Aust. N. Z. J. Stat. **43**(4), 2001, 481–492

TWO-STAGE WELSH'S TRIMMED MEAN FOR THE SIMULTANEOUS EQUATIONS MODEL

LIN-AN CHEN[1*], KUO-YUAN LIANG[2] AND CHWEN-CHI LIU[3]

National Chiao Tung University, Shih Hsin University and Feng Chia University

Summary

This paper discusses the large sample theory of the two-stage Welsh's trimmed mean for the limited information simultaneous equations model. Besides having asymptotic normality, this trimmed mean, as the two-stage least squares estimator, is a generalized least squares estimator. It also acts as a robust Aitken estimator for the simultaneous equations model. Examples illustrate real data analysis and large sample inferences based on this trimmed mean.

Key words: Aitken estimator; generalized least squares estimator; simultaneous equations model; trimmed mean.

1. Introduction

The conventional method of two-stage least squares is commonly used in econometrics, with simultaneous equations models. Two justifications are frequently associated with its popularity. First, from a computational perspective, it requires only the least squares technique. Second, it is well known that a two-stage least squares estimator (2SLSE) can be interpreted as an Aitken estimator (see e.g. Fomby, Hill & Johnson, 1984 p. 478; Amemiya, 1985 p. 239). More specifically, it implies that after linear transformations of the model, the 2SLSE is a generalized least squares estimator.

It is also well known that the 2SLSE is highly sensitive to even a very small departure from normality and to the presence of outliers. Therefore, many robust estimators have been proposed as alternatives to the 2SLSE for simultaneous equations systems (see e.g. Amemiya, 1982; Powell, 1983; Krasker, 1985; Chen & Portnoy, 1996).

In this article, we extend the Welsh's trimmed mean (Welsh, 1987) for linear regression to the simultaneous equations model. Large sample statistical inferences based on this trimmed mean and real data analysis are also provided. We are interested in two aspects of this estimator. First, because its asymptotic distribution is independent of the choice of initial estimator, this trimmed mean can be obtained simply on the basis that an initial estimate can be easily computed. In contrast, the robust estimators above rely on the estimation of regression quantiles and so are computationally much more difficult. Second, we show that this trimmed mean can be interpreted as a robust Aitken estimator in the simultaneous equations model.

We introduce the two-stage Welsh's trimmed mean in Section 2 and develop its large sample distribution in Section 3. Its ability to serve as a generalized least squares estimator is

Received March 1999; revised May 2000; accepted August 2000.
* Author to whom correspondence should be addressed.
[1] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. e-mail: lachen@stat.nctu.edu.tw
[2] School of Management, Shih Hsin University, Taipei, Taiwan.
[3] Institute of Economics, Feng Chia University, Taichung, Taiwan.
Acknowledgments. The authors thank the Technical Editor and the Associate Editor and referee whose valuable comments led to a substantial improvement of the exposition.

© Australian Statistical Publishing Association Inc. 2001. Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden MA 02148, USA

proved in Section 4. Examples and large sample inferences are presented in Sections 5 and 6, respectively. Finally, the proof of Theorem 4.1 is given in the appendix.

2. Two-stage Welsh's trimmed mean

Consider the simultaneous equations model

$$\boldsymbol{y} = \boldsymbol{Y}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{\beta}_2 + \boldsymbol{\tau}, \tag{1}$$

where $\boldsymbol{Y} = [\boldsymbol{y}\ \boldsymbol{Y}_1]$ denotes an $n \times p_0$ observation matrix of p_0 endogenous variables (i.e. dependent variables), $\boldsymbol{Z}_1$ denotes an $n \times p_1$ observation matrix of p_1 exogenous variables (i.e. independent variables) including an intercept term, and $\boldsymbol{\tau}$ denotes a vector of independent and identically distributed (iid) disturbance variables. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ denote the parameter vector which is to be estimated.

Let the reduced form of the simultaneous equations model be

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Pi} + \boldsymbol{V},$$

where $\boldsymbol{Z} = [\boldsymbol{Z}_1\ \boldsymbol{Z}_2]$ denotes the set of all exogenous variables, $\boldsymbol{Z}_2$ denotes an $n \times p_2$ matrix, and rows of V are vectors of iid random variables $(v_1, \dots, v_{p_0})$ with zero mean vector and positive definite covariance matrix. Let $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_1\ \boldsymbol{\Pi}_2]$ and $\boldsymbol{V} = [\boldsymbol{V}_1\ \boldsymbol{V}_2]$ be partitioned to correspond with the dimensions of $[\boldsymbol{y}\ \boldsymbol{Y}_1]$, so that the reduced form can be represented as $[\boldsymbol{y}\ \boldsymbol{Y}_1] = \boldsymbol{Z}[\boldsymbol{\Pi}_1\ \boldsymbol{\Pi}_2] + [\boldsymbol{V}_1\ \boldsymbol{V}_2]$.

For the simultaneous equations model, the regression quantile of Koenker & Bassett (1978) can be applied, to construct the two-stage trimmed least squares estimator established by Chen & Portnoy (1996).

Let $\hat{\boldsymbol{\Pi}}_2$ be an estimator of $\boldsymbol{\Pi}_2$. Replacing $\boldsymbol{Y}_1$ by $\boldsymbol{Z}\hat{\boldsymbol{\Pi}}_2$ and using the parameter restriction $\boldsymbol{\tau} = \boldsymbol{V}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = (1, -\boldsymbol{\beta}_1)$, the simultaneous equations model can be rewritten as

$$\boldsymbol{y} = \boldsymbol{D}_n\boldsymbol{\beta} + \boldsymbol{V}_1 - \boldsymbol{Z}(\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1, \tag{2}$$

where $\boldsymbol{D}_n = [\boldsymbol{Z}\hat{\boldsymbol{\Pi}}_2\ \ \boldsymbol{Z}_1]$. Let $\boldsymbol{y} = (y_1, \dots, y_n)$ and $\boldsymbol{d}_i^{\mathsf{T}}$ denote the ith row of $\boldsymbol{D}_n, i = 1, \dots, n$. The two-stage estimation techniques treat $\boldsymbol{V}_1 - \boldsymbol{Z}(\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1$ as regression errors.

The Welsh's trimmed mean is defined on a Winsorized observation with its construction based on an initial estimator of $\boldsymbol{\beta}$. It can also be seen that the trimmed mean developed in this section has asymptotic distribution independent of the initial estimator (see Welsh, 1987 for this property in the trimmed mean for the linear regression model). Compared with the trimmed least squares estimator of Koenker & Bassett (1978), which requires computing of the regression quantiles, this estimator has the advantage of computational ease if the initial estimator can be calculated by using the least squares estimator.

Let $\hat{\boldsymbol{\beta}}_0$ be an initial estimator of $\boldsymbol{\beta}$, treating $\boldsymbol{V}_1 - \boldsymbol{Z}(\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1$ in (2) as regression errors. The regression residuals are

$$\boldsymbol{e}_i = y_i - \boldsymbol{d}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}_0 \qquad (i = 1, \dots, n).$$

For large sample analysis in this paper, the initial estimator $\hat{\boldsymbol{\beta}}_0$ needs to satisfy the assumption $n^{1/2}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}) = O_p(1)$ (assumption (A5) in the next section). Choices of $\hat{\boldsymbol{\beta}}_0$ include consistent root n estimators such as the 2SLSE, the two-stage ℓ_1-norm estimator, minimizing a sum

© Australian Statistical Publishing Association Inc. 2001

of absolute values of residual terms that depend upon an initial estimator, and the Koenker–Bassett two-stage trimmed least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$. Amemiya (1982) demonstrated the asymptotic properties of the two-stage ℓ_1-norm estimator under the particular normal distribution of the error terms and Powell (1983) provided a general asymptotic theory for it. Chen & Portnoy (1996) gave a general asymptotic theory for the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$.

For $0 < \alpha < 0.5$, let $\hat{\eta}(\alpha)$ and $\hat{\eta}(1-\alpha)$ represent the αth and $(1-\alpha)$th empirical quantiles of the regression residuals, respectively. A Winsorized observation defined by Welsh (1987) is

$$y_i^* = y_i I\big(\hat{\eta}(\alpha) \le e_i \le \hat{\eta}(1-\alpha)\big) + \hat{\eta}(\alpha)\big(I\big(e_i < \hat{\eta}(\alpha)\big) - \alpha\big) \\ + \hat{\eta}(1-\alpha)\big(I\big(e_i > \hat{\eta}(1-\alpha)\big) - \alpha\big) \qquad (i = 1, \ldots, n). \tag{3}$$

Let $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^{\mathsf{T}}$, and denote the initial-estimator-based trimming matrix by $\boldsymbol{A} = \mathrm{diag}(a_1, \ldots, a_n)$, where $a_i = I(\hat{\eta}(\alpha) \le e_i \le \hat{\eta}(1-\alpha))$.

The Welsh's trimmed mean (Welsh, 1987) from the linear regression model to the simultaneous equations model is defined as

$$\hat{\boldsymbol{\beta}}_{\mathrm{W}} = (\boldsymbol{D}_n^{\mathsf{T}} \boldsymbol{A} \boldsymbol{D}_n)^{-1} \boldsymbol{D}_n^{\mathsf{T}} \boldsymbol{y}^* .$$

3. Asymptotic normality of two-stage Welsh's trimmed mean

Let f_j, F_j and F_j^{-1} represent the probability density function (pdf), cumulative distribution function (cdf) and inverse cdf, respectively, of v_j, for $j = 1, \ldots, p_0$. The following are some assumptions concerning the design matrix $\boldsymbol{Z}$, the error variables $v_1, \ldots, v_{p_0}$ and the first-stage estimator.

(A1) $n^{-1}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{Z} = \boldsymbol{Q} + o(1)$, where $\boldsymbol{Q}$ is positive definite; and the matrix $\boldsymbol{J} = \begin{bmatrix} \boldsymbol{\Pi}_2 & \begin{matrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{matrix} \end{bmatrix}$, is full rank.

(A2) $n^{-1}\sum_{i=1}^n z_{ij}^4 = O(1)$ for all j.

(A3) $n^{-1/4}\max_{ij}|z_{ij}| = O(1)$.

(A4) For $j = 1, \ldots, p_0$, f_j is symmetric at zero; and f_j and f_j' are both bounded away from 0 in a neighbourhood of $F_j^{-1}(\lambda)$ for $\lambda \in (0, 1)$.

(A5) Denote the partition $\hat{\boldsymbol{\Pi}}_2 = [\hat{\boldsymbol{\pi}}_2, \ldots, \hat{\boldsymbol{\pi}}_{p_0}]$. For $j = 2, \ldots, p_0$,

$$n^{1/2}(\hat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j) = \boldsymbol{Q}^{-1} n^{-1/2} \sum_{i=1}^n z_i \psi_j(v_{ji}) + o_p(1) ,$$

where function ψ_j satisfies $\mathrm{E}(\psi_j(v_j)) = 0$.

(A6) $n^{1/2}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}) = O_p(1)$.

Assumptions (A1)–(A4) are standard as given in Ruppert & Carroll (1980), Koenker & Portnoy (1987) and Welsh (1987). We let $\boldsymbol{\Sigma} = \boldsymbol{J}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{J}$. Assumption (A1) shows that $\boldsymbol{\Sigma}$ is also positive definite. Examples of $\hat{\boldsymbol{\Pi}}_2$ include the non-robust least squares estimator producing $\psi_j(v_j) = v_j$; the robust ℓ_1-norm estimator producing $\psi_j(v_j) = f_j^{-1}(0)(0.5 - I(v_j < 0))$; and the trimmed mean studied by Ruppert & Carroll (1980) producing $\psi_j(v_j) = \phi(v_j)$ with

© Australian Statistical Publishing Association Inc. 2001

$$\phi(v_j) = \begin{cases} F_j^{-1}(\alpha) & \text{if } v_j < F_j^{-1}(\alpha), \\ v_j & \text{if } F_j^{-1}(\alpha) \le v_j \le F_j^{-1}(1-\alpha), \\ F_j^{-1}(1-\alpha) & \text{if } v_j > F_j^{-1}(1-\alpha). \end{cases}$$

Theorem 3.1.

$$n^{1/2}(\hat{\boldsymbol\beta}_{\mathrm{W}} - \boldsymbol\beta) = \frac{1}{1-2\alpha} n^{-1/2} \boldsymbol\Sigma^{-1} \sum_{i=1}^{n} \tilde{\boldsymbol d}_i \phi(v_{1i}) - \boldsymbol\Sigma^{-1} \boldsymbol H^* n^{1/2}(\hat{\boldsymbol\Pi}_2 - \boldsymbol\Pi_2)\boldsymbol\beta_1 + o_p(1),$$

where $\tilde{\boldsymbol d}_i^{\mathsf T}$ is the ith row of matrix $[\boldsymbol Z \boldsymbol\Pi_2 \quad \boldsymbol Z_1]$ and $\boldsymbol H^* = \boldsymbol J^{\mathsf T} \boldsymbol Q$.

Corollary 3.2. *The Welsh's trimmed mean asymptotically has a normal distribution with mean 0 and covariance matrix given by $\sigma_{\mathrm W}^2 \boldsymbol\Sigma^{-1}$, where $\sigma_{\mathrm W}^2 = \boldsymbol\gamma^{\mathsf T} \mathrm{D}(\boldsymbol\phi^*) \boldsymbol\gamma \boldsymbol\Sigma^{-1}$ and $\boldsymbol\phi^* = ((1-2\alpha)^{-1}\phi(v_1), \psi_2(v_2), \ldots, \psi_{p_0}(v_{p_0}))$. For the case in which $\hat\Pi_2$ is the ℓ_1-norm estimator, $\mathrm{D}(\boldsymbol\phi^*) = \boldsymbol\Delta \mathrm{D}(\boldsymbol\phi_1)\boldsymbol\Delta$, where $\boldsymbol\Delta = \mathrm{diag}((1-2\alpha)^{-1}, f_2^{-1}(0), \ldots, f_{p_1}^{-1}(0))$ and $\boldsymbol\phi_1 = (\phi(v_{11}), \frac12 - I(v_{21} < 0), \ldots, \frac12 - I(v_{p_1 1} < 0))$. And $\mathrm{D}(\boldsymbol\phi_1) = [c_{ij}]$, where*

$$\begin{aligned} c_{11} &= 2\alpha\big(F_1^{-1}(\alpha)\big)^2 + \mathrm{E}\big(v_{11}^2 I\big(F_1^{-1}(\alpha) < v_{11} < F_1^{-1}(1-\alpha)\big)\big), \\ c_{1j} = c_{j1} &= -\big[F_1^{-1}(\alpha)\mathrm{E}\big(I\big(v_{11} < F_1^{-1}(\alpha), v_{ji} < 0\big)\big) \\ &\qquad + F_1^{-1}(1-\alpha)\mathrm{E}\big(I\big(v_{11} > F_1^{-1}(1-\alpha), v_{j1} < 0\big)\big) \\ &\qquad + \mathrm{E}\big(v_{11} I\big(F_1^{-1}(\alpha) < v_{11} < F_1^{-1}(1-\alpha), v_{j1} < 0\big)\big)\big], \\ c_{jj} = \tfrac14, &\quad c_{jk} = \mathrm{E}\big(I(v_{j1} < 0, v_{k1} < 0)\big) - \tfrac14 \qquad (j,k = 2, \ldots, p_1). \end{aligned}$$

Let $\hat{\boldsymbol\Pi}_2$ denote the matrix for the α-trimming estimator of Π_{2j}, based on either the Welsh's trimmed mean or the regression quantiles. Then we have matrix $\mathrm{D}(\boldsymbol\phi^*) = (1-2\alpha)^{-2}\mathrm{D}(\boldsymbol\phi)$, where $\boldsymbol\phi = (\phi(v_{11}), \ldots, \phi(v_{p_1}))$. When we further assume that $f_{jk}(x,y) = f_{jk}(-x,-y)$ for $(x,y) \in \mathbb{R}^2$, $j,k = 1, \ldots, p_1$, then $\mathrm{D}(\boldsymbol\phi^*) = (1-2\alpha)^{-2}\boldsymbol H$, where $\boldsymbol H = [h_{ij}]$, with

$$h_{jj} = 2\alpha\big(F_j^{-1}(\alpha)\big)^2 + \mathrm{E}\big(v_{j1}^2 I\big(F_j^{-1}(\alpha) < v_{j1} < F_j^{-1}(1-\alpha)\big)\big),$$

and for $j \ne k$ with $j,k = 1, \ldots, p_1$,

$$\begin{aligned} h_{jk} = \mathrm{E}\big(&v_{j1}v_{k1} I\big(F_j^{-1}(\alpha) < v_{j1} < F_j^{-1}(1-\alpha),\ F_k^{-1}(\alpha) < v_{k1} < F_k^{-1}(1-\alpha)\big)\big) \\ &+ 2F_j^{-1}(\alpha)F_k^{-1}(\alpha)\mathrm{E}\big(I\big(v_{j1} < F_j^{-1}(\alpha),\ v_{k1} < F_k^{-1}(\alpha)\big)\big) \\ &+ 2F_j^{-1}(\alpha)F_k^{-1}(1-\alpha)\mathrm{E}\big(I\big(v_{j1} < F_j^{-1}(\alpha),\ v_{k1} > F_k^{-1}(1-\alpha)\big)\big). \end{aligned}$$

4. Two-stage Welsh's trimmed mean as a robust Aitken estimator

Let (1) be premultiplied by matrix $\boldsymbol Z^{\mathsf T}$ to yield

$$\boldsymbol Z^{\mathsf T}\boldsymbol y = \boldsymbol Z^{\mathsf T}[\boldsymbol Y_1\ \boldsymbol Z_1]\boldsymbol\beta + \boldsymbol Z^{\mathsf T}\boldsymbol\tau\,. \qquad (4)$$

© Australian Statistical Publishing Association Inc. 2001

Denote the variance of variables τ by σ_τ^2. As interpreted by Fomby *et al.* (1984), the new explanatory variables in (4) consist essentially of sample cross moments between the endogenous variables and the exogenous variables — the former as they appear in (1), the latter as they appear in the entire system. The new explanatory variables divided by sample size n converge in probability to a non-stochastic limit and thus are uncorrelated with the error term appearing in (4). On the other hand, the covariance matrix of $\boldsymbol{Z}^\top\boldsymbol{\tau}$ is $\sigma_\tau^2\boldsymbol{Z}^\top\boldsymbol{Z}$ which makes the generalized least squares estimation appropriate.

Define the transformation $\boldsymbol{Z}^\top\boldsymbol{y}^*$ where $\boldsymbol{y}^*$ is given by (3). We see from (4) that

$$\boldsymbol{Z}^\top\boldsymbol{y}^* = \boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{D}_n\boldsymbol{\beta} + \boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{V}_1 - \boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{Z}(\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1 + \boldsymbol{Z}^\top\big(\hat{\boldsymbol{\eta}}(\alpha)\delta_\alpha - \hat{\boldsymbol{\eta}}(1-\alpha)\delta_{1-\alpha}\big), \quad (5)$$

where vector $\boldsymbol{\delta}_\lambda = (I(e_1 < \hat{\eta}(\lambda)) - \lambda, \ldots, I(e_n < \hat{\eta}(\lambda)) - \lambda)$.

The induced model for $\boldsymbol{Z}^\top\boldsymbol{y}^*$ is obtained through the large sample representations of $\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{Z}(\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1$ and $(\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{V}_1, \boldsymbol{Z}^\top(\hat{\boldsymbol{\eta}}(\alpha)\delta_\alpha - \hat{\boldsymbol{\eta}}(1-\alpha)\delta_{1-\alpha}))$, where the former produces errors in terms of $v_{2i}, \ldots, v_{p_0 i}$ and the latter produces errors in terms of v_{1i}.

Theorem 4.1. *For $\hat{\boldsymbol{\Pi}}_2$ satisfying assumption (A5), the following is an induced Aitken simultaneous equations model*

$$\boldsymbol{Z}^\top\boldsymbol{y}^* = \boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{D}_n\boldsymbol{\beta} + (1-2\alpha)\sum_{i=1}^n z_i\boldsymbol{\gamma}^\top\boldsymbol{\psi}^* + o_p(n^{1/2}).$$

The least squares estimator for the above induced model is

$$\hat{\boldsymbol{\beta}}_{\mathrm{WLS}} = (\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{y}.$$

Then $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{WLS}} - \boldsymbol{\beta})$ has normal asymptotic distribution with zero mean and covariance matrix

$$\boldsymbol{W} = \sigma_{\mathrm{W}}^2(\boldsymbol{J}^\top\boldsymbol{Q}^2\boldsymbol{J})^{-1}\boldsymbol{J}^\top\boldsymbol{Q}^3\boldsymbol{J}(\boldsymbol{J}^\top\boldsymbol{Q}^2\boldsymbol{J})^{-1},$$

which implies that the following estimator is a robust Aitken estimator, called Welsh's generalized trimmed mean,

$$\hat{\boldsymbol{\beta}}_{\mathrm{GW}} = (\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\hat{\boldsymbol{Q}}_\lambda\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\hat{\boldsymbol{Q}}_\lambda\boldsymbol{Z}^\top\boldsymbol{y}^*,$$

where $\hat{\boldsymbol{Q}}_\lambda$ satisfies $n^{-1}\hat{\boldsymbol{Q}}_\lambda = \lambda\boldsymbol{Q} + o_p(1)$ for some positive constant λ.

Corollary 4.2. *The least squares estimator for the induced model is*

$$\hat{\boldsymbol{\beta}}_{\mathrm{KBLS}} = (\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^\top\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{y},$$

and then $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{KBLS}} - \boldsymbol{\beta})$ has normal asymptotic distribution with zero mean and covariance matrix $\boldsymbol{W}$ larger than that of $\hat{\boldsymbol{\beta}}_{\mathrm{GW}}$.

Accordingly:

(1) The Welsh's generalized trimmed mean provides another robust Aitken estimator for the limited information simultaneous equations model.

© Australian Statistical Publishing Association Inc. 2001

(2) $\hat{\boldsymbol{\beta}}_{\mathrm{GW}}$ has asymptotic distribution exactly the same as that of the Koenker–Bassett (KB) trimmed least squares estimator (see Chen & Portnoy, 1996), which is then independent of the choice of initial estimator $\hat{\boldsymbol{\beta}}_0$.

(3) The least squares estimator for the induced model for $\boldsymbol{Z}^{\top}\boldsymbol{y}^*$ is

$$(\boldsymbol{D}_n^{\top}\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^{\top}\boldsymbol{A}\boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^{\top}\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^{\top}\boldsymbol{y}^*$$

which has asymptotic distribution exactly the same as that of $\hat{\boldsymbol{\beta}}_{\mathrm{KBLS}}$. This then implies that, in this induced transformed model, the least squares estimation is less efficient than the generalized least squares estimation.

The choice $\hat{\boldsymbol{Q}}_\lambda = \boldsymbol{Z}^{\top}\boldsymbol{A}\boldsymbol{Z}$ for $\hat{\boldsymbol{\beta}}_{\mathrm{GW}}$ satisfying $n^{-1}\hat{\boldsymbol{Q}}_\lambda = (1-2\alpha)\boldsymbol{Q} + o_p(1)$ generates the two-stage Welsh's trimmed mean which implies Theorem 4.3.

Theorem 4.3. *The two-stage Welsh's trimmed mean* $\hat{\boldsymbol{\beta}}_{\mathrm{W}}$ *is a robust Aitken estimator.*

The choice $\hat{\boldsymbol{Q}}_\lambda = \boldsymbol{Z}^{\top}\boldsymbol{Z}$ makes the following estimator

$$\left(\boldsymbol{D}_n^{\top}\boldsymbol{A}\boldsymbol{Z}(\boldsymbol{Z}^{\top}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\top}\boldsymbol{A}\boldsymbol{D}_n\right)^{-1}\boldsymbol{D}_n^{\top}\boldsymbol{A}\boldsymbol{Z}(\boldsymbol{Z}^{\top}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\top}\boldsymbol{y}^*$$

also a robust Aitken estimator.

5. Example and simulation

In this section we contrast the results for different estimates when applied to two datasets. First, consider the model for estimating supply of the commercial banks' loans to business firms in the United States for 1979–1984 (monthly data). The supply model is

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \tau\,.$$

This is a simultaneous equations model with endogenous variables y_1 and y_2 and exogenous variables z_1 and z_2, where $y_1 =$ total commercial loans (billions of US dollars), $y_2 =$ average prime rate charged by banks, $z_1 =$ 3-months treasury bill rate, $z_2 =$ total bank deposits. Economic theory expects β_2 to be negative, and β_1 and β_3 positive.

In this simultaneous equations model, there are also two available instrumental variables, z_3 and z_4. These variables represent an AAA corporate bond rate and an industrial production index, respectively. The reduced form model then takes a bivariate regression of (y_1, y_2) on exogenous variables $z_1, \dots, z_4$ associated with the intercept term. For details about the specifications of this model and the data, see Maddala (1988 p.314).

Maddala analysed this dataset through the least squares estimate (LSE) and two-stage least squares estimate (2SLSE). He concluded that quantity supplied is more responsive to changes in interest rates (see these two estimates in Table 1) than is suggested by the LSE. In Table 1, we also display the two-stage ℓ_1 estimate $(2\ell_1)$, two-stage trimmed least squares estimator based on regression quantile $(\hat{\beta}_{\mathrm{KB}}(\alpha))$, for $\alpha = 0.05(0.05)0.25$, and the two-stage Welsh's trimmed mean for number of the observations trimmed, $n_t = 1, 2, \dots, 6$.

Comparing these estimates in Table 1, we conclude the following:

(a) For β_1, β_2 and β_3, all robust estimates carry the expected sign. This is consistent with the LSE estimates made by Maddala. However, magnitudes of all robust estimates, in absolute terms, are smaller than those obtained using the 2SLSE method.

© Australian Statistical Publishing Association Inc. 2001

TABLE 1

Estimates for commercial loan data

Estimates	β_0	β_1	β_2	β_3
LSE	−77.414	2.415	−1.888	0.331
2SLSE	−87.988	6.905	−7.081	0.334
$2\ell_1$	−88.071	5.102	−4.706	0.335
$\hat{\beta}_{KB}(\alpha)$				
$\alpha = 0.05$	−90.723	6.504	−6.407	0.336
$= 0.10$	−84.083	5.766	−5.730	0.333
$= 0.15$	−92.925	5.888	−5.533	0.338
$= 0.20$	−86.117	5.727	−5.660	0.335
$= 0.25$	−91.375	6.165	−5.773	0.334
$\hat{\beta}_W$				
$n_t = 1$	−87.529	6.056	−5.928	0.334
$= 2$	−86.496	6.329	−6.299	0.334
$= 3$	−86.806	6.204	−5.923	0.332
$= 4$	−84.808	6.378	−6.180	0.331
$= 5$	−86.277	5.839	−5.508	0.332
$= 6$	−84.518	6.095	−5.865	0.331

(b) From the residuals computed from the two-stage ℓ_1-norm estimates, we see a few suspect outliers. The 2SLSEs of β_1 and β_2 are slightly larger than the robust estimates, while the two-stage ℓ_1-norm estimates are slightly smaller. From the theory of estimation, we can expect the 2SLSE to produce the worst estimates and the two-stage ℓ_1-norm to be inefficient.

(c) An important advantage of the two-stage Welsh's trimmed mean is that we can have the actual percentage of trimming close to any specified α. After performing a sequential trimming of $\hat{\beta}_{KB}(\alpha)$ and $\hat{\beta}_W$, the two-stage Welsh's trimmed means remain quite stable in the first six trimming estimates. We then expect that the two-stage Welsh's trimmed mean with a small number of trimmed observations is appropriate for estimating the parameters.

Next, we consider macroeconomic data for 1970–1984 in the United States (see Gujarati, 1988 p.568) where the income model is

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \tau,$$

where y_1 = income, y_2 = stock of money, z_1 = investment expenditure, z_2 = government expenditure on goods and services. This model states that income is determined by the endogenous variable y_2 and two exogenous variables z_1 and z_2. Table 2 displays the various estimates for parameters of this model.

Based on Table 2, we can conclude the following:

(a) Estimates of 2SLSE with negative signs are larger and those with positive signs are smaller than most of the corresponding robust estimates. This reveals the non-robustness of this least-squares-type estimation method.

(b) The two-stage ℓ_1-norm is quite satisfactory for these data.

(c) Again, the stability of the two-stage Welsh's trimmed mean reveals that smaller number trimming is appropriate for these data; however, this property is not shown by trimmed least squares.

© Australian Statistical Publishing Association Inc. 2001

TABLE 2

Estimates for money income data

Estimate	β_0	β_1	β_2	β_3
2SLSE	−0.034	−0.227	1.379	4.090
$2\ell_1$	−0.052	−1.125	1.636	4.654
$\hat{\beta}_{\mathrm{KB}}(\alpha)$				
$\alpha = 0.05$	−0.053	−1.168	1.704	4.655
$= 0.10$	−0.033	−0.125	1.370	4.009
$= 0.15$	−0.046	−0.761	1.395	4.538
$= 0.20$	−0.044	−0.638	1.323	4.479
$= 0.25$	−0.054	−1.208	1.603	4.755
$= 0.30$	−0.054	−1.237	1.673	4.714
$\hat{\beta}_{\mathrm{W}}$				
$n_t = 1$	−0.045	−0.691	1.245	4.626
$= 2$	−0.050	−1.032	1.599	4.639
$= 3$	−0.051	−1.080	1.626	4.629
$= 4$	−0.051	−1.093	1.635	4.627
$= 5$	−0.052	−1.108	1.613	4.656
$= 6$	−0.052	−1.145	1.646	4.659

To study two-stage estimators for the simultaneous equations model with asymmetric error distributions, we performed a Monte Carlo simulation for the simple simultaneous equations model $y = \beta_0 + \beta_1 y_1 + \beta_2 z_1 + \beta_3 z_2 + \tau$ with reduced form $[\boldsymbol{y}\ \boldsymbol{y}_1] = [\mathbf{1}\ \boldsymbol{z}_1\ \boldsymbol{z}_2][\boldsymbol{\Pi}_1\ \boldsymbol{\Pi}_2] + [\boldsymbol{v}_1\ \boldsymbol{v}_2]$. We let (u_1, u_2) denote a vector of independent exponential random variables with mean 1. We assume that the error vector in the reduced form follows the following mixture model

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{cases} \begin{bmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 - 1 \\ u_2 - 1 \end{bmatrix} & \text{with probability } 1-\delta, \\ s \begin{bmatrix} u_1 - 1 \\ u_2 - 1 \end{bmatrix} & \text{with probability } \delta. \end{cases}$$

This ensures that (v_1, v_2) has an asymmetric distribution with mean 0 and probability $(1-\delta)$ from a distribution with covariance matrix $\left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$, and probability δ from a distribution with covariance matrix $s^2 I_2$, where large values of s may produce outliers.

We take $(\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2)$ such that $\beta_j = 0.5$, $j = 0, 1, 2, 3$, and we use sample size $n = 40$ and samples (z_1, z_2) randomly generated from a bivariate normal distribution. With 1000 replications, we generate observations (y, y_1, z_1, z_2), obeying the assumptions above, and, for estimating parameters β_j, $j = 0, 1, 2, 3$, we compute the two-stage ℓ_1-norm estimates, the two-stage trimmed LSE and the two-stage Welsh's trimmed mean. Table 3 displays the results in terms of average mean squared errors (MSE).

The outliers produced by asymmetric distributions are, in general, unbalanced with respect to the population mean. Therefore none of the estimators is very efficient for estimating the population mean, though the two-stage ℓ_1-norm estimator is relatively less efficient because it, in fact, estimates the population median. On the other hand, both the two-stage Welsh's trimmed mean and the two-stage trimmed least squares estimator are quite promising and are very competitive in estimating the population mean for this asymmetric distribution.

© Australian Statistical Publishing Association Inc. 2001

TABLE 3

MSE for two-stage estimators

Estimator	$\delta = 0$	$\delta = 0.1$ $s = 3$	$\delta = 0.1$ $s = 10$	$\delta = 0.2$ $s = 3$	$\delta = 0.2$ $s = 10$
$2\ell_1$	20.70	7.047	23.98	10.370	35.81
$\hat{\beta}_{\mathrm{KB}}(\alpha)$					
$\alpha = 0.1$	0.405	0.372	0.436	4.374	1.985
$= 0.2$	0.551	1.94	3.371	0.526	0.814
$= 0.3$	0.395	2.506	2.143	0.911	0.685
$\hat{\beta}_{\mathrm{W}}$					
$n_t = 2$	0.155	0.187	1.563	0.173	2.180
$= 4$	0.286	0.396	2.870	0.431	1.436
$= 6$	0.673	0.745	1.632	0.904	0.952
$= 8$	0.993	1.080	0.687	1.480	0.683

6. Large sample inference

Here we sketch some large-sample methods for confidence ellipsoids and hypothesis testing based on the two-stage Welsh's trimmed mean. First assume that we have a statistic V which is a consistent estimator of the asymptotic covariance matrix $\boldsymbol{\gamma}^{\mathsf{T}}\mathrm{D}(\boldsymbol{\phi}^*)\boldsymbol{\gamma}\boldsymbol{\Sigma}^{-1}$. For $0 < \lambda < 1$, let

$$d_\lambda(r, s) = \frac{r}{1-2\alpha}\, c_{1-\lambda}(\mathrm{F}_{r,s}),$$

where $c_q(\mathrm{F}_{r,s})$ denotes the q-quantile of the $\mathrm{F}_{r,s}$ distribution. Suppose for some integer ℓ, $\boldsymbol{K}$ is an $\ell \times p$ matrix of rank ℓ. Let m be the number of residuals e_i lying outside the interval $(\hat{\eta}(\alpha), \hat{\eta}(1-\alpha))$. Then

$$\Pr\big((\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{\beta})^{\mathsf{T}}\boldsymbol{K}^{\mathsf{T}}(\boldsymbol{K}\boldsymbol{V}^{-1}\boldsymbol{K}^{\mathsf{T}})^{-1}\boldsymbol{K}(\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{\beta}) \le d_u(\ell, n-m-p)\big) \approx 1 - u.$$

If $\boldsymbol{K} = \boldsymbol{I}_p$, the confidence ellipsoid for $\boldsymbol{\beta}$ is given by

$$(\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{\beta})^{\mathsf{T}}\boldsymbol{V}^{-1}(\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{\beta}) \le d_u(\ell, n-m-p).$$

Moreover, if we test $H_0\colon \boldsymbol{K}\boldsymbol{\beta} = \boldsymbol{v}$ by rejecting H_0 whenever

$$(\boldsymbol{K}\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{v})^{\mathsf{T}}(\boldsymbol{K}\boldsymbol{V}^{-1}\boldsymbol{K}^{\mathsf{T}})^{-1}(\boldsymbol{K}\hat{\boldsymbol{\beta}}_{\mathrm{W}} - \boldsymbol{v}) \ge d_u(\ell, n-m-p),$$

then this test has an asymptotic size of u.

We still need to have an estimator of asymptotic covariance matrix $\boldsymbol{\gamma}^{\mathsf{T}}\mathrm{D}(\boldsymbol{\phi}^*)\boldsymbol{\gamma}\boldsymbol{\Sigma}^{-1}$. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{J}}$ denote the two-stage Welsh's trimmed mean of $\boldsymbol{\beta}_1$ and the matrix $\boldsymbol{J}$ in (A1), replacing $\boldsymbol{\Pi}_2$ by $\hat{\boldsymbol{\Pi}}_2$, respectively. The estimator of $\boldsymbol{\Sigma}$ is given by $\hat{\boldsymbol{\Sigma}} = n^{-1}\hat{\boldsymbol{J}}^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{Z}\hat{\boldsymbol{J}}$. It remains to estimate the matrix $\mathrm{D}(\boldsymbol{\phi}^*)$. Suppose that here $\hat{\boldsymbol{\Pi}}_2$ is the trimmed least squares estimator for Π_2. Denote the residual matrix as:

$$\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p_1 1} \\ e_{12} & e_{22} & \cdots & e_{p_1 2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{p_1 n} \end{bmatrix} = [\boldsymbol{y} - \hat{\boldsymbol{D}}_n\hat{\boldsymbol{\beta}}_{\mathrm{W}} \quad \boldsymbol{Y}_1 - \boldsymbol{Z}\hat{\boldsymbol{\Pi}}_2].$$

© Australian Statistical Publishing Association Inc. 2001

Also, let $\hat{\eta}_j(\lambda)$ be the λth empirical quantile based on residuals e_{ji}, $i = 1, \ldots, n$ and define

$$\hat{\phi}_j(a) = \begin{cases} \hat{\eta}_j(\alpha) & \text{if } a < \hat{\eta}_j(\alpha), \\ a & \text{if } \hat{\eta}_j(\alpha) \le a \le \hat{\eta}_j(1-\alpha), \\ \hat{\eta}_j(1-\alpha) & \text{if } a > \hat{\eta}_j(1-\alpha). \end{cases}$$

An estimator of C where $\hat{\Pi}_2$ is the trimmed least squares estimator is given by $\hat{\mathrm{D}}(\phi^*) = (1-2\alpha)^{-2} n^{-1} \sum_{i=1}^n \hat{\boldsymbol{\phi}}\hat{\boldsymbol{\phi}}^{\mathsf{T}}$, where $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1(e_{1i}), \ldots, \hat{\phi}_{p_1}(e_{p_1 i}))$. On the other hand, if we consider that $\hat{\Pi}_2$ is the ℓ_1-norm estimator of Π_2, let $\hat{\mathrm{D}}(\boldsymbol{\phi}_1) = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\phi}}_1 \hat{\boldsymbol{\phi}}_1^{\mathsf{T}}$, where $\hat{\boldsymbol{\phi}}_1 = (\hat{\phi}_1(e_{1i}), \frac{1}{2} - I(e_{2i} < 0), \ldots, \frac{1}{2} - I(e_{p_1 i} < 0))$. We also let $\hat{f}_j(0)$ be the estimator of $f_j(0)$ for $j = 2, \ldots, p_1$; it can be the estimator by Koenker & Portnoy (1987), Welsh (1991) or Chen (1997). Then $\hat{\mathrm{D}}(\boldsymbol{\phi}^*)$, the estimator of $\mathrm{D}(\boldsymbol{\phi}^*)$ when we use ℓ_1-norm estimator for $\boldsymbol{\Pi}_2$, equals

$$\mathrm{diag}\big((1-2\alpha)^{-1}, \hat{f}_2^{-1}(0), \ldots, \hat{f}_{p_1}^{-1}(0)\big)\hat{\mathrm{D}}(\boldsymbol{\phi}_1)\,\mathrm{diag}\big((1-2\alpha)^{-1}, \hat{f}_2^{-1}(0), \ldots, \hat{f}_{p_1}^{-1}(0)\big).$$

7. Appendix

Proof of Theorem 4.1. Let

$$H(t) = n^{-1/2}\sum_{i=1}^n \big(\gamma - I(v_{1i} < F_1^{-1}(\gamma) + n^{-1/2} z_i^{\mathsf{T}} \boldsymbol{t})\big).$$

From Jurečková (1984), we have

$$H(T) - H(0) - f_1\big(F_1^{-1}(\gamma)\big) n^{-1}\sum_{i=1}^n z_i^{\mathsf{T}} \boldsymbol{T} = o_p(1)$$

for any random vector $\boldsymbol{T}$ with $\boldsymbol{T} = O_p(1)$. From Ruppert & Carroll (1980), we have

$$n^{-1/2}\sum_{i=1}^n \big(\gamma - I(e_i < \hat{\eta}(\gamma)\big) = o_p(1). \tag{6}$$

By rearrangement, the following equation holds

$$e_i - \hat{\eta}(\gamma) = v_{1i} - F_1^{-1}(\gamma) - n^{-1/2} z_i^{\mathsf{T}} \boldsymbol{T}_n(\gamma)$$

with
$$\boldsymbol{T}_n(\gamma) = n^{1/2}\big[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}) - (\hat{\boldsymbol{\Pi}}_2 - \boldsymbol{\Pi}_2)\boldsymbol{\beta}_1 + \big(\hat{\eta}(\gamma) - F_1^{-1}(\gamma)\big)\boldsymbol{a}\big], \tag{7}$$

where $\boldsymbol{a}$ is a vector of zeros except for the first element which is 1. The method of Jurečková (1977 proof of Lemma 5.2) and (7) also show that for $\delta > 0$, there exist positive values s, k and N_0 such that

$$\Pr\bigg(\inf_{|t|\ge k} n^{-1/2}\bigg|\sum_{i=1}^n \big(\gamma - I\big(v_{1i} < F_1^{-1}(\gamma) + z_i^T t\big)\big)\bigg| < s\bigg) < \delta \tag{8}$$

© Australian Statistical Publishing Association Inc. 2001

for $n \geq N_0$. From (6) and (8), it is seen that $T_n(\gamma) = O_p(1)$ which implies that $\hat{\eta}(\gamma)$ is consistent for $F^{-1}(\gamma)$, as

$$\hat{\eta}(\gamma) = F_1^{-1}(\gamma) + o_p(1).$$

Let
$$M(\boldsymbol{t}, \gamma) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{z}_i v_{1i} I\big(v_{1i} \leq F_1^{-1}(\gamma) + n^{-1/2} \boldsymbol{z}_i^{\top} \boldsymbol{t}\big).$$

Now
$$n^{-1/2} \boldsymbol{Z}^{\top} \boldsymbol{A} \boldsymbol{V}_1 = M\big(\boldsymbol{T}_n(1-\alpha), 1-\alpha\big) - M\big(\boldsymbol{T}_n(\alpha), \alpha\big). \tag{9}$$

From Ruppert & Carroll (1980) and Jurečková (1984), we have

$$M(\boldsymbol{T}, \gamma) - M(\boldsymbol{0}, \gamma) = F_1^{-1}(\gamma) f_1\big(F_1^{-1}(\gamma)\big) n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_i^{\top} \boldsymbol{T} + o_p(1) \tag{10}$$

for any sequence $\boldsymbol{T}$ with $\boldsymbol{T} = O_p(1)$. Then (9) and (10) induce the following,

$$\begin{aligned} n^{-1/2} \boldsymbol{Z}^{\top} \boldsymbol{A} \boldsymbol{V}_1 = {} & F_1^{-1}(1-\alpha) f_1\big(F_1^{-1}(1-\alpha)\big) n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{T}_n(1-\alpha) \\ & - F_1^{-1}(\alpha) f_1\big(F_1^{-1}(\alpha)\big) n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{T}_n(\alpha) \\ & + n^{-1/2} \sum_{i=1}^{n} z_i v_{1i} I\big(F_1^{-1}(\alpha) \leq v_{1i} \leq F_1^{-1}(1-\alpha)\big) + o_p(1). \end{aligned} \tag{11}$$

Similarly,

$$\begin{aligned} & n^{-1/2} \hat{\eta}(\alpha) \boldsymbol{Z}^{\top} \boldsymbol{\delta}_{\alpha} + n^{-1/2} \hat{\eta}(1-\alpha) \boldsymbol{Z}^{\top} \boldsymbol{\delta}_{1-\alpha} \\ & \quad = F_1^{-1}(\alpha) f_1\big(F_1^{-1}(\alpha)\big) n^{-1} \sum_{i=1}^{n} z_i T_n(\alpha) - F_1^{-1}(1-\alpha) f_1\big(F_1^{-1}(1-\alpha)\big) \\ & \qquad n^{-1} \sum_{i=1}^{n} z_i T_n(1-\alpha) + F_1^{-1}(\alpha) n^{-1/2} \sum_{i=1}^{n} z_i \big(I(v_{1i} < F_1^{-1}(\alpha)\big) - \alpha\big) \\ & \qquad + F_1^{-1}(1-\alpha) n^{-1/2} \sum_{i=1}^{n} z_i \big(I\big(v_{1i} > F_1^{-1}(1-\alpha)\big) - \alpha\big) + o_p(1). \end{aligned} \tag{12}$$

Combining (11) and (12), we have

$$n^{-1/2} \boldsymbol{Z}^{\top} \boldsymbol{A} \boldsymbol{V}_1 + n^{-1/2} \hat{\eta}(\alpha) \boldsymbol{Z}^{\top} \boldsymbol{\delta}_{\alpha} + n^{-1/2} \hat{\eta}(1-\alpha) \boldsymbol{Z}^{\top} \boldsymbol{\delta}_{1-\alpha} = n^{-1/2} \sum_{i=1}^{n} z_i^{\top} \boldsymbol{\phi}(v_{1i}) + o_p(1). \tag{13}$$

Then the induced form of Theorem 4.1 follows from (5), (13) and because $n^{-1} \boldsymbol{Z}^{\top} \boldsymbol{A} \boldsymbol{Z} = (1-2\alpha) \boldsymbol{Q} + o_p(1)$ and $\hat{\boldsymbol{\Pi}}_2 = \boldsymbol{\Pi}_2 + o_p(1)$.

© Australian Statistical Publishing Association Inc. 2001

References

AMEMIYA, T. (1982). Two stage least squares deviations estimators. *Econometrica* **50**, 689–711.

AMEMIYA, T. (1985). *Advanced Econometrics.* Massachusetts: Harvard University Press.

CHEN, L-A. (1997). An efficient class of weighted trimmed means for linear regression models. *Statistica Sinica* **7**, 669–686.

CHEN, L-A. & PORTNOY, S. (1996). Regression quantiles and trimmed least squares estimators for structural equations models. *Comm. Statist. Simulation Comput.* **25**, 1005–1032.

FOMBY, T.B., HILL, R.C. & JOHNSON, S.R. (1984). *Advanced Econometric Methods.* New York: Springer-Verlag.

GUJARATI, D.N. (1988). *Basic Econometrics.* New York: McGraw-Hill.

JUREČKOVÁ, J. (1977). Asymptotic relation of M-estimates and R-estimates in a linear regression model. *Ann. Statist.* **5**, 464–472.

JUREČKOVÁ, J. (1984). Regression quantiles and trimmed least squares estimators under a general design. *Kybernetika* **20**, 345–357.

KOENKER, R.W. & BASSETT, G.W. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

KOENKER, R. & PORTNOY, S. (1987). L-estimators for linear models. *J. Amer. Statist. Assoc.* **82**, 851–857.

KRASKER, W.S. (1985). Two stage bounded-influence estimators for simultaneous equations models. *J. Bus. Econom. Statist.* **4**, 432–444.

MADDALA, G.S. (1988). *Introduction to Econometrics.* New York: Macmillan Publishing Company.

POWELL, J.L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica* **51**, 1569–1576.

RUPPERT, D. & CARROLL, R.J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75**, 828–838.

WELSH, A.H. (1987). The trimmed mean in the linear model. *Ann. Statist.* **15**, 20–36.

WELSH, A.H. (1991). Asymptotically efficient adaptive L-estimators in the linear model. *Statist. Sinica* **1**, 203–228.

© Australian Statistical Publishing Association Inc. 2001