient) For all $M > 0$,*

$$\sup_{T^{\frac{1}{2}} \|\beta - \beta_0\| \leq M} \| T^{-\frac{1}{2}} \Sigma x_t \psi(\epsilon_t(\beta)) - T^{-\frac{1}{2}} \Sigma x_t \psi(\epsilon_t(\beta_0)) + D_T T^{\frac{1}{2}} (\beta - \beta_0) \| = o_p(1) \tag{4.8}$$

Corollary 2 is also equivalent to Lemma 4.1 of Bickel (1975). The proof is immediate from Lemma 1 of Weiss (1986b) and is omitted. But note that it follows directly from a linearization of $T^{-\frac{1}{2}} \Sigma x_t \psi(\epsilon_t(\beta))$ which forms an integral part of Huber's approach.

The main difficulty with applying Theorem 1 for inference is the estimation of D_T. If η_t is homoscedastic then $D_T = 2h(0) A_T$. Following Powell (1984), $h(0)$ may be estimated by

$$\hat{h}(0) = (2\hat{k}_T T)^{-1} \Sigma\, 1(-\hat{k}_T < \epsilon_t(\hat{\beta}_{LA}) < \hat{k}_T) \tag{4.9}$$

where $1(\cdot)$ is the indicator function and $\hat{k}_T$ is a suitable constant converging to zero. For example, let $\hat{k}_T = c_0 T^{-\gamma} \hat{\sigma}^2$ where $c_0 > 0$, $\gamma = .25$ and $\hat{\sigma}^2 = T^{-1} \Sigma \epsilon_t (\hat{\beta}_{LA})^2$ takes the scale of the data into account. In (4.9), the right-hand-side approximates the two-sided derivative of the distribution function of the errors at zero. Similarly, some other kernel estimator could be used to estimate $h(0)$.

The extension of (4.9) to the heteroscedastic model is

$$\hat{D}_T(\hat{\beta}_{LA}) = (\hat{k}_T T)^{-1} \Sigma\, 1(-\hat{k}_T < \epsilon_t(\hat{\beta}_{LA}) < \hat{k}_T) x_t x_t' \tag{4.10}$$

It can be shown that under the appropriate conditions, $\hat{h}(0) - h(0) \xrightarrow{p} 0$ or $\hat{D}_T - D_T \xrightarrow{p} 0$. Details may be found in Powell (1984) or Weiss (1986b).

The estimation of $h(0)$ or D_T allows the estimation of the covariance matrix of $\hat{\beta}_{LA}$, *i.e.*, $[2h(0)]^{-2}A_T$ or $D_T^{-1}A_T D_T^{-1}$ with and without homoscedasticity respectively, and the derivation of the Wald and LM tests based on $\hat{\beta}_{LA}$. These have also been analysed in Weiss (1986b) (and originally in Koenker and Bassett (1982)). Hence we discuss these only briefly here, highlighting the comparison of OLS and LAE-based tests.

For simplicity, we consider the linear hypotheses

$$H_0 : R\beta_0 = r \tag{4.11}$$

where R is $q \times k$ with rank $q < k$, and r is $q \times 1$. The Wald and LM tests based on $\hat{\beta}_{LA}$ are given in the next theorem. Denote the Wald test statistics by $\xi_W(i)$, $i = LS, LA$ for OLS and LAE respectively and the LM test statistics by $\xi_{LM}(i)$, $i = LS, LA$. Also let $\tilde{\beta}_i$, $i = LS, LA$ be the OLS (LAE) estimator obtained by minimizing $Q_{LS}(Q_{LA})$ subject to the restrictions $R\beta = r$, and define $\tilde{D}_T$ as the estimate of D_T evaluated at $\tilde{\beta}_{LA}$.

THEOREM 3: *(LAE-based Wald and LM Tests)*

a) Wald Test

$$\xi_W(LA) = T(R\hat{\beta}_{LA} - r)'(R\hat{D}_T^{-1}\hat{A}_T\hat{D}_T^{-1}R')^{-1}(R\hat{\beta}_{LA} - r) \tag{4.12}$$

and $\xi_W(LA)$ is distributed as a noncentral χ_q^2 with noncentrality parameter

$$\mu_{LA} = \gamma' K_2' D^{-1} R'(RD^{-1}AD^{-1}R')^{-1}RD^{-1}K_2\gamma \tag{4.13}$$

b) LM Test

$$\xi_{LM}(LA) = T^{-1}\bar{\Psi}(\tilde{\beta}_{LA})'X\tilde{D}_T^{-1}R'(R\tilde{D}_T^{-1}\hat{A}_T\tilde{D}_T^{-1}R')^{-1}R\tilde{D}_T^{-1}X'\bar{\Psi}(\tilde{\beta}_{LA}) \tag{4.14}$$

and

$$\xi_{LM}(LA) - \xi_W(LA) = o_p(1).$$

Hence $\xi_{LM}(LA)$ is also distributed as a noncentral χ^2_q with concentrality parameter μ_{LA}. For the LAE-based LR test, Koenker and Bassett suggested using

$$\xi_{LR}(LA) = [Q_{LA}(\tilde{\beta}_{LA}) - Q_{LA}(\hat{\beta}_{LA})]/h_T(0) \tag{4.15}$$

However as is usual, $\xi_{LR}(LA)$ is not χ^2_q unless the errors are homoscedastic.

With homoscedasticity, $\xi_W(LA)$ and $\xi_{LM}(LA)$ also simplify. For example, in the simple exclusion hypothesis, $\beta_2 = 0$, where $\beta' = (\beta_1' : \beta_2')$, $\xi_{LM}(LA)$ reduces to

$$\bar{\Psi}(\tilde{\beta}_{LA})'X(X'X)^{-1}X'\bar{\Psi}(\tilde{\beta}_{LA}) + o_p(1) \tag{4.16}$$

which is estimated by TR^2 from the regression of $\bar{\Psi}(\tilde{\beta}_{LA})$ on X. Similarly, $\xi_W(LA)$ becomes $4h_T(0)^2T(R\hat{\beta}_{LA} - r)'(R\hat{A}_T^{-1}R')^{-1}(R\hat{\beta}_{LA} - r)$. The LM test is particularly convenient because the estimation of $h(0)$ is not required. Also, as Engle (1982a) notes, a variety of testing problems can be viewed in terms of omitted variables.

The Wald and LM tests based on $\hat{\beta}_{LS}$ follow directly from Gallant and White (1986). Alternatively, in (4.12) and (4.14), we simply replace $\hat{\beta}_{LA}$ by $\hat{\beta}_{LS}$, D_T by A_T, A_T by B_T, $\bar{\Psi}(\tilde{\beta}_{LA})$ by $\epsilon(\tilde{\beta}_{LS})$ and K_{2T} by K_{1T} to obtain

$$\xi_W(LS) = T(R\hat{\beta}_{LS} - r)'(R\hat{A}_T^{-1}\hat{B}_T\hat{A}_T^{-1}R')^{-1}(R\hat{\beta}_{LS} - r) \tag{4.17}$$

$$\xi_{LM}(LS) = T^{-1}\epsilon(\tilde{\beta}_{LS})'X\hat{A}_T^{-1}R'(R\hat{A}_T^{-1}\tilde{B}_T\hat{A}_T^{-1}R')^{-1}R\hat{A}_T^{-1}X'\epsilon(\tilde{\beta}_{LS}) \tag{4.18}$$

and $\xi_W(LS) - \xi_{LM}(LS) = o_p(1)$, where $\tilde{B}_T$ is B_T evaluated at $\tilde{\beta}_{LS}$. Further, both are within $o_p(1)$ distributed as noncentral χ^2_q with noncentrality parameter

$$\mu_{LS} = \gamma'K_1'A^{-1}R'(RA^{-1}BA^{-1}R')^{-1}RA^{-1}K_1\gamma \tag{4.19}$$

In the exclusion hypothesis $\beta_2 = 0$ and with homoscedasticity, the LM test statistic is obtained as TR^2 from the regression of the residuals on X, rather than the signs of the residuals on X.

The OLS and LAE-based tests may be compared through their asymptotic relative efficiency (ARE). The ARE of $\xi_W(LA)$ to $\xi_W(LS)$ 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 alternatives" (Koenker and Bassett (1982)). Unfortunately, it is difficult to make this comparison in general, as inspection of (4.13) and (4.19) demonstrates. However it is easy to find examples in which either OLS or LAE-based tests are preferable.

Similarly, it is obviously important to test for heteroscedasticity. Tests along the lines of those in Bruesch and Pagan (1979), Engle (1982b) or Glejser (1969), and based on either $\epsilon_t(\hat{\beta}_{LS})^2$ or $|\epsilon_t(\hat{\beta}_{LA})|$, may be derived. In this case, an additional complication is that the models for heteroscedasticity specified under the alternative may be different. Rather than pursuing this (some comments may be found in Weiss (1986b)), we now turn to tests based on comparing $\hat{\beta}_{LS}$ and $\hat{\beta}_{LA}$, rather than comparing tests based on each.

5. COMPARING OLS AND LAE

In this section we derive the Hausman test based on the difference between $\hat{\beta}_{LA}$ and $\hat{\beta}_{LS}$. Following Newey (1985), this test is equivalent to GMM tests based on the moment conditions from either OLS or LAE. We concentrate on the latter, leading to tests involving the signs of the residuals. Also, since GMM tests are discussed extensively in Newey (1985), many features of the tests are left implicit.

The basis of the results is the asymptotic linearization (4.8). In particular, since $T^{\frac{1}{2}}(\hat{\beta}_{LA}-\beta_0)$ is $O_p(1)$,

$$T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LA}) = T^{-\frac{1}{2}}X'\bar{\Psi}(\beta_0) - D_T T^{\frac{1}{2}}(\hat{\beta}_{LA}-\beta_0) + o_p(1). \tag{5.1}$$

Similarly,

$$T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LS}) = T^{-\frac{1}{2}}X'\bar{\Psi}(\beta_0) - D_T T^{\frac{1}{2}}(\hat{\beta}_{LS}-\beta_0) + o_p(1) \tag{5.2}$$

Therefore, since $T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LA}) = o_p(1)$,

$$T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LS}) = D_T T^{\frac{1}{2}}(\hat{\beta}_{LA}-\hat{\beta}_{LS}) + o_p(1) \tag{5.3}$$

That is, a Hausman test based on $(\hat{\beta}_{LA} - \hat{\beta}_{LS})$ is asymptotically equivalent to a GMM test based on the moment conditions $T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LS})$. The GMM test has the advantage of being easier to calculate since the LAE estimator itself is not needed. Note also that we can define the two-step estimator

$$\dot{\beta} = \hat{\beta}_{LS} + \dot{D}_T^{-1}T^{-1}X'\bar{\Psi}(\hat{\beta}_{LS}) \tag{5.4}$$

where $\dot{D}_T = \hat{D}_T(\hat{\beta}_{LS})$ is a consistent estimate of D_T. From (5.3), $\dot{\beta}$ is asymptotically equivalent to the LAE estimator although unfortunately the estimation of D_T is still required.

Next, from (5.2) and (3.5),

$$\begin{aligned} T^{-\frac{1}{2}}X'\bar{\Psi}(\hat{\beta}_{LS}) &= T^{-\frac{1}{2}}X'\bar{\Psi}(\beta_0) - D_T A_T^{-1} T^{-\frac{1}{2}} X'\epsilon(\beta_0) + o_p(1) \\ &= [I : -D_T A_T^{-1}]\left[T^{-\frac{1}{2}}\left(\Psi(\beta_0)'X : \epsilon(\beta_0)'X\right)'\right] + o_p(1) \end{aligned} \tag{5.5}$$

By the CLT

$$V_T^{-\frac{1}{2}}T^{-\frac{1}{2}}\left(\Psi(\beta_0)'X : \epsilon(\beta_0)'X\right)' - V_T^{-\frac{1}{2}}(K_{2T}' : K_{1T}')'\gamma \xrightarrow{d} N(0, I_{2k}) \tag{5.6}$$

where

$$V_T = \begin{pmatrix} A_T & V_{12T} \\ V_{12T} & B_T \end{pmatrix} \tag{5.7}$$

and

$$V_{12T} = T^{-1}\Sigma E\left(|\eta_t| x_t x_t'\right) \tag{5.8}$$

V, the limit of V_T, is assumed to be positive definite. The test statistics for the GMM and Hausman tests, ξ_M and ξ_H respectively, follow directly from (5.3), (5.5) and (5.6) and are given in the next theorem.

THEOREM 4: *(GMM and Hausman tests):*

$$\xi_M = T^{-1}\bar{\Psi}(\hat{\beta}_{LS})'XW_T^{-1}X'\bar{\Psi}(\hat{\beta}_{LS}) \tag{5.9}$$

and ξ_M is asymptotically distributed as a noncentral χ_k^2 with noncentrality parameter

$$\mu_M = \gamma'(K_2' : K_1')(I : -DA^{-1})'W^{-1}(I : -DA^{-1})(K_2' : K_1')'\gamma \tag{5.10}$$

where $W = \lim_{T\to\infty} W_T$,

$$W_T = (I : -D_T A_T^{-1}) V_T (I : -D_T A_T^{-1})' \tag{5.11}$$

and W *rank* k. *Further,*

$$\xi_H = T(\hat{\beta}_{LA} - \hat{\beta}_{LS})' D_T W_T^{-1} D_T (\hat{\beta}_{LA} - \hat{\beta}_{LS}) \tag{5.12}$$

and

$$\xi_H - \xi_M = o_p(1) \tag{5.13}$$

Clearly the asymptotic distributions of ξ_H and ξ_M are unchanged if W_T and D_T are replaced by consistent estimates. $\hat{D}_T$ may be based on $\hat{\beta}_{LS}$ rather than $\hat{\beta}_{LA}$, and the estimate of V_{12} is

$$\hat{V}_{12} = T^{-1}\Sigma|\epsilon_t(\hat{\beta}_{LS})|x_t x_t' \tag{5.14}$$

An additional asymptotically equivalent test is obtained using the first-order conditions for OLS. That is,

$$T^{-\frac{1}{2}} X'\epsilon(\hat{\beta}_{LS}) = T^{-\frac{1}{2}} X'\epsilon(\hat{\beta}_{LA}) - A_T T^{\frac{1}{2}} (\hat{\beta}_{LS} - \hat{\beta}_{LA}) + o_p(1), \tag{5.15}$$

or, using (5.3),

$$\begin{aligned} T^{-\frac{1}{2}} X'\epsilon(\hat{\beta}_{LA}) &= A_T T^{\frac{1}{2}} (\hat{\beta}_{LA} - \hat{\beta}_{LS}) + o_p(1) \\ &= A_T D_T^{-1} T^{\frac{1}{2}} X'\bar{\Psi}(\hat{\beta}_{LS}) + o_p(1) \end{aligned} \tag{5.16}$$

Hence a test based on $T^{-\frac{1}{2}}X'\epsilon(\hat{\beta}_{LA})$, measuring by how much $\hat{\beta}_{LA}$ violates the OLS orthogonality conditions, is asymptotically equivalent to one testing whether $\hat{\beta}_{LS}$ violates the LAE orthogonality conditions.

As noted above, many other features of GMM tests are derived in Newey (1985). For example, in any GMM test there will be misspecification directions in which the noncentrality parameter is zero, and the power of the test is equal to its size. In the current model, this may occur if ℓ, the dimension of c_T and γ, exceeds the rank of $D^{-1}K_2 - A^{-1}K_1$.

Then γ may lie in the nullspace of $D^{-1}K_2 - A^{-1}K_1$ and from (5.10), the test will have noncentrality parameter zero. That is, if rank $(D^{-1}K_2 - A^{-1}K_1) \leq k$, then the dimension of the nullspace is greater than or equal to $\ell - k$. Hence, if $\ell - k > 0$, or $2k < \ell + k$, then the test may have noncentrality parameter zero. The last inequality allows a degrees of freedom interpretation of the problem. The number of moment conditions (*i.e.*, degrees of freedom available for estimating parameters) is $2k$, and this is less than the total number of unknown parameters in β and c, *i.e.*, $k + \ell$. Since the application of other results in Newey (1985) is straightforward, rather than pursuing these, we now turn to some special cases.

6. Special Cases

As noted in the Introduction, the exact forms of the GMM and Hausman tests depend upon the presence or absence of heteroscedasticity. We therefore consider the special case of no heteroscedasticity, leading to the test suggested by Brown and Kildea (1979). Following this, we discuss an explicitly asymmetric local alternative, which highlights some of the properties of the test.

i) Homoscedasticity.

In general, the test statistic ξ_M requires the estimation of W_T and hence D_T. However, in certain examples, this is not the case. In the absence of heteroscedasticity, $D_T = 2h(0)A_T$, $B_T = \sigma^2 A_T$ and $V_{12} = \omega A_T$, where $\sigma^2 = E(\eta_t^2)$ and $\omega = E(|\eta_t|)$. W_T becomes

$$W_T = qA_T \tag{6.1}$$

where $q = (1 - 4h(0)\omega + 4h(0)^2\sigma^2)$. Hence

$$\xi_M = q^{-1}\bar{\Psi}(\hat{\beta}_{LS})'X(X'X)^{-1}X'\bar{\Psi}(\hat{\beta}_{LS}) + o_p(1) \tag{6.2}$$

which is the test statistic derived by Brown and Kildea (although they also assumed fixed X's). The estimation of $h(0)$ is still necessary, although if the errors are normal, then $q = (1 - 2/\pi)$ while in the double exponential, $q = 1$.

Further, except for q, ξ_M is obtained as TR^2 from the regression of $\bar{\Psi}(\hat{\beta}_{LS})$, the signs of the residuals from OLS, on X. This follows because $\bar{\Psi}(\hat{\beta}_{LS})'\bar{\Psi}(\hat{\beta}_{LS}) = T$. Similarly, for the test based on $T^{-\frac{1}{2}}X'\epsilon(\hat{\beta}_{LA})$, in the absence of heteroscedasticity we simply regress $\epsilon(\hat{\beta}_{LA})$ on X and keep the explained sum of squares. This is then multiplied by $4\hat{h}_T(0)^2/\hat{q}_T$, where $\hat{q}_T$ is the obvious estimate of q, to give the test statistic.

Brown and Kildea suggested ξ_M as a test for symmetry in the absence of heteroscedasticity rather than a test for misspecification as is implicit here. To see why this is the case, we next consider the noncentrality parameter in the absence of heteroscedasticity. Our reasons for considering an explicitly asymmetric local alternative are then obvious.

If the misspecified alternative implies that the conditional distribution of η_t has non-zero mean and median, but is still symmetric (as is suggested by Assumption 1) then the effects on OLS and LAE should be similar. In particular, with homoscedasticity, the noncentrality parameter may be zero implying the test has power equal to its size. For example, suppose that

$$\eta_t = \varsigma_t + z_t' c_T \tag{6.3}$$

where ς_t is symmetric about zero, conditional on I_{t-1}, z_t is a vector of omitted variables measurable with respect to I_{t-1}, and $c_0 = 0$. Suppose further than the distributions of ς_t, x_t and z_t do not depend upon c_T. Then

$$E_T(x_t\eta_t) = E(x_t z_t' c_T) \tag{6.4}$$

and at $c = c_0$,

$$\partial E_T(x_t\eta_t)/\partial c = E(x_t z_t'). \tag{6.5}$$

Alternatively,

$$E_T[x_t\psi(\eta_t)] = E[x_t - 2x_t H_t(-z_t' c_T)] \tag{6.6}$$

where H_t is the conditional distribution of ς_t. Then

$$\begin{aligned} \partial E_T[x_t\psi(\eta_t)]/\partial c &= 2E[h_t(-z_t' c_T) x_t z_t'] \\ &\longrightarrow 2E[h_t(0) x_t z_t'] \end{aligned} \tag{6.7}$$

Therefore, under homoscedasticity,

$$
\begin{aligned}
D_T^{-1} K_{2T}\gamma &= [2h(0)]^{-1} A_T^{-1} 2h(0) T^{-1} \Sigma E(x_t z_t')\gamma \\
&= A_T^{-1} K_{1T}\gamma
\end{aligned}
\tag{6.8}
$$

where $h(0)$ is the height of the common density of ς_t at zero, and the noncentrality parameter is zero. Another example is the simultaneous equation system considered by Hausman (1978):

$$
\begin{aligned}
y_{1t} &= x_{1t}\gamma + u_{1t} \\
y_{2t} &= y_{1t}\beta + u_{2t}
\end{aligned}
\tag{6.9}
$$

Suppose that (u_{1t}, u_{2t}) are zero-mean bivariate normal. Then

$$
u_{2t} \mid u_{1t} \sim N\left(\frac{\sigma_{12}}{\sigma_1^2} u_{1t}, \left(1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right)\sigma_2^2\right)
$$

where $\sigma_{12} = \mathrm{cov}(u_{1t}, u_{2t})$ and $\sigma_i^2 = \mathrm{var}(u_{it})$, $i = 1, 2$. The conditional density is symmetric, and with homoscedasticity, the noncentrality parameter is again zero.

Of course, with heteroscedasticity, the argument in (6.8) for example is not appropriate and the noncentrality parameter may be non-zero, giving the test power. These features however suggest that for many alternatives, the test may not be powerful and conform to the notion that a test comparing mean-based and median-based estimators may be more useful against asymmetry. Whether or not this is in fact the case is a subject for further research, *e.g.*, a Monte Carlo experiment. Here, we now turn to an alternative of asymmetry.

ii) Asymmetry.

Suppose that the conditional distribution of η_t is symmetric about zero but instead of (2.2) we have

$$
y_t = x_t'\beta_0 + \eta_t^* \tag{6.10}
$$

where η_t^* is explicitly asymmetric:

$$\eta_t^* = \eta_t(1 + 1(\eta_t > 0)\gamma_1/\sqrt{T}) \tag{6.11}$$

where γ_1 is a positive scalar (*e.g.*, Newey and Powell (1985), Antille, Kersting and Zucchini (1982), and Boos (1982)). Then since $\text{sign}(\eta_t^*) = \text{sign}(\eta_t)$, it follows that $\text{median}(\eta_t^*) = \text{median}(\eta_t) = 0$, and

$$E(\psi(\eta_t^*) \mid I_{t-1}) = 0 \tag{6.12}$$

where the expectation is taken over the conditional density of η_t, which does not depend upon T. Also,

$$E(\eta_t^* \mid I_{t-1}) = \gamma_1\omega_t/(2\sqrt{T}) \tag{6.13}$$

where $\omega_t = E(|\eta_t| \mid I_{t-1})$. Therefore $K_{2T} = 0$ while

$$\begin{aligned} K_{1T} &= (2T)^{-1}\Sigma\partial E(x_t\omega_t\gamma_1/\sqrt{T})/\partial(\gamma_1/\sqrt{T}) \\ &= (2T)^{-1}\Sigma E(x_t\omega_t). \end{aligned} \tag{6.14}$$

With homoscedasticity, $\omega_t = \omega$ for all t, and the mean term in the asymptotic distribution of $\hat{\beta}_{LS}$, *i.e.*, $A_T^{-1}K_{1T}\gamma$, becomes $\omega\gamma_1/2[T^{-1}\Sigma E(x_t x_t')]^{-1}T^{-1}\Sigma E(x_t) = \omega\gamma_1 R/2$, where $R = (1\ 0\ldots0)'$. Hence only the constant term is affected. Otherwise, the asymmetry will affect all the parameters.

Equation (6.13) implies that the asymmetry is equivalent to having omitted the variable ω_t from the regression equation. That is, define the error

$$\eta_t^{\#} = \eta_t^* - \gamma_1\omega_t/(2\sqrt{T}) \tag{6.15}$$

Then (6.10) can be written

$$y_t = x_t'\beta_0 + \gamma_1\omega_t/(2\sqrt{T}) + \eta_t^{\#} \tag{6.16}$$

where $E(\eta^{\#} \mid I_{t-1}) = 0$. Regressing y_t on x_t implies that ω_t is an omitted variable. If, as in many models for heteroscedasticity, ω_t (or equivalently $\sigma_t^2 = E(\eta_t^2|I_{t-1})$) depends upon regressors in x_t, then this effect corresponds to ordinary omitted variables. Alternatively, in an ARCH-type model, we might have

$$\omega_t = \alpha_0 + \alpha_1 |\eta_{t-1}| \tag{6.17}$$

or (with normality),

$$\begin{aligned}\omega_t &= (2/\pi)^{\frac{1}{2}} \sigma_t \\ &= (2/\pi)^{\frac{1}{2}} (\alpha_0 + \alpha_1 \eta_{t-1}^2)^{\frac{1}{2}} \end{aligned} \tag{6.18}$$

where $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$. In the former, the omitted variable is the lagged error, while in the latter, the asymmetric alternative is equivalent to having omitted the conditional standard deviation of the error from the equation explaining the mean. Models such as this arise in the analysis of risk, *e.g.*, Pagan and Ullah (1986).

Since only the constant term is affected when the errors are homoscedastic, an alternative to ξ_M is the the Hausman test is based upon $R'(\hat{\beta}_{LA} - \hat{\beta}_{LS})$. But from (5.3),

$$T^{\frac{1}{2}} R'(\hat{\beta}_{LA} - \hat{\beta}_{LS}) = R' D_T^{-1} T^{-\frac{1}{2}} X' \bar{\Psi}(\hat{\beta}_{LS}) + o_p(1) \tag{6.19}$$

Then as in Theorem 4,

$$\xi_C \equiv T(\hat{\beta}_{LA} - \hat{\beta}_{LS})' R (R' D_T^{-1} W_T D_T^{-1} R)^{-1} R' (\hat{\beta}_{LA} - \hat{\beta}_{LS}) \tag{6.20}$$

$$= q^{-1} \bar{\Psi}(\hat{\beta}_{LS})' X (X'X)^{-1} R (R'(X'X)^{-1} R)^{-1} R' (X'X)^{-1} X' \bar{\Psi}(\hat{\beta}_{LS}) + o_p(1) \tag{6.21}$$

ξ_C is within $o_p(1)$ distributed as noncentral χ_1^2 with noncentrality parameter

$$\begin{aligned}\mu_C &= k_1 (R' A^{-1} R)^{-1} \\ &= k_1 \lim_{T \to \infty} \left[1 - T^{-1} \Sigma E_T(\tilde{x}_t') [\Sigma E_T(\tilde{x}_t \tilde{x}_t')]^{-1} \Sigma E_T(\tilde{x}_t)\right] \end{aligned} \tag{6.22}$$

where $k_1 = q^{-1} h(0)^2 \omega^2 \gamma_1^2$ and $x_t' = (1 : \tilde{x}_t')$. From (6.21), the test statistic is estimated by regressing $\bar{\Psi}(\hat{\beta}_{LS})$ on X and multiplying the square of the estimated t-statistic on the first coefficient by $q^{-1} s^2$, where $s^2 = 1 - q\xi_M / T$ is the mean squared error from the regression.

The alternative considered by Brown and Kildea was asymmetric contamination leading to bias in the distribution of $T^{\frac{1}{2}}(\hat{\beta}_{LA}-\beta_0)$ rather than in the distribution of $T^{\frac{1}{2}}(\hat{\beta}_{LS}-\beta_0)$. However, with homoscedasticity, this is equivalent to the case just considered. That is, let

$$y_t = x_t'\beta_0 + \eta_t^{\#} \tag{6.23}$$

Then since $E(\eta_t^{\#} \mid I_{t-1}) = 0$ while $E\big(\psi(\eta_t^{\#}) \mid I_{t-1}\big) \neq 0$, $K_{1T} = 0$ while $K_{2T} = \omega h(0)T^{-1}\Sigma E(x_t)$. Clearly this transformation does not affect the test statistic. But now, since $E\big(\psi(\eta_t^{\#}) \mid I_{t-1}\big) \neq 0$, the contamination distorts the balance of residual signs from OLS. This motivates the use of $\bar{\Psi}(\hat{\beta}_{LS})$ in a test statistic. Of course, the same intuition applies in the LAE-based LM test for omitted variables, (4.16). There, the balance of signs is upset by the omission of the variables corresponding to β_2.

Finally, we compare ξ_M and ξ_C as tests for asymmetry. Under the local alternative (6.11), ξ_M has a χ^2 distribution with k degrees of freedom and noncentrality parameter $\mu_M = k_1$. On the other hand, the distribution of ξ_C has 1 degree of freedom and noncentrality parameter $\mu_C \leq k_1$. From (6.22), the relationship between ξ_M and ξ_C depends upon the x_t, and whether the effect of the decrease in the number of degrees of freedom from ξ_M to ξ_C is greater than the effect of the decrease in noncentrality parameter. If $E(x_t)$ is small, then ξ_C is the more powerful test. This dependence on $E(x_t)$ is equivalent to the behavior of a t-test on the constant term in a simple regression model with $k = 2$ and x_t fixed (*e.g.*, Johnston (1984, equation 2-46)). In the current model, a linear transformation of the $\tilde{x}_t$ can be made to set $\hat{\beta}_{1LS} = 0$, where $\beta_{1LS} = R'\beta_{LS}$. Then ξ_C becomes

$$4h(0)^2q^{-1}T\hat{\beta}_{1LA}^{*2}[1 + T\bar{x}^2/\Sigma(x_t^* - \bar{x})^2]^{-1} \tag{6.24}$$

where the x_t^* are the transformed $\tilde{x}_t$, $\bar{x} = T^{-1}\Sigma x_t^*$ and $\hat{\beta}_{1LA}^*$ is the LAE estimate of $\beta_1 = R'\beta_0$ in the transformed model. But $4h(0)^2q^{-1}T\hat{\beta}_{1LA}^2$ is also the first of the four terms obtained when the quadratic form in the Hausman test (5.12) is expanded. Since only the constant term is affected by the asymmetry, the other three terms do not contribute to the noncentrality parameter. The denominator in (6.24) represents the differences in the noncentrality parameters since $[1 + T\bar{x}^2/\Sigma(x_t^* - \bar{x})^2]^{-1} = 1 - T\bar{x}^2/\Sigma x_t^{*2}$. Presumably of course, in most cases the test with the smaller number of degrees of freedom would be preferred.

Similarly, with heteroscedasticity and for the more general local alternative (2.2), Hausman tests based on subsets of the elements of $(\hat{\beta}_{LA} - \hat{\beta}_{LS})$, equivalent to (6.20) but with different selection matrices R, can be derived. Again, the relative powers of the tests would depend upon the actual process.

7. Concluding Comments

In this paper, we have discussed some of the relationships between OLS and LAE estimation. These can be summarized as follows. Firstly, the estimation and testing results for OLS and LAE are essentially equivalent and in each case they are derived from a linearization of the first-order conditions. In general however, comparing tests for the same hypotheses based on each is difficult.

Next, the comparison of OLS and LAE leads to a GMM test based on the signs of the residuals from OLS. The first-order conditions for LAE imply that the residuals round the median will be "balanced," *e.g.*, the sum of their signs converges to zero since a constant is included in x_t. This is not necessarily so for OLS although it should be approximately true if the errors are symmetric. Hence alternatives which disturb the relationship between the mean and median will affect the balance of the signs of the OLS residuals. Misspecification leading to asymmetry in the errors is an alternative which will lead to this. On the other hand the test is unlikely to be powerful against alternatives which imply that the mean and median of the errors are equal. In fact in some situations, *e.g.*, with homoscedasticity, the test has power equal to its size. Newey and Powell (1985) report the results of a Monte Carlo experiment which confirms that the Hausman test is useful against asymmetric alternatives, at least in a simple model.

We have also demonstrated that asymmetry can affect more than just the constant in the regression equation. The effect is equivalent to having omitted a variable. Because of this, it is important to test for asymmetry, particularly in the presence of heteroscedasticity.

Alternatively, it is possible to extend the model. For example, we might suppose that the errors are autocorrelated or heteroscedastic, as here, but the form of the error process is known. Then some type of GLS estimator would be appropriate. Similarly, with heteroscedasticity of unknown form, an estimator along the lines of that suggested by Cragg (1983) would be useful. In the simultaneous equation model, Amemiya (1982) has suggested a two stage LAE estimator. Presumably, this can be extended to the current dynamic framework.

Finally, in these and other cases, it should be remembered that OLS estimators are optimal with normal homoscedastic errors. As the distribution becomes long-tailed, the extra effort in an estimation method such as LAE may be warranted by the increase in the precision of the estimates. Poirier, Tello and Zin (1986) have proposed an LM test which might be used to give an indication of the preferable estimation method.

Mathematical Appendix

Proof of Theorem 1. The proof follows that of Theorem 2 of Weiss (1986b) (which in turn is modeled on those of Powell (1984, Theorem 2) and Huber (1967, Lemma 3 and Theorem 3)). There are two main modifications.

i) In Weiss (1986b), $q_t(\beta) \equiv x_t\psi(\epsilon_t(\beta))$, and

$$\begin{aligned}\lambda_T(\beta) &= T^{-1}\Sigma E_T\big[x_t\psi(\epsilon_t(\beta))\big] \\ &= 2T^{-1}\Sigma E_T\big[x_t\{\frac{1}{2} - H_{t,T}(x_t'(\beta-\beta_0))\}\big]\end{aligned}$$

But since $H_{t,T}$ and $u_{t,T}$ are continuous in c,

$$\begin{aligned}&E_T\big[x_t\{\frac{1}{2} - H_{t,T}(x_t'(\beta-\beta_0))\}\big] \\ &\longrightarrow E\big[x_t\{\frac{1}{2} - H_t(x_t'(\beta-\beta_0))\}\big]\end{aligned}$$

as $T \to \infty$ by the Lebesgue Dominated Convergence Theorem. Then applying a mean value expansion to H_t gives

$$H_t(x_t'(\beta-\beta_0)) = H_t(0) + h_t(x_t'(\beta^*-\beta_0))x_t'(\beta-\beta_0)$$

where β^* lies between β and β_0. Hence

$$\begin{aligned}\lambda_T(\beta) &= -2T^{-1}\Sigma E\big[h_t(x_t'(\beta^*-\beta_0))x_tx_t'\big](\beta-\beta_0) + o(1) \\ &= -D_T(\beta-\beta_0) + O(\|\beta-\beta_0\|^2) + o(1)\end{aligned}$$

since h_t is Lipschitz continuous. For T large enough, $|\lambda_T(\beta)| \geq a\|\beta-\beta_0\|$ for some $a > 0$.

ii) The results in Weiss (1986b, Theorem 2) therefore imply that

$$T^{-\frac{1}{2}}\Sigma x_t\psi(\epsilon_t(\beta_0)) - D_TT^{\frac{1}{2}}(\hat{\beta}-\beta_0) \xrightarrow{p} 0$$

The CLT is then applied to

$$T^{-\frac{1}{2}}\Sigma x_t\psi(\epsilon_t(\beta_0)) = T^{-\frac{1}{2}}\Sigma\{x_t\psi(\epsilon_t(\beta_0)) - E_T[x_t\psi(\epsilon_t(\beta_0))]\} + T^{-\frac{1}{2}}\Sigma E_T[x_t\psi(\epsilon_t(\beta_0))]$$

Following Newey (1985, Lemma A.8), since the expectation is continuous in c and at c_0, $E[x_t\psi(\epsilon_t(\beta_0))] = E\big[x_t E\big(\psi(\epsilon_t(\beta_0)) \mid I_{t-1}\big)\big] = 0$,

$$A_T^{-1}T^{-\frac{1}{2}}\Sigma\{x_t\psi(\epsilon_t(\beta_0)) - E_T[x_t\psi(\epsilon_t(\beta_0))]\} \xrightarrow{d} N(0, I_k)$$

Also following Newey (1985, Lemma 1) and applying a mean value expansion around $c = c_0$,

$$\begin{aligned} E_T\big[x_t\psi(\epsilon_t(\beta_0))\big] &= 2E_T[x_t\{\frac{1}{2} - H_{t,T}(0)\}] \\ &\longrightarrow -2E[x_t\partial H_t/\partial c']\gamma/\sqrt{T} \end{aligned}$$

where $\partial H_t/\partial c$ is the derivative of $H_{t,T}(0)$ with respect to c, evaluated at $c = c_0$. Hence $T^{-\frac{1}{2}}\Sigma x_t\psi(\epsilon_t(\beta_0))$ is asymptotically normal with mean $K_2\gamma/\sqrt{T}$.

Proof of Theorem 3.

a) **Wald Test.** Standard. See, for example, Theorem 3.5 of White (1980).

b) **LM Test.** The argument for $\hat{\beta}_{LA}$ can be applied to $\tilde{\beta}_{LA}$ to show that $T^{\frac{1}{2}}(\tilde{\beta}_{LA} - \beta_0)$ is $O_p(1)$. Then (4.8) implies that

$$T^{-\frac{1}{2}}\Sigma x_t\psi\big(\epsilon_t(\tilde{\beta}_{LA})\big) - T^{-\frac{1}{2}}\Sigma x_t\psi\big(\epsilon_t(\beta_0)\big) + D_T T^{\frac{1}{2}}(\tilde{\beta}_{LA} - \beta_0) = o_p(1),$$

or that

$$R'D_T^{-1}T^{-\frac{1}{2}}\Sigma x_t\psi\big(\epsilon_t(\tilde{\beta}_{LA})\big) - R'D_T^{-1}T^{-\frac{1}{2}}\Sigma x_t\psi\big(\epsilon_t(\beta_0)\big) = o_p(1)$$

since $T^{\frac{1}{2}}(R'\tilde{\beta}_{LA} - r) = 0$. The result then follows from the distribution of $T^{-\frac{1}{2}}\Sigma x_t\psi(\epsilon_t(\beta_0))$.

Proof of Theorem 4. The distribution of ξ_M follows from (5.5) and (5.6). The expression for ξ_H and the asymptotic equivalence of ξ_M and ξ_H follows from (5.3). W has rank k because V is positive definite and $(I : -DA^{-1})$ has rank k (see Johnston (1984, p. 153)).

FOOTNOTES

1. Note that while many of the moment conditions pertain to the regressors, the (possibly) dynamic nature of the model may imply similar conditions on the errors. In contrast, in the static linear model, it is not necessary to assume that any moments of the errors are finite. However the usual moments of the regressors must still exist, and we note that this represents a somewhat artificial situation for economic data.

2. See, for example, White and Domowitz (1984) for a definition of mixing. Again, the actual definition is not crucial to our development.

3. It would also be repetitious to state the assumptions when they are standard. Other references to this type of material include White (1984) and White and Domowitz (1984).

4. The particular LLN and CLT we have in mind are those used in Weiss (1986b). These are essentially restatements of those proved in Newey (1985, Lemmas A.7 and A.8). The important feature of these results is that they allow for the drift in the data generation process (2.2) which is a consequence of c depending on T.

5. Of course, the extension to allowing variables other than x_t and lags of η_t to affect the variance is straightforward. But since we are not dealing with the model for the heteroscedasticity explicitly, we retain the simpler framework.

References

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

Antille, A., G. Kersting and W. Zucchini. Testing Asymmetry. *Journal of the American Statistical Association* 77 (1982), 639-646,

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

Boos, D.D. A Test for Asymmetry Associated with the Hodges-Lehmann Estimator. *Journal of the American Statistical Association* 77 (1982), 647-651.

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

Brown, B.M. and D.G. Kildea. Outlier-Detection Tests and Robust Estimators Based on Signs of Residuals. *Communications in Statistics* A8 (1979): 257-269.

Cragg, J.C. More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form. *Econometrica* 51 (1983): 751-763.

Domowitz, I. New Directions in Non-linear Estimation with Dependent Observations. *Canadian Journal of Economics* 18 (1985): 1-27.

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

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

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.

Hausman, J.A. Specification Tests in Econometrics. *Econometrica* 46 (1978): 1251-1272.

Huber, P.J. The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In L.M. LeCam 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.

Johnston, J. *Econometric Methods.* Third Edition. New York: McGraw Hill, 1984.

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

Newey, W.K. Generalized Method of Moments Specification Testing. *Journal of Econometrics* 29 (1985): 229-256.

Newey, W.K. and J.L. Powell. Asymmetric Least Squares Estimation. Mimeograph, Princeton University, Department of Economics (1985).

Pagan, A.R. and A. Ullah. The Econometric Analysis of Risk Terms. Mimeograph, Australian National University (1986).

Poirier, D.J., Tello, M.D. and S.E. Zin. A Diagnostic Test for Normality within the Power Exponential Family. *Journal of Business and Economic Statistics* 4 (1986): 359-373.

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

Ruppert, D. and R.J. Carroll. Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association* 75 (1980): 828-838.

Weiss, A.A. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory* 2 (1986a): 107-131.

Weiss, A.A. Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation. MRG Working Paper #M8630, University of Southern California, Department of Economics (1986b).

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

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

REPORTS IN THIS SERIES

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.
M8632: TIMUR KURAN: The Craft Guilds of Tunis and Their Amins: A Study in Institutional Atrophy.
M8633: ROBERT KALABA and LEIGH TESFATSION: Time-Varying Regression Via Flexible Least Squares.
M8634: HERMAN C. QUIRMBACH: Comparative Statics for Oligopoly: Demand Shift Effects.
M8635: HARRISON H.C. CHENG: The Coalitional Approach to the Core Theory.
M8636: DARRELL DUFFIE and WAYNE SHAFER: Equilibrium in Incomplete Markets: I. A Basic Model of Generic Existence.
M8637: DARRELL DUFFIE and WAYNE SHAFER: Equilibrium in Incomplete Markets: II. Generic Existence in Stochastic Economies.
M8638: REZAUL KHANDKER: A Model of Layoff, Search and Job Choice and Its Estimation.
M8639: MASSOUD M. SAGHAFI and RAYMOND HUBBARD: Advertising and Cigarette Industry Sales in the United States.
M8640: JACQUES SILBER: Asymmetric Price Changes and Indexation Schemes.
M8641: JACQUES SILBER: Measures of Economic Welfare and the Skewness of Income Distributions.
M8642: DANIEL F. SPULBER: Contract Damages and Competition.
M8643: MONGI AZABOU, TIMUR KURAN and MUSTAPHA NABLI: The Wholesale Produce Market of Tunis and Its Porters: A Tale of Market Degeneration.
M8644: HANS-WALTER LORENZ: On the Uniqueness of Limit Cycles in Business Cycle Theory.
M8645: MASSOUD M. SAGHAFI: Optimal Price-Quantity Decisions for Demand-Related Commodities.
M8646: RICHARD H. DAY, KYOO-HONG KIM and DIANE MACUNOVICH: Demoeconomic Dynamics: A Classical Analysis.
M8647: MOHSEN ATTARAN and MASSOUD M. SAGHAFI: Concentration Trends in the U.S. Manufacturing Sector: 1970-1984.
M8648: ANDREW A. WEISS: A Comparison of Ordinary Least Squares and Least Absolute Error Estimation.