Econometrica, Vol. 52, No. 3 (May, 1984)

THE STRUCTURE OF SIMULTANEOUS EQUATION ESTIMATORS: A GENERALIZATION TOWARDS NONNORMAL DISTURBANCES

BY INGMAR R. PRUCHA AND HARRY H. KELEJIAN

A general linear simultaneous equation system with a multivariate Student t disturbance vector is considered. The normal equations of the corresponding maximum likelihood estimator are used as estimator generating equations to introduce a new class of estimators. Properties of large subclasses of these estimators are determined for disturbance vectors other than the multivariate Student t.

1. INTRODUCTION[1]

IN A SEMINAL PAPER Hendry [9] unified nearly all existing estimation theory for linear simultaneous equation systems. The starting point of his analysis was the full information maximum likelihood estimator derived under the assumption of normally distributed disturbances, henceforth the NFIML estimator. Hendry then used the normal equations of the NFIML estimator as estimator generating equations to define a wide class of estimators, henceforth NFIML_A estimators. The NFIML_A estimators can be interpreted as numerical approximations to the NFIML estimator. Hendry then showed that virtually all known estimators for linear systems belong to the NFIML_A class; furthermore, based on easily determined characteristics of the approximation, the NFIML_A class can be subdivided into asymptotically equivalent classes. Thus Hendry's [9] approach both unified and simplified existing estimation theory in that it made case by case analysis of large sample properties of estimators unnecessary.

Despite the wide use of NFIML_A estimators typical arguments supporting the assumption of normally distributed disturbances are not fully convincing.[2] Even if one assumes that the disturbance terms are composed as a sum of independent and additive variables the usual appeal to a central limit theorem can at most imply that the distribution of the disturbances is approximately normal—the stress is on the word approximately.[3] Observed disturbance distributions exhibit, in fact, often thicker tails than is consistent with normality.[4]

The possibility of nonnormal disturbances is consequential. As has been

[1]Certain parts of this paper were first derived in Kelejian and Prucha [14], and have been presented at the 4th World Congress of the Econometric Society in Aix-en-Provence, 1980. We would like to thank Benedikt Poetscher and the Editor and referees of this journal for helpful comments. We also benefited from discussions with Manfred Deistler, Phoebus Dhrymes, and James Ramsey at an early stage of our research. We retain, however, full responsibility for any shortcomings. The support of computer time through the facilities of the Computer Science Center of the University of Maryland is gratefully acknowledged. Detailed proofs are suppressed at the suggestion of the Editor, but are available on request from the authors.

[2]See Bartels [2] and Goldfeld and Quandt [6] on this account.

[3]See Huber [10, p. 74].

[4]See Judge *et al.* [13, pp. 297–321] and the references cited therein.

pointed out in the robustness literature the (asymptotic) efficiency of estimators derived under the assumption of normally distributed disturbances is generally sensitive to deviations from this assumption.[5] In particular, if the actual disturbance distribution has thick tails "outliers" are more frequent; estimators which correspond to the normality assumption tend to place too much weight on those "outliers."[6] In addition, the normal distribution is completely defined by the first two moments; thus, estimators which correspond to it also fail to utilize sample information beyond those moments.

The above considerations motivate interest in estimators which correspond to a more general disturbance distribution which allows for thicker tails and, preferably, contains the normal distribution as a special case. The multivariate t is such a distribution. The thickness of the tails of the t distribution is characterized by the degrees of freedom parameter. The normal distribution is obtained as a limiting case as the degrees of freedom parameter tends to infinity.

In general terms, we first derive the maximum likelihood estimator corresponding to a disturbance distribution which is more general than the normal; we then determine the large sample properties of feasible counterparts to this estimator when the disturbance distribution is only taken to be symmetric given the existence of certain moments.

More specifically, we first derive the full information maximum likelihood estimator, henceforth the TFIML estimator, for a system of linear simultaneous equations under the assumption that the disturbance distribution is multivariate t with known degrees of freedom. As expected, the NFIML estimator corresponds to the limiting case of this estimator as the degrees of freedom parameter tends to infinity. Somewhat less expectedly, it turns out that the TFIML estimator has an instrumental variable form, a result which generalizes Hausman's [**7**, **8**] findings for the multivariate normal case. It also turns out that the TFIML estimator is a robust estimator in that it places less weight on large disturbance values than does the NFIML estimator.

After obtaining and interpreting the TFIML estimator we drop the multivariate t assumption. In doing so, we redefine the degrees of freedom parameter such that it has meaning as a measure of the thickness of the tails for general disturbance distributions without, however, changing its original interpretation in case the disturbance distribution is in fact multivariate t.[7] We then use the

[5]See Huber [**11**] and the references cited therein. For a nice review of the robustness literature, see Judge et al. [**13**, pp. 303–308], and Maddala [**15**, pp. 305–314]. In the single equation case various robust procedures have been proposed as an alternative to OLS. In the simultaneous equation context the literature on robust procedures is rather limited. Amemiya [**1**] considers the two-stage-least-absolute-deviation estimator. Fair [**3**] introduced a very general class of robust estimators and performed several (specific) Monte Carlo experiments. Analytically the (asymptotic) properties of the latter estimators have not yet been derived.

[6]As a consequence such estimators will vary substantially from sample to sample and so their variances will be large.

[7]In particular we define the degrees of freedom parameter as a function of second and first absolute moments. It should however be noted that our analysis is also valid for alternative (reasonable) definitions.

normal equations of the TFIML estimator as estimator generating equations to define a class of feasible estimators, henceforth TFIML_{A} estimators, which can be viewed as numerical approximations to the TFIML estimator. The (redefined) degrees of freedom parameter is replaced by an estimator so that a priori knowledge of this parameter is no longer required. Somewhat expectedly, the TFIML_{A} class contains the NFIML_{A} class as a subclass; it therefore includes not only virtually all known estimators, but corresponding generalizations as well, e.g., generalized three stage least squares.

We prove the consistency of a large subclass of TFIML_{A} estimators and derive the asymptotic distribution of a somewhat smaller subclass of linearized TFIML_{A} estimators for symmetric disturbance distributions with, respectively, finite fourth and fifth moments.[8] The members of the latter subclass are shown to be asymptotically efficient for the case in which the disturbance distribution is multivariate t or multivariate normal; we henceforth refer to those estimators as linearized TFIML_{E} estimators. This implies that in case of multivariate t or normal disturbances (feasible) linearized TFIML_{E} estimators use sample information which relates to both the first two moments as well as the thickness of the tails of the disturbance distribution in a fully efficient manner.

We hypothesize that linearized TFIML_{E} estimators are asymptotically efficient relative to the NFIML_{A} class (including the NFIML estimator) for a wide class of symmetric disturbance distributions. At this point a formal general description of this class is not available. We do, however, calculate for the single equation case the relative efficiency of linearized TFIML_{E} estimators as compared to the NFIML or OLS estimator for various disturbance distributions. In all cases considered linearized TFIML_{E} estimators are efficient relative to the OLS estimator, except when the disturbance distribution is normal, in which case linearized TFIML_{E} estimators and the OLS estimator are asymptotically equivalent. Indeed, in some cases we find the asymptotic variances of the OLS estimator to be one hundred times larger than those of linearized TFIML_{E} estimators.

Clearly our choice of the multivariate t family as a "generating" distribution is not the only possible one. Other long-tailed distributions may serve as well and further research in that direction is needed. Our results suggest however that the choice of the multivariate t family was a reasonable one.

The paper is organized as follows: Section 2 gives the specification of the model. The TFIML estimator and its large sample distribution are given in Section 3. In Section 4 the class of TFIML_{A} estimators is defined. The asymptotic properties of wide subclasses of those estimators are analyzed for general symmetric disturbance distributions. Concluding remarks and suggestions for further research are given in Section 5.

[8] We conjecture that the present moment requirements can be reduced by using more elaborate proofs. For such a reduction in the single equation case see Poetscher and Prucha [**16**]. We note that the present moment requirements are satisfied by, e.g., ϵ-contaminated normal distributions often used to model the existence of outliers.

2. MODEL ASSUMPTIONS AND NOTATION

We consider the following linear simultaneous equation model:

$$Y = YB + ZC + U \tag{2.1}$$

where Y and Z are $T \times M$ and $T \times K$ matrices of the M endogenous and K exogenous variables of the model over T periods; U is the $T \times M$ matrix of disturbances; B and C are $M \times M$ and $K \times M$ matrices of parameters. As a normalization rule the diagonal elements of B are taken to be zero. We further assume that $I - B$ is nonsingular so that (2.1) has the reduced form representation

$$Y = Z\Pi + V, \qquad \Pi = C(I-B)^{-1}, \qquad V = U(I-B)^{-1}. \tag{2.2}$$

It will be convenient to adopt the following notation: Let N be some matrix; then $n_{i.}$ and $n_{.j}$ denote the ith row and jth column respectively. Similarly $n^{i.}$ and $n^{.j}$ denote the ith row and jth column of the inverse matrix N^{-1}. The (i, j)th elements of N and N^{-1} will be denoted by n_{ij} and n^{ij} respectively.

We assume that the vectors of disturbances $u_{t.}$ are distributed i.i.d. with zero mean and nonsingular bounded covariance matrix Σ. We further assume that the distribution of the disturbances is symmetric with finite fourth moments. The exogenous variables are taken to be nonstochastic and uniformly bounded, i.e. $\sup_{t,i}(z_{ti}) < \infty$. Further we assume $\lim_{T\to\infty} Z'Z/T = Q$ where Q is a finite positive definite matrix.

We assume that every equation is identified subject to zero-type parameter restrictions. Hence the ith structural equation of (2.1) can be expressed as

$$y_{.i} = Yb_{.i} + Zc_{.i} + u_{.i} = Y_i\beta_i + Z_i\gamma_i + u_{.i} \qquad (i = 1, \ldots, M), \tag{2.3}$$

where Y_i and Z_i denote, respectively, the $T \times M_i$ and $T \times K_i$ matrices of observations on the endogenous and exogenous variables that appear as regressors in the ith equation; β_i and γ_i are the corresponding $M_i \times 1$ and $K_i \times 1$ vectors of unrestricted (nonzero) parameters.

It will be convenient for our later discussion to introduce selector matrices L_{i1} and L_{i2} such that $Y_i = YL_{i1}$, $Z_i = ZL_{i2}$, $b_{.i} = L_{i1}\beta_i$, and $c_{.i} = L_{i2}\gamma_i$ for $i = 1, \ldots, M$. The selector matrices L_{i1} and L_{i2} are of order $M \times M_i$ and $K \times K_i$ respectively; their elements are zeros and ones in appropriate places.

It will also be convenient to adopt notation concerning block diagonal matrices and the vectorization of matrices. Let the matrices N_i, $i = 1, \ldots, M$, be of order $r_i \times s_i$; then we define $\operatorname{diag}_M(N_i) = \operatorname{diag}(N_1, \ldots, N_M)$ as the $\sum r_i \times \sum s_i$ block diagonal matrix whose ith block is N_i. Let N be an $r \times s$ matrix; then we define $\operatorname{vec}(N) = [n'_{.1}, \ldots, n'_{.s}]'$. Given this notation, the equations of (2.1) can be written as

$$(2.4) \qquad y = X\delta + u$$

where $y = \text{vec}(Y)$, $X = \text{diag}_M(X_i)$, $X_i = [Y_i, Z_i]$, $u = \text{vec}(U)$, and $\delta = [\beta_1', \gamma_1', \ldots, \beta_M', \gamma_M']'$. The elements of B and C relate to those of δ as $\text{vec}[(B', C')'] = \text{diag}_M(L_i)\delta$ where $L_i = \text{diag}(L_{i1}, L_{i2})$.

3. MAXIMUM LIKELIHOOD ESTIMATION FOR STUDENT t DISTURBANCES

In this section we give and interpret the full information maximum likelihood (TFIML) estimator of δ and Σ under the assumption that the disturbances are distributed multivariate t with known degrees of freedom $v > 5$. This distributional assumption will not be maintained in the subsequent sections of the paper.

Assume the disturbances are multivariate t, and so their density is given by

$$(3.1) \qquad f(u_{t.}) = \frac{v^{v/2}\Gamma[(M+v)/2]|\Theta|^{1/2}}{\Pi^{M/2}\Gamma[v/2]}\left[v + u_{t.}\Theta u_{t.}'\right]^{-(M+v)/2},$$

$$\Theta = \frac{v}{v-2}\Sigma^{-1},$$

where $\Gamma(\cdot)$ denotes the Gamma function.[9] Note that the density of the multivariate normal distribution represents a limiting case of (3.1) as $v \to \infty$. The sample log-likelihood function of (2.1) subject to (3.1) is given by

$$(3.2) \qquad \mathscr{L}(B, C, \Sigma; Y, Z) = \text{constant} + \frac{T}{2}\log|\Sigma^{-1}| + T\log|I - B| + \frac{M+v}{2}\sum_{t=1}^{T}\log(w_t),$$

$$w_t = \left[1 + u_{t.}\left[(v-2)\Sigma\right]^{-1}u_{t.}'\right]^{-1}, \qquad u_{t.} = y_{t.} - y_{t.}B - z_{t.}C.$$

The TFIML estimators of δ and Σ, say $\hat{\delta}$ and $\hat{\Sigma}$, are defined as the maximizing values of (3.2) for given observations on Y and Z. The proofs of the following two theorems are available on request from the authors.[10]

THEOREM 3.1: *Consider the model of Section* 2 *and assume* (3.1). *Then the*

[9] Compare Raiffa and Schlaifer [**17**, pp. 256–258]. For ease of notation we do not distinguish between the random variables and the values they assume.

[10] The manipulations involved in proving Theorem 3.1 are somewhat similar to those of Hausman [**7, 8**]. In proving Theorem 3.2 we use standard results on the $\sqrt{T}$-consistency of maximum likelihood estimators. The exact asymptotic distribution is derived via an expansion of the normal equations using standard central limit theorems. An alternative derivation of the asymptotic variance covariance matrix via the information matrix is given in Kelejian and Prucha [**14**].

TFIML *estimators* $\hat{\delta}$ *and* $\hat{\Sigma}$ *satisfy the following system of normal equations*:

(3.3a)
$$\hat{\delta}=\left[\hat{X}'(\hat{\Sigma}^{-1}\otimes\hat{W})X\right]^{-1}\hat{X}'(\hat{\Sigma}^{-1}\otimes\hat{W})y,$$
$$\hat{X}=\operatorname{diag}_M(\hat{X}_i),\qquad \hat{X}_i=[\hat{Y}_i,Z_i],\qquad \hat{Y}_i=Z\hat{\Pi}_i,\qquad \hat{\Pi}_i=\hat{\Pi}L_{i1},$$
$$\hat{\Sigma}=T^{-1}\frac{M+v}{v-2}\left[Y-Y\hat{B}-Z\hat{C}\right]'\hat{W}\left[Y-Y\hat{B}-Z\hat{C}\right],$$
$$\hat{W}=\operatorname{diag}_T(\hat{w}_t),$$

(3.3b) $$\hat{w}_t=\left[1+\hat{u}_{t.}\hat{\Sigma}^{-1}\hat{u}_{t.}'/(v-2)\right]^{-1},\qquad \hat{u}_{t.}=y_{t.}-y_{t.}\hat{B}-z_{t.}\hat{C},$$

(3.3c) $$\hat{\Pi}=\hat{C}(I-\hat{B})^{-1},\qquad \operatorname{vec}\left[(\hat{B}',\hat{C}')'\right]=\operatorname{diag}_M(L_i)\hat{\delta}.$$

THEOREM 3.2: *Under the assumptions of Theorem* 3.1 *the* TFIML *estimators* $\hat{\delta}$ *and* $\hat{\Sigma}$ *are consistent*; *further* $\sqrt{T}(\hat{\delta}-\delta)\overset{\text{i.d.}}{\to}N[0,\Phi_{\hat{\delta}}(v,M)]$ *where*

(3.4) $$\Phi_{\hat{\delta}}(v,M)=\frac{v-2}{v}\,\frac{v+M+2}{v+M}\left[R'(\Sigma^{-1}\otimes Q)R\right]^{-1},$$
$$R=\operatorname{diag}_M(\Pi L_{i1},L_{i2}).$$

That is, the asymptotic distribution of $\hat{\delta}$ *is* $N[\delta,\Phi_{\hat{\delta}}(v,M)/T]$.[11]

REMARK 3.1: The normal equations of the full information maximum likelihood (NFIML) estimators, say $\tilde{\delta}$ and $\tilde{\Sigma}$, for normally distributed disturbances can be obtained by replacing $\hat{w}_t$ and $(M+v)/(v-2)$ by unity everywhere in (3.3).[12] This result corresponds to the fact that the multivariate normal distribution is a limiting case of the multivariate t as $v\to\infty$.

REMARK 3.2: Define the instrument matrix $\hat{P}=(\hat{\Sigma}^{-1}\otimes\hat{W})\hat{X}$, and note that the TFIML estimator can be expressed as $\hat{\delta}=(\hat{P}'X)^{-1}\hat{P}'y$. Hence the TFIML estimator can be viewed as an instrumental variable estimator of the stacked model (2.4). This generalizes a similar result of Hausman [**7**, **8**] obtained for the NFIML estimator.

[11] Note that in (3.4) the asymptotic covariance matrix of $\hat{\delta}$ involves the number of equations M; furthermore, M does not cancel even if Σ is diagonal. The reason for this is that if the disturbances are distributed multivariate t they are stochastically dependent even if Σ is diagonal. It may also appear that formula (3.4) does not involve moments of w_t. This is not the case. We have calculated those moments under density (3.1) explicitly as functions of v and Σ—compare also Corollary 4.1. A simple consistent estimator for $\Phi_{\hat{\delta}}(v,M)$ is

$$\frac{v-2}{v}\,\frac{v+M+2}{v+M}\left[T^{-1}\hat{X}'(\hat{\Sigma}^{-1}\otimes I_T)\hat{X}\right]^{-1}.$$

[12] Compare Hausman [**7**, **8**] and Hendry [**9**]. Note that Hendry's normal equations are written in implicit rather than explicit form.

REMARK 3.3: The asymptotic distribution of the NFIML estimator is $N[\delta, \Phi_{\tilde{\delta}}(M)/T]$ with $\Phi_{\tilde{\delta}}(M) = [R'(\Sigma^{-1} \otimes Q)R]^{-1}$ for all disturbance distributions satisfying the assumptions of Section 2.[13] Note that the multivariate t distribution specified above satisfies those conditions. Upon comparing the asymptotic covariance matrices of the TFIML and NFIML estimator we see that they are proportional:

$$\Phi_{\hat{\delta}}(v, M) = \kappa(v, M)\Phi_{\tilde{\delta}}(M) \qquad \text{with} \quad \kappa(v, M) = \frac{v-2}{v}\,\frac{v+M+2}{v+M}.$$

We note that for finite degrees of freedom, v, the proportionality factor is less than unity so that there is gain in efficiency with the use of the TFIML estimator. We also note that the proportionality factor decreases as the numbers of equations increases. Since $v > 5$, for large models a 40 per cent increase in efficiency constitutes the upper bound. However, note that the above formula for the asymptotic variance covariance matrix of the TFIML estimator and hence the gain factor is specific to the distributional assumption (3.1). For distributions other than the multivariate t the gain in efficiency may be significantly higher, as shown in the subsequent sections.

REMARK 3.4: Premultiplying the tth value of the endogenous and exogenous variables and the disturbances of our model with $w_{*t}^{1/2} = [((M + v)/(v - 2))w_t]^{1/2}$ where $w_t = [1 + u_{t.}\Sigma^{-1}u_{t.}'/(v - 2)]^{-1}$ yields the following transformed model

$$(3.5) \qquad y_* = X_*\delta + u_*$$

with $y_* = (I_M \otimes W_*^{1/2})y$, $X_* = (I_M \otimes W_*^{1/2})X$, $u_* = (I_M \otimes W_*^{1/2})u$, $W_*^{1/2} = \text{diag}_T(w_{*t}^{1/2})$. Observe that w_t is an even function of $u_{t.}$ and $0 \leq w_t \leq 1$. A typical element of the transformed disturbance vector is $u_{*tj} = [((M + v)/(v - 2))w_t]^{1/2}u_{tj}$. It is not difficult to show that the transformed disturbances have the same first two moments as the disturbances of the original model: $E(u_{*tj}) = 0$ and $E(u_{*ti}u_{*tj}) = \sigma_{ij}$.[14] However despite this fact, inspection of the structure of w_t shows that in the transformed model relatively less weight is given to observations in time periods with large disturbances (or outliers). This is particularly true for small values of the degrees of freedom parameter v, i.e. for distributions that have relatively fat tails as compared to the normal distribution. An observable form of the transformed model (3.5) is obtained by defining $w_{*t}^{1/2} = [((M + v)/(v - 2))\hat{w}_t]^{1/2}$ where $\hat{w}_t$ is based on the TFIML estimator defined in (3.3). We may now apply the NFIML formula to this model. It turns out (as can easily be checked) that the TFIML estimator of the original model and the NFIML estimator of the transformed model are identical.[15] In interpreting this result

[13] This well known result follows easily from Schönfeld's [**19**] central limit theorem, for example.
[14] See Kelejian and Prucha [**14**, p. 26].
[15] Similarly we can view the TFIML estimator as a minimum distance estimator of the transformed model (3.5) or as a generalized minimum distance estimator of the original model. Also, the TFIML estimator can be shown to be a member of the class of robust estimators suggested by Fair [**3**]; see Kelejian and Prucha [**14**] for details. For an interpretation of the single equation TFIML estimator as a member of Huber's [**10**] class of M estimators, see Poetscher and Prucha [**16**].

observe that the distribution of the transformed disturbances has relatively thinner tails than that of the original disturbances and that correspondingly the NFIML formula is based on a distribution with relatively thinner tails than that underlying the TFIML formula.

4. A CLASS OF ESTIMATORS INTERPRETABLE AS NUMERICAL APPROXIMATIONS TO TFIML

In the following we use the normal equations of the TFIML estimator to generate a wide class of estimators, say TFIML_A estimators, that can be interpreted as numerical approximations to the TFIML estimator. In particular the assumption of multivariate t distributed disturbances is no longer maintained; rather the analysis is performed under the general distributional assumptions of Section 2. As a consequence we have to redefine the degrees of freedom parameter v such that the class of TFIML_A estimators remains well specified even for general disturbance distributions. It seems desirable to define v in such a way that if the disturbance distribution is multivariate t the interpretation of v as a degrees of freedom parameter still holds.

For the case in which the disturbance distribution is multivariate t the ratio of the second to the squared first absolute moment is ($i = 1, \dots, M$)

$$(4.1) \qquad \sigma_{ii}/m_i^2 = g(v), \qquad g(v) = \left\{\pi\Gamma(v/2)^2\right\}/\left\{(v-2)\Gamma\left[(v-1)/2\right]^2\right\},$$

where $m_i = E[|u_{ti}|]$. Denote the average of these ratios as $a = M^{-1}\sum_{i=1}^{M}\sigma_{ii}/m_i^2$; then $a = g(v)$. The function $g(v)$ is strictly monotone decreasing with $g(v=2) = \infty$ and $g(v=\infty) = \pi/2$. This last value corresponds to the normal distribution. We now use the above relationship to define $\mu = v^{-1}$ for general disturbance distributions as the following function of first absolute and second moments:

$$(4.2) \qquad \mu = \begin{cases} 1/\left[g^{-1}(a)\right] & \text{if } a > \pi/2, \\ 0 & \text{if } a \leq \pi/2. \end{cases}$$

For all disturbance distributions with "fatter" tails than the normal $a > \pi/2$ and so $\mu = 1/v(a)$ where $v(a) = g^{-1}(a)$ is the solution value of $g(v) - a = 0$. Since $g(\cdot)$ is strictly monotone decreasing this value is unique.[16] For disturbance distributions with "thinner" tails than the normal $\mu = 0$.

The above definition of v is not the only possible one. In principle we could define v as a function of any two moments, or quantiles, or set v equal to some fixed number.[17] Our results below hold for any definition of v as long as its estimator has certain (reasonable) properties.

[16] Practically we may find the solution value of $g(v) - a = 0$ by pretabulating $g(v)$ for different values of v or by applying some of the readily available solution algorithms for implicit equations.

[17] An interesting generalization of this class could be obtained by allowing v to differ over equations. We conjecture that asymptotic results similar to the ones given below could also be derived for this more general class. However, this is a subject of future research.

4.1. *The Estimator Generating Equations: A Generalization*

We now use the normal equations of the TFIML estimator (3.3) as estimator generating equations in defining the following general class of estimators.

DEFINITION 4.1: Let $\check{\Pi}$, $\check{\Sigma}^{-1}$, $\check{\Sigma}_*^{-1}$, $\check{\delta}$, and $\check{\mu}$ be any estimators of Π, Σ^{-1}, Σ^{-1}, δ, and $\mu = v^{-1}$, respectively, where $\check{\Sigma}^{-1}$ and $\check{\Sigma}_*^{-1}$ need not be the same; then any instrumental variable estimator of the form (4.3) is said to be a TFIML_A estimator:

(4.3a) $$\hat{\delta}_A = \left[\check{X}'(\check{\Sigma}^{-1} \otimes \check{W})X\right]^{-1}\check{X}'(\check{\Sigma}^{-1} \otimes \check{W})y,$$

$$\check{X} = \text{diag}_M(\check{X}_i), \qquad \check{X}_i = \left[\check{Y}_i, Z_i\right], \qquad \check{Y}_i = Z\check{\Pi}_i, \qquad \check{\Pi}_i = \check{\Pi}L_{i1},$$

(4.3b) $$\check{W} = \text{diag}_T(\check{w}_t), \qquad \check{w}_t = \left\{1 + \left[\check{\mu}/(1 - 2\check{\mu})\right]\check{u}_{t.}\check{\Sigma}_*^{-1}\check{u}_{t.}'\right\}^{-1},$$

$$\check{u}_{t.} = y_{t.} - y_{t.}\check{B} - z_{t.}\check{C}, \qquad \text{vec}\left[(\check{B}', \check{C}')'\right] = \text{diag}_M(L_i)\check{\delta}.$$

For a given set of data the estimator $\hat{\delta}_A$ can be viewed as a function of $\check{\Pi}$, $\check{\Sigma}^{-1}$, $\check{\Sigma}_*^{-1}$, $\check{\delta}$, and $\check{\mu}$. We denote this dependence more compactly as

(4.3′) $$\hat{\delta}_A = F\left(\check{\Pi}, \check{\Sigma}^{-1}, \check{\delta}, \check{\Sigma}_*^{-1}, \check{\mu}\right).$$

REMARK 4.1: Note that we define TFIML_A estimators as a function of an estimator of $\mu = 1/v$. Hence a priori knowledge of v (as required in Section 3) is no longer assumed. Hendry's [9] NFIML_A class results as a special subclass of the TFIML_A class for $\check{W} = I_T$ or equivalently $\check{\Sigma}_*^{-1} = 0$: $\hat{\delta}_A = [\check{X}'(\check{\Sigma}^{-1} \otimes I_T) X]^{-1}\check{X}'(\check{\Sigma}^{-1} \otimes I_T)y = F(\check{\Pi}, \check{\Sigma}^{-1}, ., 0, .)$. Hendry showed that virtually all known estimators belong to the NFIML_A class; hence they also belong to the TFIML_A class. In addition the TFIML_A class contains generalizations of virtually all known estimators. As an illustration, the 3SLS estimator is, using obvious notation, given by $\delta_{3SLS} = F(\Pi_{OLS}, \Sigma_{2SLS}^{-1}, ., 0, .)$. The estimator $\delta_{G3SLS} = F(\Pi_{OLS}, \Sigma_{2SLS}^{-1}, \delta_{2SLS}, \Sigma_{2SLS}^{-1}, \mu_{2SLS})$, where μ_{2SLS} is the moment estimator of μ based on 2SLS residuals, would be a typical example for a generalized version of the 3SLS estimator.

4.2. *Consistency of* TFIML_A *Estimator*

We now give a general theorem concerning the consistency of TFIML_A estimators.[18]

THEOREM 4.1: *Consider the model* (2.1) *subject to the* (*general distribution*) *assumptions of Section* 2. *Let* $\check{\delta}_C$ *be any consistent estimator for* δ. *Let* $\check{\Pi}$, $\check{\Sigma}^{-1}$, $\check{\Sigma}_*^{-1}$

[18] Roughly speaking, in proving the theorem we first show that $\text{plim}(\hat{\delta}_C - \bar{\delta}_C) = 0$ where $\bar{\delta}_C = F(\text{plim}\,\check{\Pi}, \text{plim}\,\check{\Sigma}^{-1}, \delta, \text{plim}\,\check{\Sigma}_*^{-1}, \text{plim}\,\check{\mu})$. We then prove the consistency of $\bar{\delta}_C$. The detailed proof is available on request.

and $\check{\mu}$ *be any (possibly inconsistent) estimators for* Π, Σ^{-1} *and* μ, *respectively, with bounded probability limits and in particular* plim $\check{\mu} < 1/2$. *Let* $\check{\Sigma}_*^{-1}$ *be a positive semidefinite matrix and* plim $T^{-1}\check{X}'(\check{\Sigma}^{-1} \otimes I_T)X$ *a finite nonsingular matrix. Then the* TFIML_A *estimator* $\hat{\delta}_C = F(\check{\Pi}, \check{\Sigma}^{-1}, \hat{\delta}_C, \check{\Sigma}_*^{-1}, \check{\mu})$ *is consistent for* δ.

REMARK 4.2: The above theorem essentially makes case by case studies of the consistency of TFIML_A estimators superfluous. Roughly speaking, the theorem states that TFIML_A estimators (as instrumental variable estimators) are consistent even if inconsistent estimators are used in the construction of the instruments. There is one exception: The disturbances entering the weighting matrix $\check{W}$ must be estimated consistently. This ensures that plim $T^{-1}Z'\check{W}U = 0$. Since the variances are bounded, this implies that there exists some $\rho > 0$ such that $\mu \leq 1/2 - \rho$ (compare (4.1) and (4.2)). Hence any $\check{\mu}$ that is consistent satisfies the assumptions of the Theorem. A sufficient but not necessary condition for plim $T^{-1}\check{X}'(\check{\Sigma}^{-1} \otimes I_T)X$ to be finite and nonsingular is that $\check{\Pi}$ and $\check{\Sigma}$ are consistent (see, e.g., Schmidt [**18**, p. 205]).

4.3. *Asymptotic Distribution of Linearized* TFIML_A *Estimators*

Consider the subclass of TFIML_A estimators defined as fixpoints of the form

$$\hat{\delta}_F = F\big(\check{\Pi}, \check{\Sigma}^{-1}, \hat{\delta}_F, \check{\Sigma}_*^{-1}, \check{\mu}\big). \tag{4.4'}$$

Clearly the TFIML estimator (3.3) is a member of this class. We are particularly interested in linearized versions of certain members of this class. To derive those estimators we rewrite (4.4′), using the explicit notation of (4.3), as

$$G[\hat{\delta}_F] = \check{X}'(\check{\Sigma}^{-1} \otimes \tilde{W})\tilde{u} = 0, \qquad \tilde{u} = \text{vec}(\tilde{U}) = y - X\hat{\delta}_F, \tag{4.4}$$

$$\tilde{W} = \text{diag}_T(\tilde{w}_t), \qquad \tilde{w}_t = \big[1 - (\check{\mu}/(1 - 2\check{\mu}))\tilde{u}_{t.}\check{\Sigma}_*^{-1}\tilde{u}_{t.}'\big]^{-1}.$$

A linearized version of $\hat{\delta}_F$ can be defined in terms of the first step of a Gauss–Newton iteration scheme designed to find a solution to (4.4). In general, starting at $\check{\delta}$, such a linearized estimator is given by $\hat{\delta}_{FL} = \check{\delta} - [\partial G(\check{\delta})/\partial\delta]^{-1}G(\check{\delta})$. The following subclass of linearized estimators will be of particular interest.

DEFINITION 4.2: Consider the model of Section 2. Let $\check{\Pi}$, $\check{\Sigma}^{-1}$, $\check{\Sigma}_*^{-1}$, and $\check{\mu}$ be consistent estimators with $\check{\Sigma}_*^{-1}$ positive semidefinite; further, let the elements of $\check{\delta} - \delta$, $\check{\Sigma}_*^{-1} - \Sigma^{-1}$ and $\check{\mu} - \mu$ be of $O_p(T^{-1/2})$.[19] Then in the notation of (4.3) any

[19] These speed of convergence assumptions are not restrictive at all. Any of the usual estimators will satisfy these conditions. See Remark 4.4 for further details.

estimator of the form

$$\text{(4.5a)} \qquad \hat{\delta}_E = \check{\delta} + \left[\check{X}'(\check{\Sigma}^{-1} \otimes \check{W})X - \frac{2\check{\mu}}{1 - 2\check{\mu}} \check{X}'\check{V}X \right]^{-1} \check{X}'(\check{\Sigma}^{-1} \otimes \check{W})\check{u},$$

$$\text{(4.5b)} \qquad \check{V} = \left[\check{V}_{ij} \right]_{i,j=1,\ldots,M}, \qquad \check{V}_{ij} = \operatorname{diag}_{t=1}^{T}(\check{v}_{ij,t}), \qquad \check{v}_{ij,t} = \check{\sigma}^{i\cdot}\check{u}_{t.}'\check{w}_t^2\check{u}_{t.}\check{\sigma}^{\cdot j},$$

with $\check{u} = y - X\check{\delta}$ is called a linearized TFIML$_{\text{E}}$ estimator (the subscript E stands for efficient—the exact meaning of which will become clear below).

The inverse matrix on the right-hand side of (4.5a) is readily seen to be equal to $-[G(\check{\delta})/\partial\delta]^{-1}$. We now give a general theorem concerning the consistency and asymptotic distribution of linearized TFIML$_{\text{E}}$ estimators.[20]

THEOREM 4.2: *Consider the model* (2.1) *subject to the general distribution assumptions of Section* 2. *Assume further that the fifth moment of the disturbance distribution exists and that the matrix* $\Sigma - \lambda\Omega$ *is nonsingular where* $\Omega = E(u_{t.}'w_t^2u_{t.})$, $\lambda = 2\mu/[(1 - 2\mu)\theta]$, *and* $\theta = E(w_t)$. *Let* $\hat{\delta}_E$ *be any linearized* TFIML$_{\text{E}}$ *estimator*; *then* $\hat{\delta}_E$ *is consistent and* $\sqrt{T}(\hat{\delta}_E - \delta) \xrightarrow{\text{i.d.}} N(0, \Phi_{\hat{\delta}})$ *where*

$$\text{(4.6)} \qquad \Phi_{\hat{\delta}} = H^{-1}\left[R'(\Sigma^{-1}\Omega\Sigma^{-1} \otimes Q)R \right] H^{-1}/\theta^2,$$

$$H = R'\left[\Sigma^{-1}(\Sigma - \lambda\Omega)\Sigma^{-1} \otimes Q \right] R.$$

REMARK 4.3: According to the above theorem the whole class of linearized TFIML$_{\text{E}}$ estimators has the same asymptotic distribution. Hence, case by case discussions of the large sample distributions of those estimators are unnecessary. Note also that while it is possible to iterate on the formula (4.5) those iterations will not affect the large sample distribution and hence will not lead to more efficient estimators.

REMARK 4.4: For practical purposes we need, for general disturbance distributions, a consistent estimator of the variance covariance matrix $\Phi_{\hat{\delta}}$. In introducing such an estimator we shall for later reference also consider issues concerning the speed of convergence. Let $\check{U}$ be based on any consistent estimator of the structural parameters, say $\check{\delta}$, such that $\check{\delta} = \delta + O_p(T^{-1/2})$.[21] It is then readily seen that $\check{\Sigma} = T^{-1}\check{U}'\check{U} = \Sigma + O_p(T^{-1/2})$ and therefore $\check{\Sigma}^{-1} = \Sigma^{-1} + O_p(T^{-1/2})$.[22] Similarly we have $\check{m}_i = T^{-1}\Sigma_{t=1}^{T}|\check{u}_{ti}| = m_i + O_p(T^{-1/2})$ as an estimator for the first absolute moment. As a consequence $\check{a} = M^{-1}\Sigma_{i=1}^{M}\check{\sigma}_{ii}/\check{m}_i^2 = a + O_p(T^{-1/2})$. Further, because the Gamma function is continuously differ-

[20]The proof is tedious, and involves a second order Taylor series expansion of $\check{w}_t$ around the true parameter values. The second order terms are found to be asymptotically negligible. Details are available on request.

[21]Note that virtually all known estimators, as the 2SLS estimator, converge of $O_p(T^{-1/2})$.

[22]The latter and several of the subsequent results follow immediately from Lemma 5.1.4. and Corollary 5.1.6 in Fuller [4, p. 184 and p. 192].

entiable so is the function $\mu = \mu(a)$ defined in (4.2) at all possible argument values, except at the "normal" point $a = \pi/2$. At that point $\mu(a)$ is continuous with finite left and right hand first order derivatives. It therefore follows that $\check{\mu} = \mu(\check{a}) = \mu + O_p(T^{-1/2})$. Let $\check{W} = \text{diag}_T(\check{w}_t)$ based on $\check{\delta}$, $\check{\Sigma}^{-1}$, and $\check{\mu}$. As a byproduct of the proof of Theorem 4.2 it is readily seen via a Taylor expansion of $\check{w}_t$ around the true parameter values that $\check{\theta} = T^{-1}\sum_{t=1}^{T}\check{w}_t = \theta + o_p(T^{-1/4})$ and $\check{\Omega} = T^{-1}\check{U}'\check{W}\check{U} = \Omega + o_p(T^{-1/4})$; hence also $\check{\lambda} = 2\check{\mu}/[(1-2\check{\mu})\check{\theta}] = \lambda + o_p(T^{-1/4})$. Assume further that $\check{\Pi}$ is a consistent estimator. Then in the notation of (4.3) the following estimator is consistent for $\Phi_{\hat{\delta}}$:

$$(4.7)\qquad \hat{\Phi}_{\hat{\delta}} = \check{H}^{-1}\left\{T^{-1}\check{X}'\left[\check{\Sigma}^{-1}\check{\Omega}\check{\Sigma}^{-1}\otimes I_T\right]\check{X}\right\}\check{H}^{-1}/\check{\theta}^2,$$

$$\check{H} = T^{-1}\check{X}'\left[\check{\Sigma}^{-1}(\check{\Sigma}-\check{\lambda}\check{\Omega})\check{\Sigma}^{-1}\otimes I_T\right]\check{X}.$$

For practical purposes it is hence not necessary to evaluate $E(w_t)$ or $E(u'_{t.}w_t^2u_{t.})$ analytically for different disturbance distributions.

REMARK 4.5: The assumption that the matrix $\Sigma - \lambda\Omega = E(u'_{t.}(1-\lambda w_t)u_{t.})$ is nonsingular is satisifed for both the Normal and the Student t distribution.[23] We think that the assumption is satisfied for all symmetric unimodal distributions with differentiable density; however this remains to be shown for the systems case.[24] As a safety strategy we may use a sort of "pretest" estimator: Clearly $\Sigma - \lambda\Omega$ is nonsingular if $\Delta = \{\det(I - \lambda\Omega\Sigma^{-1})\}^{1/M}$ is not equal to zero. Consider the estimator $\check{\Delta} = \{\det(I - \check{\lambda}\check{\Omega}\check{\Sigma}^{-1})\}^{1/M}$ and suppose $\check{\Sigma}^{-1} = \Sigma^{-1} + o_p(T^{-1/4})$, $\check{\Omega} = \Omega + o_p(T^{-1/4})$, and $\check{\lambda} = \lambda + o_p(T^{-1/4})$.[25] Then also $\check{\Delta} = \Delta + o_p(T^{-1/4})$. Now the basic idea of the "pretest" estimator is to choose an interval for $\check{\Delta}$ around zero. If $\check{\Delta}$ falls into the interval we select the NFIML estimator $\tilde{\delta}$ (or any other asymptotically equivalent estimator); otherwise we select one of the linearized TFIML$_E$ estimators, say $\hat{\delta}_E$. Consider, e.g., the interval $[-\check{s}/\ln T, \check{s}/\ln T]$ with $\check{s} = s + o_p(1)$ and $s > 0$.[26] Observe that the limits of the interval converge of slower order to zero that $\check{\Delta}$ converges to the true value Δ. Consequently the probability limit of the indicator function

$$(4.8)\qquad D(\check{\Delta}) = \begin{cases} 1 & \text{if } |\check{\Delta}| > \check{s}/\ln T, \\ 0 & \text{otherwise,} \end{cases}$$

is one if $\Delta \neq 0$ and zero if $\Delta = 0$. Now suppose we define the "pretest" estimator, say $\hat{\hat{\delta}}$, as

$$(4.9)\qquad \hat{\hat{\delta}} = D(\check{\Delta})\hat{\delta}_E + \left(1 - D(\check{\Delta})\right)\tilde{\delta};$$

[23] In particular we have $\Sigma - \lambda\Omega = \rho\Sigma$ with $\rho = 1$ for the normal and $\rho = (v+M)/(v+M+2)$ for the Student t distribution. Compare Corollary 4.1.

[24] For a verification of this assertion in the single equation case, see Poetscher and Prucha [**16**].

[25] For the existence of such estimators, see Remark 4.4.

[26] The subsequent results do not depend on the particular choice of s as long as $s > 0$. One possibility would be to choose s to be some fraction of the asymptotic standard deviation of $\check{\Delta}$.

then clearly $\sqrt{T}(\hat{\hat{\delta}} - \delta) \overset{\text{i.d.}}{\rightarrow} N(0, \Phi_{\hat{\hat{\delta}}})$ where $\Phi_{\hat{\hat{\delta}}} = \Phi_{\hat{\delta}}$ if $\Delta \neq 0$ and $\Phi_{\hat{\hat{\delta}}} = \Phi_{\tilde{\delta}}$ if $\Delta = 0$. $\Phi_{\tilde{\delta}}$ denotes the asymptotic covariance matrix of the NFIML estimator given in Section 3. Suppose $\hat{\Phi}_{\hat{\delta}}$ and $\hat{\Phi}_{\tilde{\delta}}$ are consistent estimators for $\Phi_{\hat{\delta}}$ and $\Phi_{\tilde{\delta}}$ respectively; then $\hat{\Phi}_{\hat{\hat{\delta}}} = D(\check{\Delta})\hat{\Phi}_{\hat{\delta}} + (1 - D(\check{\Delta}))\hat{\Phi}_{\tilde{\delta}}$ is a consistent estimator for $\Phi_{\hat{\hat{\delta}}}$.

In case the disturbances are distributed multivariate t with $v = \mu^{-1}$ degrees of freedom it is readily seen that

$$E(w_t) = \frac{v}{v + M} \quad \text{and} \quad E\left(w_t^2 u_{ti} u_{tj}\right) = \frac{v - 2}{v + M} \frac{v}{v + M + 2} \sigma_{ij}.$$

In case the disturbances are multivariate normal we have $\mu = 0$ and consequently $E(w_t) = 1$ and $E(w_t^2 u_{ti} u_{tj}) = \sigma_{ij}$. This implies the following corollary.

COROLLARY 4.1: *Consider the model* (2.1) *subject to the assumptions of Theorem* 4.2. *Assume that the disturbance distribution is multivariate t with v degrees of freedom. Let $\hat{\delta}_E$ be any linearized* TFIML$_{\text{E}}$ *estimator. Then* $\sqrt{T}(\hat{\delta}_E - \delta) \overset{\text{i.d.}}{\rightarrow} N(0, \Phi_{\hat{\delta}})$ *where $\Phi_{\hat{\delta}}$ defined in* (4.6) *simplifies to*

$$\Phi_{\hat{\delta}} = \frac{v - 2}{v} \frac{v + M + 2}{v + M} \left[R'(\Sigma^{-1} \otimes Q) R \right]^{-1}.$$

If in particular the disturbance distribution is multivariate normal, then $\Phi_{\hat{\delta}} = [R'(\Sigma^{-1} \otimes Q)R]^{-1}$.

REMARK 4.6: The corollary implies that linearized TFIML$_{\text{E}}$ estimators have, in case of multivariate t distributed disturbances, the same asymptotic distribution as the TFIML estimator and in case of multivariate normal distributed disturbances the same as the NFIML estimator (compare Section 3 for the covariance matrices of the TFIML and NFIML estimator). Hence linearized TFIML$_{\text{E}}$ estimators are efficient in both cases. Asymptotically no penalty is imposed for not knowing v a priori. Clearly the NFIML estimator is inefficient in the multivariate t case.

4.4. *Gain in Efficiency*: *The Single Equation Case*

In the following we compare for the single equation model, $y = Z\delta + u$, the asymptotic covariance matrix of linearized TFIML$_{\text{E}}$ estimators to that of the NFIML (i.e. OLS) estimator. We consider a variety of disturbance distributions.

Let $\tilde{\delta}$ be the OLS estimator; then for all disturbance distributions satisfying the conditions of Section 2 $\sqrt{T}(\tilde{\delta} - \delta) \overset{\text{i.d.}}{\rightarrow} N(0, \Phi_{\tilde{\delta}})$ with $\Phi_{\tilde{\delta}} = \sigma^2 Q^{-1}$ (compare Remark 3.3). Let $\hat{\delta}_E$ be any linearized TFIML$_{\text{E}}$ estimator; then under the assumptions of Theorem 4.2 $\sqrt{T}(\hat{\delta}_E - \delta) \overset{\text{i.d.}}{\rightarrow} N(0, \Phi_{\hat{\delta}})$. The variance covariance matrix $\Phi_{\hat{\delta}}$ is defined in (4.6) and reduces to

$$(4.10) \qquad \Phi_{\hat{\delta}} = \kappa \Phi_{\tilde{\delta}}, \qquad \kappa = \left[E(w_t^2 u_t^2)/\sigma^2 \right] \Big/ \left[E(w_t) - \frac{2\mu}{1 - 2\mu} E(w_t^2 u_t^2)/\sigma^2 \right]^2.$$

The parameter μ is implicitly defined by (4.2) with $a = E(u_t^2)/[E(|u_t|)]^2$. Hence, for the single equation case, the asymptotic covariance matrix of linearized TFIML_E estimators is seen to be proportional to that of the OLS estimator. We shall refer to the proportionality factor κ as the efficiency parameter.

In computing the efficiency parameter κ for a particular disturbance distribution we first have to calculate $E(|u_t|)$ and $E(u_t^2)$. This implies a certain value for μ. This value is then used to calculate $E(w_t)$ and $E(w_t^2u_t^2)$. Rather than evaluate the above expectations analytically, we calculate them by numerical integration techniques. The accuracy of the numerical procedure was checked for the Student t distribution against the analytically implied value $\kappa = (v-2)(v+3)/(v(v+1))$. It was found that the numerical results were accurate up to the first five decimal places.

Table I gives the values of the efficiency parameter κ for the Student t, the Normal, the Laplace, and the Logistic distribution.[27] We also evaluate κ for the case in which those distributions are "contaminated" by an ϵ-fraction of a Normal distribution with relatively fat tails. Such distributions are typically considered in the literature on robust estimators; see, e.g., Huber [**10**]. The contamination may model the occurrence of outliers. Table II gives the efficiency

TABLE I

EFFICIENCY PARAMETER κ OF TFIML_E ESTIMATORS RELATIVE TO THE OLS ESTIMATOR.
Disturbance Density: $f(u) = (1-\epsilon)f_*(u \mid m) + \epsilon f_N(u \mid cm)$.

$f_*(u \mid m)$	$\epsilon = 0$	$\epsilon = 0.05$, $c = 2$	$\epsilon = 0.05$, $c = 4$	$\epsilon = 0.05$, $c = 10$	$\epsilon = 0.1$, $c = 2$	$\epsilon = 0.1$, $c = 4$	$\epsilon = 0.1$, $c = 10$
$f_T(u \mid v = 5, m)$	.800	.77	.54	.17	.75	.43	.10
$f_T(u \mid v = 10, m)$	.945	.90	.63	.19	.86	.49	.11
$f_T(u \mid v = 15, m)$	.975	.93	.65	.20	.89	.50	.12
$f_N(u \mid m)$	1	.96	.69	.21	.93	.53	.12
$f_{LAP}(u \mid m)$	.66	.65	.45	.14	.65	.37	.09
$f_{LOG}(u \mid m)$	.91	.87	.60	.19	.84	.47	.11

NOTE: $f_N(u \mid m)$, $f_{LAP}(u \mid m)$, $f_{LOG}(u \mid m)$, and $f_T(u \mid v, m)$ denote, respectively, the probability density functions of the Normal, Laplace, Logistic, and Student $t(v)$ distribution with zero mean and first absolute moment m. Note that in all cases the densities are completely characterized by specifying m. In using the first absolute moment as a measure of spread we follow an often used convention in the roubustness literature. In the actual calculation we set m equal to one. Note however that the above results are independent of the particular value of m.

TABLE II

EFFICIENCY PARAMETER κ OF TFIML_E ESTIMATORS RELATIVE TO THE OLS ESTIMATOR.

Disturbance Density: $f(u) \propto [(1+bu^2)^{(b+3)/2} + cu^2]^{-1}$, $b = v - 2$.

	$c = 1$	$c = 10$	$c = 100$	$c = 1000$
$v = 5$	.66	.26	.05	.01
$v = 10$	.81	.33	.08	.02
$v = 15$	.85	.35	.09	.03

[27] See Johnson and Kotz [**12**].

parameter κ for a "generalized" version of the Student t distribution (the latter distribution is obtained for $c = 0$). The results of Tables I and II are quite encouraging. We note that in case of normally distributed disturbances linearized TFIML_E estimators are asymptotically equivalent to the OLS estimator. In all other cases we observe gains in efficiency due to the use of linearized TFIML_E estimators. In a variety of cases those gains are very substantial.

5. CONCLUSIONS

A number of issues remain to be considered. First, the paper should be viewed as a pilot study in the analysis of a general class of simultaneous equation estimators that, as indicated by large sample results, are less sensitive to deviations from the usual normality assumption, and provide the potential for substantial gains in efficiency. The effects of different definitions of the "degrees of freedom" parameter (including different definitions for different equations) on the efficiency of linearized TFIML_E estimators needs further analysis. Monte Carlo studies may be particularly useful in that respect.

Second, we may try to reduce the present moment requirements by using more elaborate techniques of proof. Progress along these lines has already been made in the single equation case; see Poetcher and Prucha [**16**].

Third, we believe that the matrix $\Sigma - \lambda\Omega$ is nonsingular for all symmetric unimodal distributions which have a differentiable density. However this remains to be determined. If this does not hold, the concept of "pretest" estimators needs further consideration.

Finally, it would be interesting to derive classes of estimators from other families of multivariate distributions and compare their properties to those of TFIML_A estimators.

University of Maryland

Manuscript received September, 1982; revision received May, 1983.

REFERENCES

[1] AMEMIYA, T.: "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50(1982), 689–711.

[2] BARTELS, R.: "On the Use of Limit Theorem Arguments in Economic Statistics," *American Statistician*, 31(1977), 85–87.

[3] FAIR, R. C.: "On the Robust Estimation of Econometric Models," *Annals of Economic and Social Measurement*, 3(1974), 667–678.

[4] FULLER, W. A.: *Introduction to Statistical Time Series*. New York: Wiley, 1976.

[5] GALLANT, A. R.: "Three-Stage Least-Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," *Journal of Econometrics*, 5(1977), 71–88.

[6] GOLDFELD, S. M., AND R. E. QUANDT: "Econometric Modelling with Non-Normal Disturbances," *Journal of Econometrics*, 17(1981), 141–155.

[7] HAUSMAN, J. A.: "Full Information Instrumental Variable Estimation of Simultaneous Equation Systems," *Annals of Economic and Social Measurement*, 3/4(1974), 641–652.

[8] ———: "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models," *Econometrica*, 43(1975), 727–738.

[9] HENDRY, D. F.: "The Structure of Simultaneous Equation Estimators," *Journal of Econometrics*, 4(1976), 51–88.

[10] HUBER, P. J.: "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35(1964), 73–101.
[11] ———: *Robust Statistics*. New York: Wiley, 1981.
[12] JOHNSON, N. L., AND S. KOTZ: *Distributions in Statistics: Continuous Univariate Distributions*. New York: Wiley, 1970.
[13] JUDGE, G. G., W. E. GRIFFITHS, R. C. HILL, AND T. C. LEE: *The Theory and Practice of Econometrics*. New York: Wiley, 1980.
[14] KELEJIAN, H. H., AND I. R. PRUCHA: "The Structure of Simultaneous Equation Estimators: A Generalization Towards Robustness," Department of Economics, University of Maryland, Working Paper 26, 1980.
[15] MADDALA, G. S.: *Econometrics*. New York: McGraw-Hill, 1977.
[16] POETSCHER, B. M., AND I. R. PRUCHA: "A Class of Partially Adaptive One-Step M-Estimators," Institute for Econometrics, University of Technology Vienna, 1983.
[17] RAIFFA, H., AND R. SCHLAIFER: *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University, 1961.
[18] SCHMIDT, P.: *Econometrics*. New York: Marcel Dekker, 1976.
[19] SCHOENFELD, P.: "A Useful Central Limit Theorem for m-Dependent Random Variables," *Metrika*, 17(1971), 116–128.