Asymptotic probability concentrations and finite sample properties of modified LIML estimators for equations with more than two endogenous variables

Dennis Oberhelman\textsuperscript{a}, K. Rao Kadiyala\textsuperscript{b,∗}

\textsuperscript{a}Darla Moore School of Business, University of South Carolina, Columbia, SC 29208, USA
\textsuperscript{b}Krannert Graduate School of Management, Purdue University, 1310 Krannert West Lafayette, IN 47907, USA

Abstract

This paper investigates the distributional properties of a class of modified limited information maximum-likelihood (LIML) estimators. It is shown that the asymptotic distributions of these estimators are more concentrated than those of the modified LIML estimators suggested by Fuller. Additionally, the results of an extensive Monte Carlo investigation of the finite sample properties of the proposed estimators show that when the equation of interest has more than two endogenous variables, the LIML estimator is often highly inefficient so that substantial gains in precision are realized by using the modified estimators in place of the LIML estimator. © 2000 Elsevier Science S.A. All rights reserved.

\textit{JEL classification:} C14; C22; C24; C41; C5

\textit{Keywords:} Asymptotic mean-squared error; Asymptotic probability concentration; Modified LIML; Monte carlo; Simultaneous equations models; Small $\sigma$ expansion

∗Corresponding author. Tel.: 765-494-4476; fax: 765-496-1778.
E-mail address: kadiyala@mgmt.purdue.edu (K. Rao Kadiyala).
1. Introduction

This paper uses Monte Carlo simulation to investigate the finite sample properties of a modification of the limited information maximum-likelihood estimator (LIMLE). It differs from the bulk of previous Monte Carlo studies of $k$-class estimators in that it is not restricted to models containing just two endogenous variables and allows for the presence of lagged endogenous variables. Specifically, our study uses small $\sigma$ results on the sampling properties of the modified LIMLE together with some results discussed in Phillips (1980, 1983b) to design a series of Monte Carlo experiments that allow us to systematically investigate the finite sample properties of the LIML and modified LIML estimators when the equation of interest contains three endogenous variables. Among other things, this permits us to study the sensitivity of the distributions of these estimators to the degree of correlation in the concentration matrix – an issue which has received little attention in empirical studies concerning the distributions of $k$-class estimators.

The paper is divided into three main sections. In Section 2 we present the formal model and introduce modifications of the LIMLE suggested by Fuller (1977) and Kadiyala and Oberhelman (1992). Section 3 summarizes some important small $\sigma$ results concerning these estimators. Finally, Section 4 presents the results of the Monte Carlo study.

2. The model

Let

$$y = Y_1 \beta + X_1 \gamma + u_1,$$

be one of $G$ simultaneous equations, where $y$ is a $T \times 1$ vector of observations on an endogenous variable, $Y_1$ is a $T \times G_1$ matrix of observations on $G_1$ additional endogenous variables, and $X_1$ is a $T \times K_1$ matrix of observations on $K_1$ exogenous variables. $\gamma$ and $\beta$ are, respectively, $K_1 \times 1$ and $G_1 \times 1$ vectors of parameters to be estimated. Finally, $u_1$ is a $T \times 1$ vector of independently and identically normally distributed errors with common mean 0 and common variance $\sigma^2$.

It is assumed that Eq. (1) is part of a system of $G \geq G_1 + 1$ equations which are characterized by, in addition to the variables appearing in (1), a $T \times (G - G_1 - 1)$ matrix of observations on $G_2 = (G - G_1 - 1)$ endogenous variables and a $T \times (K - K_1)$ matrix of observations on $K_2 = K - K_1$ exogenous variables. Let the $T \times K$ matrix of observations on the exogenous variables be denoted by $X_\star$ and let the $T \times K_2$ matrix of observations on the exogenous variables excluded from (1) be denoted by $X_2$, so that $X_\star = [X_1 | X_2]$. 
We assume that \( X_\ast X_\ast / T \) converges to a positive-definite matrix as \( T \) becomes large and define \( Y = [y|Y_1] \).

The reduced form of the endogenous variables equals \( Y = X_\ast (\pi_1|\Pi_2) + (v_1|V_2) \). where \( \pi_1' = (\pi_{11}'|\pi_{21}') \) and \( (\Pi_2' = \Pi_{12}'|\Pi_{22}') \) are, respectively, \( 1 \times (K_1 + K_2) \) and \( G_1 \times (K_1 + K_2) \) matrices of reduced form coefficients. The rows of \( (v_1|V_2) \) are independently normally distributed errors with each row having mean 0 and nonsingular covariance matrix

\[
\Omega = \begin{pmatrix}
\omega_{11} & \omega_{12} \\
\omega_{12}' & \Omega_{22}
\end{pmatrix}
\]

The least-squares estimators of \( \Pi \) and \( \Omega \) are

\[
\Pi = (X_\ast X_\ast)^{-1}X_\ast Y
\]

and

\[
\hat{\Omega} = \frac{\hat{\mathcal{V}}' \hat{\mathcal{V}}}{T - K},
\]

where \( \hat{\mathcal{V}} = Y - X_\ast \hat{\Pi} \).

We make the usual assumptions concerning the identifiability and estimability of (1). In particular, we assume that \( X_\ast \) has full rank \( K \) and that \( L = K_2 - G_1 > 0 \). Then, the \( k \)-class estimator \( b(k) \) of \( (\beta'|y')' \) is given by

\[
b(k) = [Z'(I - kM)Z]^{-1}Z'(I - kM)y,
\]

where

\[
Z = [Y_1|X_1]
\]

and

\[
M = I - X_\ast (X_\ast X_\ast)^{-1}X_\ast.
\]

Many of the single equation estimators used in practice fall within the set of \( k \)-class estimators. For example, the two-stage least-squares estimator (TSLSE) has \( k = 1 \), the ordinary least-squares estimator (OLSE) has \( k = 0 \), and the limited information maximum-likelihood estimator (LIMLE) suggested by Anderson and Rubin (1949) has \( k = \lambda_0 \), where

\[
\lambda_0 = \min_{\beta} \frac{(y - Y_1 \beta)'M_1(y - Y_1 \beta)}{(y - Y_1 \beta)'M(y - Y_1 \beta)},
\]

and

\[
M_1 = I - X_1(X_1'X_1)^{-1}X_1.'
\]

If \( \sqrt{T(k - 1)} \) converges to zero as \( T \to \infty \), then it is well known that \( b(k) \) has a limiting distribution that is normal and that \( b(k) \) is asymptotically efficient.
Consequently, estimators satisfying this condition (like the TSLSE and LIMLE) are called best asymptotically normal (BAN) estimators. In finite samples the distributions of BAN estimators can differ significantly from one and another. For example, it has been demonstrated that the TSLSE is often severely biased whereas the LIMLE is close to being median unbiased. Although this result favors the LIMLE, it is mitigated by the fact that, because the LIMLE does not possess moments of any order, its distribution can be expected to have thicker tails than that of the TSLSE. Thus, in finite samples the distributions of the LIMLE and TSLSE differ significantly in location and spread, and the choice of which estimator is preferred often involves a tradeoff between these properties.

Various attempts have been made to improve upon the LIML and TSLS estimators. One particularly appealing approach is the modification to the LIMLE suggested by Fuller (1977) which ensures that the resulting estimator possesses finite moments in small samples. Fuller modified the LIMLE by defining \( k = \lambda_0 - \kappa/(T - K) \) and showed that under general conditions as long as \( \kappa > 0 \) the modified estimator possesses finite moments. He also showed that when \( \kappa = 1 \) the resulting \( k \)-class estimator is unbiased to \( O(T^{-1}) \) and demonstrated that setting \( \kappa = 4 \) yields an estimator whose mean-squared error (MSE) to \( O(T^{-2}) \) is uniformly smaller than that of any other estimator based on a smaller \( \kappa \). In a comprehensive paper on single-equation estimators, Anderson et al. (1986) use small \( \sigma \) asymptotics to show that the Fuller modification improves the LIML estimator in terms of asymptotic mean-squared error and asymptotic probability of concentration; and Rothenberg (1984) shows that the results proved by Fuller under the assumption of normality hold for any symmetric error distribution possessing higher order moments.

A paper by Kadiyala and Oberhelman (1992) considers a modification of the LIMLE in which \( k \) is a linear function of \( \lambda_0 \); and using small \( \sigma \) asymptotics, demonstrates that the Fuller estimator is inadmissible under the criterion of asymptotic MSE. In the following section we discuss the small \( \sigma \) properties of this modification. These properties provide guidance in developing the finite sample investigation of the modified LIML estimators presented in section three.

3. Small \( \sigma \) results and admissibility of modified LIML estimators

Below we consider the distributional properties of the \( k \)-class estimator, \( b(k) \), when

\[
k = \left( 1 - \frac{\kappa_1}{L + T - K} \right) \lambda_0 + \frac{\kappa_2}{T - K}.
\]
This is a convenient parameterization of the $k$-class estimator because it includes virtually all single-equation estimators of practical importance: The OLS, TSLS, LIML and Fuller estimators.

Kadiyala and Oberhelman (1992) show that $b(k)$ has finite moments if

$$\frac{\chi_1}{L + T - K} > \frac{\chi_2}{T - K}$$

and defining $e_k = b(k) - (\beta_1' / \gamma)'$ and $\Delta \chi = \chi_1 - \chi_2$, they prove the following lemmas concerning the bias and mean-squared error of $b(k)$:

**Lemma 3.1.** The bias of $b(k)$ to order $\sigma^2$ is given by

$$E(e_k) = (\Delta \chi - 1) \sigma^2 q Q + O(\sigma^3).$$

**Lemma 3.2.** The mean-squared error of $b(k)$ to order $\sigma^4$ is given by

$$E(e_k e'_k) = \sigma^2 Q + \sigma^4 \text{TR}(QC_2)Q + \sigma^4 (3 - 2\Delta \chi) \text{TR}(QC_1)Q$$

$$+ \sigma^4 \left[ \left( \frac{L + T - K + 2}{L + T - K} \right) (\Delta \chi)^2 - 6(\Delta \chi - 1) \right]$$

$$+ \frac{2\chi_2^2 L}{(L + T - K)(T - K)} QC_1 Q$$

$$+ \frac{\sigma^4}{(T - K - 2)} \left[ (L + T - K - 2)(L + 2) - 2(L + T - K - 2) \Delta \chi \right]$$

$$+ \frac{T - K - 2}{T - K} (\Delta \chi)^2 + \frac{4\Delta \chi L}{(T - K)} - \frac{4\chi_1 L}{(T - K)}$$

$$+ \frac{2\chi_2^2 L}{(L + T - K)(T - K)} QC_2 Q + O(\sigma^5),$$

where $q$, $Q$, $C_1$ and $C_2$ are as defined in Kadane (1971). That is,

$$Q = \left[ \begin{array}{cc} \Pi_2' X^*_X \Pi_2 & \Pi_2' X^*_X \Pi_1 \\ X_1 X^*_X \Pi_2 & X_1 X^*_X \Pi_1 \end{array} \right]^{-1},$$

$$q' = (1 / \sigma^2) (\omega_{12} - \beta_1' \Omega_{22} \beta_1'),$$

$$C_1 = qq'.$
Minimizing the determinantal value of the moment matrix $E(e_k e_k')$ given above with respect to $\chi_1$ and $\chi_2$, Kadiyala and Oberhelman (1992) find that the optimal values for $\chi_1$ and $\chi_2$ are:

$$\hat{\chi}_1 = -\frac{L + T - K - 2}{A} + L + T - K$$

and

$$\hat{\chi}_2 = \frac{L + T - K - 2 \left( \frac{\text{TR}(QC_2)}{\text{TR}(QC_1)} \right)}{(T - K - 2)A \left( \frac{\text{TR}(QC_2)}{\text{TR}(QC_1)} \right)}$$

where

$$A = \left[ \frac{L + T - K + 2}{L + T - K} + \frac{\text{TR}(QC_2)/\text{TR}(QC_1)}{T - K - 2} + \frac{(\text{TR}(QC_2)/\text{TR}(QC_1))^2}{(T - K)(T - K - 2)} \right].$$

and define an optimal estimator $b(k_*)$ with $k_* = (1 - \hat{\chi}_1/(L + T - K))\lambda_0 + \hat{\chi}_2/(T - K)$. Noting that $\hat{\chi}_1 \geq 4((L + T - K)/(L + T - K + 2))$ and $\hat{\chi}_2 \geq 0$, they point out that the TSLS, LIML, and Fuller estimators are not members of the family of estimators defined by $b(k_*)$ and recommend modifying the LIML by reducing the scale of the LIML root $\lambda_0$ rather than reducing the root through subtraction as suggested by Fuller. They refer to the proposed estimators as ‘scaled down’ LIML estimators and demonstrate that these estimators dominate the Fuller type estimators in terms of asymptotic mean-squared error (AMSE) to $O(\sigma^4)$.

Below, we state two theorems which formalize the results concerning the AMSE of the scaled down LIML estimators and demonstrate that these estimators also have desirable properties in terms of asymptotic probability of concentration (APC). Here APC is defined as $\text{APC} = \lim_{\sigma \to 0} 1/\sigma \text{Pr}(e_k \in S)$ and is obtained by integrating with respect to the asymptotic density of $e_k$ to $O(\sigma^3)$ over set $S$. Then, an estimator $b(k_1)$ is said to be preferred to estimator $b(k_2)$ if $\lim_{\sigma \to 0} 1/\sigma (\text{Pr}(e_k \in S) - \text{Pr}(e_k \in S)) \geq 0$ for every convex set $S$ that contains the origin. The proofs of the theorems are contained in an appendix to the paper.

**Theorem 3.1.** Let $k_1 = (1 - h/(L + T - K))\lambda_0$ and $k_2 = \lambda_0 - h/(T - K)$, where $h < 2(L + T - K)$, so that estimators $b(k_1)$ and $b(k_2)$ have the same asymptotic bias to $O(\sigma^2)$. Then, estimator $b(k_1)$ has lower AMSE to $O(\sigma^4)$ than estimator $b(k_2)$; and, for symmetric $S$, $\text{APC}(b(k_1)) \geq \text{APC}(b(k_2))$. 


Theorem 3.2. Suppose $T - K > 2$ and consider the scaled down LIML estimator $b(k_h)$ with $k_h = (1 - h/(L + T - K))j_0$ and $h < L + T - K$. If $h^*_s = (L + T - K)/(L + T - K + 2)$, then $b(k_h)$ with $h = h^*_s$ has greater APC for symmetric $S$ than any other estimator $b(k_h)$ with $h < h^*_s$, and estimator $b(k_h)$ with $h = 4h^*_s$ has lower AMSE to $O(\sigma^4)$ than any other estimator $b(k_h)$ with $h < 4h^*_s$.

Theorem 3.1 implies that the Fuller estimators are inadmissible because a scaled down LIML estimator will always have lower AMSE and higher APC than a Fuller estimator of comparable bias, and Theorem 3.2 implies that a scaled down LIML estimator with $h = 1$ (SD1E) has lower AMSE and higher APC than the LIML estimator.

Note from Lemma 3.1 that the SD1 estimator has the desirable property of being unbiased to $O(\sigma^2)$ and that for $h > 1$ the bias of the modified estimator increases with $h$. Thus, Theorem 3.2 states that biased estimators are admissible under the criterion of APC and that a biased estimator with $h \geq 4h^*_s$ is preferred in terms of AMSE. This means that the selection of the shrinkage parameter $h$ involves a tradeoff between bias (location) and variance (concentration).

In particular, if we minimize $E(e_k e'_k)$ to $O(\sigma^4)$ with respect to $h$, we find that the optimal value of $h$ is

$$\hat{h} = \frac{4 + ((L + T - K - 2)/(T - K - 2)) TR_{21}}{((L + T - K + 2)/(L + T - K)) + ((L + T - K - 2) TR_{21})/((L + T - K)(T - K - 2))},$$

and that $\hat{h}$ is an increasing function of $TR_{21} = TR(QC_2)/TR(QC_1)$ with $4h^*_s \leq \hat{h} < L + T - K$. This implies that for a sufficiently large value of $TR_{21}$ the increase in the bias of the modified estimator resulting from an increase in $h$ will be more than offset by a corresponding reduction in variance.

One situation that is of particular interest occurs when $TR_{21}$ is large enough so that $\hat{h}$ is of the order of $L$. In this situation the optimal scaled down LIML estimator resembles the TSLS estimator. Specifically, the scaled down LIMLE with $h = L$ (SDL) has the same bias to $O(\sigma^2)$ as the TSLSE, and the AMSE of the two estimators differ only in three terms. Inspecting Eq. (5) we see that when $h = L$ these three terms all approach zero as $T$ becomes large. This is not the case for the TSLS estimator ($\chi_1 = L + T - K$ and $\chi_2 = T - K$). Therefore, the SDL estimator dominates the TSLS in MSE as $T$ becomes large. A similar observation was made by Anderson et al. (1986). They demonstrated that the Fuller modified LIMLE, $b(k)$, with $\chi_1 = 0$ and $\chi_2 = -L$ has lower AMSE to $O(T^{-2})$ than the TSLSE (note Theorem 3.2 implies that the SDL is preferred over this Fuller estimator).

The above discussion concerning the TSLS and SDL estimators suggest that the classical problem of choosing between the LIML and TSLS estimators can be reformulated as choosing between scaled down LIML estimators with $h = 0$.
or $L$. Furthermore, the asymptotic results discussed above imply that a modified LIMLE based on an intermediate value of $h$ will often be preferred to one with $h = 0$ or $L$ and that the particular value selected for $h$ involves a tradeoff between bias and variance. We investigate this tradeoff below by performing a series of Monte Carlo simulations which compile the distribution of the scaled down LIML estimator for various selections of $h$ over a wide range of values for $\text{TR}_{21}$. In addition to providing a measure of bias/variance tradeoff associated with a particular value of $h$, the simulations also provide an indication of how important the modifications suggested by asymptotic theory (e.g. using the SD1E in place of the LIMLE) are in practical cases.

4. Empirical results concerning the distributions of the modified LIML estimators

In this section we summarize the results of an extensive Monte Carlo study of the distributions of the LIML and scaled down LIML estimators. The design of the study is influenced by the asymptotic results discussed above and by numerical results on the finite sampling properties of simultaneous equation estimators reported in papers by Phillips (1980,1983a) and Anderson et al. (1982).

The study by Anderson et al. (AKS) provides an analysis of the LIMLE for the structural coefficient $\beta$ when the model contains a single explanatory variable ($G_1 = 1$). Their tabulations show that the distribution of the LIMLE is somewhat thicker in the tail region than would be predicted by using a normal approximation; but, in general, they conclude that even though the moments of the LIML estimator are not finite the normal distribution is a fairly accurate approximation to the actual distribution of the LIMLE when $G_1 = 1$.

Phillips (1980,1983b) investigates the marginal distributions of instrumental variable estimators for the structural coefficients of models with $G_1 > 1$. He shows that the distributions are particularly sensitive to the degree of correlation in the concentration matrix; and finds that for comparable parameter values, the distributions of these estimators are less concentrated as $G_1$ increases and that the distributions appear to concentrate less quickly for large $T$ when the number of explanatory endogenous variables increases. He also notes that the differences in concentration observed between the cases $G_1 = 2$ and $1$ are more significant than those observed between $G_1 = 3$ and $2$.

The above observations imply that there is a greater potential for the distribution of the LIMLE to be poorly concentrated in models with $G_1 > 1$. This, taken together with observation by AKS that for $G_1 = 1$ the distribution of the LIMLE achieves a level of concentration that is comparable to the level predicted by the normal approximation, suggests that modifications of the LIMLE are more relevant in models with $G_1 > 1$. Therefore, to investigate this possibility, we consider a system of simultaneous equations whose first equation
contains two explanatory endogenous variables $Y_{11}$ and $Y_{12}$ so that

$$y = \beta_1 Y_{11} + \beta_2 Y_{12} + X_1 \gamma + \mu_1,$$  \hspace{1cm} (6)

and use Monte Carlo methods to generate the empirical distribution for $b_1(k)$, the estimator for the structural coefficient $\beta_1$, when $k = \lambda_0$ and $k = (1 - h/L + T - K)\lambda_0$. These distributions are used to quantify the improvement in concentration implied by the asymptotic results discussed above and to determine under what conditions the modifications are important in finite samples.

Following Phillips (1980) we specify, without loss of generality, that the model in Eq. (6) conforms to the standardizing transformations discussed in Basmann (1963,1974) and Phillips (1983a). The transformations reduce $T^{-1}(X'HX)^{-1}$ to an identity matrix of order $K$ and transform the variance–covariance matrix of the reduced form errors to an identity matrix. If the model conforms to these transformations, the correlation between $Y_i$ and $\mu_i$ equals $\beta_i/(1 + \beta'\beta)^2$ and the concentration matrix $\Phi = \Omega^{-1/2}_2 \Pi_2^2 M_1 X_2 \Pi_2$ is a symmetric matrix which can be written as

$$\Phi = \begin{pmatrix} \phi_{11} & \rho \sqrt{\phi_{11}\phi_{22}} \\ \rho \sqrt{\phi_{11}\phi_{22}} & \phi_{22} \end{pmatrix},$$

when $G_1 = 2$

Thus, the reduction of Eq. (6) to canonical form isolates a set of critical factors that influence the statistical properties of the LIML and modified LIML estimators. These are (1) the magnitudes of $\phi_{11}$ and $\phi_{22}$, (2) the length and direction of the standardized coefficient vector $\hat{\beta} = (\beta_1, \beta_2)'$, (3) the correlation parameter $\rho$ referred to by Phillips (1980), and (4) the degrees of freedom $T - K$ and the number of overidentification restrictions $L$. Below, we systematically investigate the sensitivity of the distribution of $b_1(k)$ to these factors.

The papers by Anderson et al. (1982), and Phillips (1980,1983b) provide guidelines for the values of the critical parameters. AKS tabulate the distribution of the LIMLE for values of the standardized structural coefficient in the range of 0–5 and values of the concentration parameter between 30 and 1000. Similar assumptions concerning the magnitudes of the structural coefficients and the magnitudes of the diagonal elements of the concentration matrix are made by Phillips (1980,1983b) when he investigates the marginal distributions of the instrumental variable estimator for models with $G_1 > 1$. In most experiments Phillips sets $|\rho| = 0.90$.

The parameter values selected by AKS are based in part on a study by Anderson et al. (1983) which used a number of classical econometric models with $G_1 = 1$ to determine typical values for the standardized structural coefficient and the noncentrality parameter. They found that a structural coefficient equal
to 1 and a concentration parameter of 100 could be classified as typical values of these parameters in empirical econometric studies. It is reasonable to assume that the standardized structural coefficients and concentration parameters will be similar in magnitude for models with $G_1 = 1$ and 2. Accordingly, we will view 1 as a baseline value for parameters $\beta_1$ and $\beta_2$ and treat values of $\phi_{11}$ and $\phi_{22}$ on the order of 100 as moderate or typical values for the concentration parameters.

In this paper we provide a summary of the general distributional properties of the $k$-class estimators when the parameters of the model in Eq. (6) assume these typical values. See Tables 1–3. The results reported in the tables are a representative subset of those reported in a more comprehensive study that follows the advice given by Mariano (1982) who recommends that Monte Carlo studies investigating the properties of $k$-class estimators employ models in canonical form and that simulations avoid the duplication of points in the canonical space while covering the canonical space as extensively as possible. In the extended version of our study, which is available from the authors upon request, we investigated how the critical parameter $\rho$ affected the distributions of the $k$-class estimators as it was allowed to vary from 0 to 0.99, how the values of the noncentrality parameters $\phi_{11}$ and $\phi_{22}$ affected the distributions as they were allowed to vary from 30 to 500, and how the distributions of the $k$-class estimators changed as the vector $\beta = (\beta_1, \beta_2)'$ assumed various lengths and orientations. This was done by using Monte Carlo methods to generate 30,000 independent samples of size $T$ for each specified combination of critical parameters and using these samples to estimate $\beta_1$ in Eq. (6) by the LIMLE, TSLE, and the scaled down LIML estimator with a shrinkage constant equal to $h(SdHE)$. These estimates were then used to generate normalized estimates for $\beta_1$ of the form $Z_k = (b_1(k) - \beta_1)/\text{ASD}$ where ASD represents the actual asymptotic standard deviation of $b_1(k)$ (i.e., the first diagonal element of matrix $\sigma^2Q$). We used the 30,000 replicated values of $Z_k$ to compile an empirical distribution for each of the normalized $k$-class estimators and used these distributions to compute selected percentile values of $Z_k$ and the relative frequencies with which the normalized estimators for $\beta_1$ fall within interval $A$: $[-0.32, 0.32]$, interval $B$: $[-0.67, 0.67]$, interval $C$: $[-1.96, 1.96]$, and interval $D$: $(-\infty, 0]$. In addition, we also computed the median value of the absolute errors $(b_1(k) - \beta_1)$ denoted in the tables by $|\text{MED}|$. The median absolute deviation of errors is computed because the LIMLE does not possess finite moments; and, as a consequence, the more traditional measure of mean absolute deviation does not exist for the LIMLE.

The median of the absolute errors together with the concentrations of probability listed under $A, B,$ and $C$ provide an indication of the overall efficiency of each of the individual estimators of $\beta_1$. Moreover, comparing the percentile values of $Z_k$ to the corresponding percentiles of the standard normal distribution, and comparing the concentrations of probability listed under $A, B, C,$ and
<table>
<thead>
<tr>
<th>Est</th>
<th>Phi</th>
<th>Beta</th>
<th>TR21</th>
<th>Z_{10}</th>
<th>Z_{25}</th>
<th>Z_{50}</th>
<th>Z_{75}</th>
<th>Z_{975}</th>
<th>MED</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>LML</td>
<td>100.00</td>
<td>1.00</td>
<td>29.00</td>
<td>-2.35</td>
<td>-1.02</td>
<td>2.46</td>
<td>7.85</td>
<td>0.40</td>
<td>0.18</td>
<td>0.36</td>
<td>0.74</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>SD1</td>
<td>100.00</td>
<td>1.00</td>
<td></td>
<td>-2.18</td>
<td>-0.86</td>
<td>1.83</td>
<td>3.18</td>
<td>0.35</td>
<td>0.20</td>
<td>0.40</td>
<td>0.83</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>SD2</td>
<td>-2.50</td>
<td>-1.56</td>
<td>-0.80</td>
<td>-0.03</td>
<td>0.74</td>
<td>1.52</td>
<td>2.44</td>
<td>0.31</td>
<td>0.22</td>
<td>0.45</td>
<td>0.89</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>SD4</td>
<td>-1.96</td>
<td>-1.27</td>
<td>-0.69</td>
<td>-0.06</td>
<td>0.58</td>
<td>1.16</td>
<td>1.84</td>
<td>0.25</td>
<td>0.27</td>
<td>0.53</td>
<td>0.96</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td>(DF = 30)</td>
<td>-1.47</td>
<td>-1.00</td>
<td>-0.58</td>
<td>-0.12</td>
<td>0.33</td>
<td>0.74</td>
<td>1.21</td>
<td>0.18</td>
<td>0.67</td>
<td>0.99</td>
<td>0.57</td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-1.60</td>
<td>-1.06</td>
<td>-0.60</td>
<td>-0.11</td>
<td>0.37</td>
<td>0.85</td>
<td>1.40</td>
<td>0.20</td>
<td>0.34</td>
<td>0.64</td>
<td>0.98</td>
</tr>
<tr>
<td>LML</td>
<td>100.00</td>
<td>1.41</td>
<td>2.00</td>
<td>-2.78</td>
<td>-1.24</td>
<td>1.16</td>
<td>2.24</td>
<td>0.40</td>
<td>0.20</td>
<td>0.36</td>
<td>0.79</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>SD1</td>
<td>100.00</td>
<td>0.00</td>
<td></td>
<td>-2.09</td>
<td>-0.54</td>
<td>1.26</td>
<td>1.86</td>
<td>0.31</td>
<td>0.22</td>
<td>0.44</td>
<td>0.88</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>SD2</td>
<td>-2.36</td>
<td>-1.65</td>
<td>-1.02</td>
<td>-0.41</td>
<td>0.23</td>
<td>0.87</td>
<td>1.67</td>
<td>0.28</td>
<td>0.24</td>
<td>0.48</td>
<td>0.92</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>SD4</td>
<td>-2.35</td>
<td>-1.66</td>
<td>-1.15</td>
<td>-0.65</td>
<td>-0.15</td>
<td>0.32</td>
<td>0.88</td>
<td>0.28</td>
<td>0.23</td>
<td>0.47</td>
<td>0.94</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>(DF = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-2.20</td>
<td>-1.73</td>
<td>-1.37</td>
<td>-0.95</td>
<td>-0.51</td>
<td>-0.07</td>
<td>0.50</td>
<td>0.38</td>
<td>0.13</td>
<td>0.31</td>
<td>0.95</td>
</tr>
<tr>
<td>LML</td>
<td>100.00</td>
<td>1.00</td>
<td>0.58</td>
<td>-1.81</td>
<td>-1.21</td>
<td>-0.72</td>
<td>-0.04</td>
<td>0.95</td>
<td>2.59</td>
<td>0.32</td>
<td>0.21</td>
<td>0.43</td>
<td>0.84</td>
</tr>
<tr>
<td>SD1</td>
<td>100.00</td>
<td>-1.00</td>
<td></td>
<td>-1.88</td>
<td>-1.30</td>
<td>-0.86</td>
<td>-0.30</td>
<td>0.39</td>
<td>1.17</td>
<td>2.14</td>
<td>0.27</td>
<td>0.49</td>
<td>0.95</td>
</tr>
<tr>
<td>SD2</td>
<td>-1.93</td>
<td>-1.39</td>
<td>-0.98</td>
<td>-0.51</td>
<td>0.02</td>
<td>0.54</td>
<td>1.19</td>
<td>0.26</td>
<td>0.25</td>
<td>0.51</td>
<td>0.97</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>SD4</td>
<td>-2.03</td>
<td>-1.55</td>
<td>-1.21</td>
<td>-0.84</td>
<td>-0.44</td>
<td>-0.07</td>
<td>0.38</td>
<td>0.33</td>
<td>0.16</td>
<td>0.38</td>
<td>0.97</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>(DF = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-2.24</td>
<td>-1.88</td>
<td>-1.58</td>
<td>-1.24</td>
<td>-0.86</td>
<td>-0.47</td>
<td>0.04</td>
<td>0.49</td>
<td>0.06</td>
<td>0.16</td>
<td>0.93</td>
</tr>
<tr>
<td>LML</td>
<td>100.00</td>
<td>1.00</td>
<td>29.00</td>
<td>-5.93</td>
<td>-2.15</td>
<td>-0.95</td>
<td>-0.01</td>
<td>0.96</td>
<td>2.25</td>
<td>0.38</td>
<td>0.19</td>
<td>0.38</td>
<td>0.77</td>
</tr>
<tr>
<td>SD1</td>
<td>100.00</td>
<td>1.00</td>
<td></td>
<td>-3.08</td>
<td>-1.76</td>
<td>-0.85</td>
<td>-0.02</td>
<td>0.83</td>
<td>1.78</td>
<td>0.39</td>
<td>0.33</td>
<td>0.21</td>
<td>0.42</td>
</tr>
<tr>
<td>SD2</td>
<td>-2.48</td>
<td>-1.53</td>
<td>-0.78</td>
<td>-0.03</td>
<td>0.73</td>
<td>1.50</td>
<td>2.44</td>
<td>0.30</td>
<td>0.23</td>
<td>0.45</td>
<td>0.89</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>SD4</td>
<td>-1.93</td>
<td>-1.28</td>
<td>-0.69</td>
<td>-0.06</td>
<td>0.58</td>
<td>1.18</td>
<td>1.83</td>
<td>0.25</td>
<td>0.27</td>
<td>0.53</td>
<td>0.96</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>(DF = 300)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-1.61</td>
<td>-1.06</td>
<td>-0.60</td>
<td>-0.12</td>
<td>0.37</td>
<td>0.85</td>
<td>1.43</td>
<td>0.20</td>
<td>0.34</td>
<td>0.64</td>
<td>0.98</td>
</tr>
<tr>
<td>LML</td>
<td>100.00</td>
<td>1.00</td>
<td>0.58</td>
<td>-1.67</td>
<td>-1.16</td>
<td>-0.69</td>
<td>-0.00</td>
<td>0.95</td>
<td>2.44</td>
<td>0.31</td>
<td>0.22</td>
<td>0.44</td>
<td>0.86</td>
</tr>
<tr>
<td>SD1</td>
<td>100.00</td>
<td>-1.00</td>
<td></td>
<td>-1.74</td>
<td>-1.26</td>
<td>-0.83</td>
<td>-0.26</td>
<td>0.42</td>
<td>1.14</td>
<td>2.09</td>
<td>0.27</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>SD2</td>
<td>-1.81</td>
<td>-1.35</td>
<td>-0.96</td>
<td>-0.48</td>
<td>0.03</td>
<td>0.53</td>
<td>1.13</td>
<td>0.25</td>
<td>0.27</td>
<td>0.53</td>
<td>0.98</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>SD4</td>
<td>-1.94</td>
<td>-1.52</td>
<td>-1.18</td>
<td>-0.81</td>
<td>-0.44</td>
<td>-0.10</td>
<td>0.30</td>
<td>0.32</td>
<td>0.16</td>
<td>0.40</td>
<td>0.98</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>(DF = 300)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-2.23</td>
<td>-1.88</td>
<td>-1.58</td>
<td>-1.23</td>
<td>-0.84</td>
<td>-0.47</td>
<td>0.03</td>
<td>0.49</td>
<td>0.06</td>
<td>0.16</td>
<td>0.92</td>
</tr>
<tr>
<td>NML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1

*(L = 9)*

Table 2

$L = 9, DF = 30$

<table>
<thead>
<tr>
<th>Est</th>
<th>Beta</th>
<th>TR21</th>
<th>Z_{0.25}</th>
<th>Z_{0.50}</th>
<th>Z_{0.75}</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Z_{0.25}</th>
<th>Z_{0.50}</th>
<th>Z_{0.75}</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Z_{0.25}</th>
<th>Z_{0.50}</th>
<th>Z_{0.75}</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>LML</td>
<td>1.00</td>
<td>29.00</td>
<td>-11.49</td>
<td>0.00</td>
<td>11.81</td>
<td>0.17</td>
<td>0.34</td>
<td>0.69</td>
<td>0.50</td>
<td>-3.07</td>
<td>0.01</td>
<td>3.20</td>
<td>0.21</td>
<td>0.41</td>
<td>0.85</td>
<td>0.50</td>
<td>-2.33</td>
<td>0.00</td>
<td>2.38</td>
<td>0.23</td>
<td>0.45</td>
<td>0.91</td>
<td>0.50</td>
</tr>
<tr>
<td>SDI</td>
<td>1.00</td>
<td></td>
<td>-2.84</td>
<td>-0.02</td>
<td>2.76</td>
<td>0.21</td>
<td>0.41</td>
<td>0.85</td>
<td>0.51</td>
<td>-2.68</td>
<td>0.00</td>
<td>2.79</td>
<td>0.22</td>
<td>0.43</td>
<td>0.87</td>
<td>0.50</td>
<td>-2.26</td>
<td>0.00</td>
<td>2.29</td>
<td>0.23</td>
<td>0.46</td>
<td>0.91</td>
<td>0.50</td>
</tr>
<tr>
<td>SD2</td>
<td></td>
<td></td>
<td>-2.16</td>
<td>-0.05</td>
<td>2.02</td>
<td>0.24</td>
<td>0.48</td>
<td>0.93</td>
<td>0.52</td>
<td>-2.45</td>
<td>0.00</td>
<td>2.51</td>
<td>0.23</td>
<td>0.45</td>
<td>0.89</td>
<td>0.50</td>
<td>-2.19</td>
<td>0.00</td>
<td>2.20</td>
<td>0.24</td>
<td>0.47</td>
<td>0.92</td>
<td>0.50</td>
</tr>
<tr>
<td>SD4</td>
<td></td>
<td></td>
<td>-1.63</td>
<td>-0.09</td>
<td>1.41</td>
<td>0.32</td>
<td>0.61</td>
<td>0.99</td>
<td>0.54</td>
<td>-2.15</td>
<td>0.00</td>
<td>2.14</td>
<td>0.25</td>
<td>0.49</td>
<td>0.93</td>
<td>0.51</td>
<td>-2.07</td>
<td>0.00</td>
<td>2.06</td>
<td>0.25</td>
<td>0.49</td>
<td>0.94</td>
<td>0.51</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>-1.21</td>
<td>-0.17</td>
<td>0.86</td>
<td>0.44</td>
<td>0.77</td>
<td>0.99</td>
<td>0.63</td>
<td>-1.77</td>
<td>0.00</td>
<td>1.66</td>
<td>0.29</td>
<td>0.57</td>
<td>0.97</td>
<td>0.53</td>
<td>-1.89</td>
<td>0.00</td>
<td>1.79</td>
<td>0.27</td>
<td>0.53</td>
<td>0.96</td>
<td>0.52</td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-1.40</td>
<td>-0.16</td>
<td>1.15</td>
<td>0.40</td>
<td>0.71</td>
<td>0.99</td>
<td>0.61</td>
<td>-1.79</td>
<td>0.00</td>
<td>1.71</td>
<td>0.29</td>
<td>0.57</td>
<td>0.96</td>
<td>0.53</td>
<td>-1.90</td>
<td>0.00</td>
<td>1.82</td>
<td>0.27</td>
<td>0.53</td>
<td>0.96</td>
<td>0.52</td>
</tr>
<tr>
<td>LML</td>
<td>1.00</td>
<td>0.58</td>
<td>-10.86</td>
<td>-0.20</td>
<td>11.51</td>
<td>0.19</td>
<td>0.39</td>
<td>0.78</td>
<td>0.57</td>
<td>-1.63</td>
<td>0.00</td>
<td>3.35</td>
<td>0.23</td>
<td>0.47</td>
<td>0.90</td>
<td>0.50</td>
<td>-1.68</td>
<td>0.01</td>
<td>2.68</td>
<td>0.25</td>
<td>0.49</td>
<td>0.93</td>
<td>0.50</td>
</tr>
<tr>
<td>SDI</td>
<td>-1.00</td>
<td></td>
<td>-3.09</td>
<td>-0.51</td>
<td>1.26</td>
<td>0.24</td>
<td>0.50</td>
<td>0.92</td>
<td>0.74</td>
<td>-1.69</td>
<td>0.00</td>
<td>2.60</td>
<td>0.24</td>
<td>0.49</td>
<td>0.94</td>
<td>0.56</td>
<td>-1.73</td>
<td>0.00</td>
<td>2.37</td>
<td>0.25</td>
<td>0.49</td>
<td>0.94</td>
<td>0.55</td>
</tr>
<tr>
<td>SD2</td>
<td></td>
<td></td>
<td>-2.60</td>
<td>-0.70</td>
<td>0.55</td>
<td>0.21</td>
<td>0.46</td>
<td>0.94</td>
<td>0.87</td>
<td>-1.76</td>
<td>0.00</td>
<td>2.02</td>
<td>0.24</td>
<td>0.50</td>
<td>0.96</td>
<td>0.63</td>
<td>-1.79</td>
<td>0.00</td>
<td>2.12</td>
<td>0.25</td>
<td>0.50</td>
<td>0.96</td>
<td>0.59</td>
</tr>
<tr>
<td>SD4</td>
<td></td>
<td></td>
<td>-2.26</td>
<td>-0.94</td>
<td>-0.02</td>
<td>0.08</td>
<td>0.27</td>
<td>0.95</td>
<td>0.98</td>
<td>-1.87</td>
<td>0.00</td>
<td>1.23</td>
<td>0.23</td>
<td>0.47</td>
<td>0.98</td>
<td>0.76</td>
<td>-1.89</td>
<td>0.00</td>
<td>1.68</td>
<td>0.24</td>
<td>0.49</td>
<td>0.97</td>
<td>0.68</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>-2.04</td>
<td>-1.23</td>
<td>-0.54</td>
<td>0.01</td>
<td>0.05</td>
<td>0.96</td>
<td>0.99</td>
<td>-2.14</td>
<td>0.10</td>
<td>1.38</td>
<td>0.09</td>
<td>0.23</td>
<td>0.95</td>
<td>0.96</td>
<td>-2.21</td>
<td>0.08</td>
<td>0.79</td>
<td>0.16</td>
<td>0.36</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>TSL</td>
<td></td>
<td></td>
<td>-2.05</td>
<td>-1.18</td>
<td>-0.21</td>
<td>0.03</td>
<td>0.13</td>
<td>0.96</td>
<td>0.99</td>
<td>-2.25</td>
<td>1.00</td>
<td>0.60</td>
<td>0.12</td>
<td>0.27</td>
<td>0.92</td>
<td>0.91</td>
<td>-2.23</td>
<td>1.00</td>
<td>1.01</td>
<td>0.16</td>
<td>0.36</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>NML</td>
<td></td>
<td></td>
<td>-1.96</td>
<td>0.00</td>
<td>-1.96</td>
<td>0.25</td>
<td>0.50</td>
<td>0.95</td>
<td>0.50</td>
<td>-1.96</td>
<td>0.00</td>
<td>1.96</td>
<td>0.25</td>
<td>0.50</td>
<td>0.95</td>
<td>0.50</td>
<td>-1.96</td>
<td>0.00</td>
<td>1.96</td>
<td>0.25</td>
<td>0.50</td>
<td>0.95</td>
<td>0.50</td>
</tr>
<tr>
<td>Df</td>
<td>Phi</td>
<td>F Tests</td>
<td>Z 0.025</td>
<td>Z 0.01</td>
<td>Z 0.005</td>
<td>MED</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>----------</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.00</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1.00</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3**

*DF = 30*

**Notes:**
- *D* indicates degree of freedom.
- *Phi* denotes phi correlation coefficient.
- *F Tests* represent Fisher's exact test.
- *Z* values are z-scores for tests of significance.
- *MED* stands for median.
- *A*, *B*, *C*, and *D* represent different categories.

**References:**
D to the normal values 0.25, 0.50, 0.95, and 0.50 provides an indication of the accuracy of the large sample approximation for the distribution of $Z_k$.

Upon investigating the results on concentration of probability and mean absolute deviation generated by the Monte Carlo experiments described above, we found for every combination of critical parameters considered that the scaled down LIML estimator with $h = 1$ was more efficient than the LIML estimator. For low values of $|\rho|$ the observed differences in efficiency were relatively minor, and the distribution of the normalized scaled down estimator for $\beta_1$ using $h = 1(Z_{SD1})$ was only marginally different from that of the normalized LIML estimator ($Z_{LML}$).

In general, when $|\rho|$ was small the properties of $Z_{LML}$ were similar to the properties observed by AKS for models with $G_1 = 1$. In particular, if $\rho = 0$ and $\beta_1$ and $\Phi_{11}$ assumed values comparable to those used for $\beta$ and the noncentrality parameter in AKS, then the percentile values in our study were nearly identical to those recorded by AKS. However, as $|\rho|$ increased in value we observed that the distribution of the normalized LIML estimator became less concentrated and that for parameter combinations consisting of low to moderate values of $\Phi_{11}$ and $\Phi_{22}$ and larger values of $|\rho|$, the distribution of $Z_{LML}$ tended to be much wider than the standard normal distribution. Moreover, the differences between the properties of $Z_{LML}$ and $Z_{SD1}$ were often substantial for larger values of $|\rho|$. Accordingly, the results summarized in Tables 1–3 concentrate on cases corresponding to larger values of $|\rho|$ by providing illustrations of the level of differences that exist between the distributions of $Z_{LML}$ and $Z_{SD1}$ when $|\rho| = 0.90$. Results are also presented in these tables for the normalized scaled down LIML with $h = 2, 4, 9$ and $L$ under the labels $Z_{SD2}$, $Z_{SD4}$, $Z_{SD9}$, and $Z_{SDL}$.

For fixed values of $|\rho|$, $\Phi_{11}$, and $\Phi_{22}$ the distribution of the normalized scaled down estimators with scaling parameter $h(Z_{SDh})$ are sensitive to changes in the length and orientation of $\beta$ and to the ratio of $\Phi_{11}|\Phi_{22}$. In our simulation experiments we found that the manner in which the distribution of $Z_{SDh}$ changed due to a change in $\beta$ or $\Phi_{11}|\Phi_{22}$ could generally be classified by the magnitude of parameter $TR_{21}$. Given this observation, we summarize the primary results of the Monte Carlo simulation described above by providing data on the results corresponding to high, moderate, and low values of $TR_{21}$.

This is done in Tables 1–3 for various values of $\Phi_{11}$, $\Phi_{22}$, $L$, and $T - K$. For instance, in Table 1 we provide distributional results for cases where $L = 9$, $\rho = 0.90$, $\Phi_{11} = \Phi_{22} = 100$, and $\beta = (1, 1)'$, $(\sqrt{2}, 0)'$, and $(1, -1)'$. These selections of ‘BETA’ and ‘PHI’ result in a value of ASD = 0.40 and values of $TR_{21}$ of 29, 2, and 0.58, respectively. Thus, the table shows how a change in $TR_{21}$ impacts the distribution of the $k$-class estimator, $b_1(k)$. In general, as demonstrated by the results in Table 1, we found that the distributions of each of the normalized estimators considered became increasingly asymmetric as $TR_{21}$
declined in value and that the bias of the estimator tended to increase with decreasing values of $TR_{21}$.

It is also apparent from Table 1 that as the difference between the scaling parameters of the scaled down LIML estimators becomes larger, the discrepancies observed between the individual distributions of the estimators become more pronounced. In particular, there are often dramatic differences observed between the distributions of the normalized LIML estimator ($h = 0$) and the normalized SDL estimator ($h = L$).

Since the distribution of each of the normalized estimators converges to the $N(0, 1)$ distribution, the differences between the individual distributions of the estimators dissipate as the concentration parameters increase in magnitude. Table 2 illustrates the rate of convergence of the estimators for values of $\phi_{11} = \phi_{22} = 50, 250, \text{ and } 500$. For given values of $\phi_{11}$ and $\phi_{22}$, the distributions of the estimators are relatively insensitive to the size of $T − K$ (see the bottom of Table 1). Thus, depending on the size of the concentration parameters, significant differences in the distributions may exist even for larger values of $T − K$.

The results in Table 3 provide examples of how the shape of the distributions of the normalized estimators change for various degrees of overidentification. In general, we observed that the width of the distributions of $Z_k$ tended to increase with $L$ while for a fixed value of $h$ the bias of the scaled down LIML estimators appeared to remain relatively stable as $L$ changed in value.

Taken as a whole, Tables 1–3 present an overview of the tradeoff between spread and location that one faces when considering a single equation estimator for $\beta$. The results in the tables point out that LIML and SDL estimators are at the two opposite extremes of this tradeoff. For instance, when the value of the shrinkage parameter $h$ is on the order of $L$, we observe that the scaled down LIMLE can be quite biased. Specifically, the results in Tables 1–3 show that as $TR_{21}$ declines the median of $Z_{SDL}$ moves away from zero, and the variance of the estimator declines. Hence, as $TR_{21}$ decreases the distribution of the estimator becomes concentrated about an aberrant value. Consequently, the approach of $Z_{SDL}$ to the $N(0, 1)$ distribution can be quite slow for smaller values of $TR_{21}$. Moreover, since the bias of the SDL estimator is proportionate to the number of overidentification restrictions (see Lemma 2.1), the poor sampling properties associated with this estimator are observed over a wider range of values for $TR_{21}$ as the number of overidentification restrictions increases. As a result it is beneficial to employ the SDL estimator in the parameter space corresponding to larger values of $TR_{21}$. Within this part of the parameter space the bias of the SDL estimator tends to be small, and the distribution of $Z_{SDL}$ is more concentrated than the $N(0, 1)$ distribution so that the SDL is more efficient than what is expected under the normal approximation.

The above observations about the normalized SDL estimator concur with those made by Anderson and Sawa (1979) concerning the normalized TSLS
estimator $Z_{TSL}$ for models with $G_1 = 1$. Comparing the distributions of $Z_{SDL}$ and $Z_{TSL}$, we found that $Z_{SDL}$ and $Z_{TSL}$ were generally quite similar in bias and that the percentile values of the distributions of the estimators were comparable for higher values of $TR_{21}$, $\Phi_{11}$ and $\Phi_{22}$. When $TR_{21}$ was low the distribution of $Z_{TSL}$ tended to be somewhat wider than that of $Z_{SDL}$. Overall, the median absolute deviations of the SDL and TSLS estimators were roughly equal. In general, it was found that choosing a scaled down LIML estimator with $h < L$ yields an estimator with better distributional properties than the TSLSE when the value of $TR_{21}$ is moderate to low; while for larger values of $TR_{21}$, the distributional properties of $Z_{SDL}$ and $Z_{TSL}$ are essentially the same. Accordingly, the traditional problem of choosing between the LIML and TSLS estimators can be reformulated as one of choosing among scaled down LIML estimator with shrinkage parameters in the range of 0–L.

In discussing how to choose between the TSLS and LIML estimators, Anderson et al. (1982) concluded that the LIMLE should be strongly preferred to the TSLS because of the unattractive properties observed above for the TSLS estimator and because the normal distribution provides an adequate approximation of the distribution of $Z_{LML}$ when $G_1 = 1$. However, for models with $G_1 > 1$ our results indicate that the distribution of $Z_{LML}$ becomes increasingly wider than the N(0, 1) distribution as $|\rho|$ increases in value. As a result, the LIMLE is often highly inefficient for larger values of $|\rho|$. Moreover, as illustrated in Tables 1–3, this inefficiency generally increases with $TR_{21}$ and $L$ and remains for moderate to larger values of $\Phi_{11}$ and $\Phi_{22}$. For instance, when the concentration parameters were on the order of 250, we consistently observed 2.5 and 97.5 percentile values for $Z_{LML}$ whose absolute values exceeded 3 when $L = 9$ and absolute values greater than 5 when $L = 18$. These values are significantly larger than those recorded by AKS for comparable levels of the concentration parameter when $G_1 = 1$ and indicate that when $G_1 > 1$, the distribution of $Z_{LML}$ narrows to the N(0,1) distribution at an appreciably slower rate than that observed for models with a single endogenous variable.

The results in Tables 1–3 show that a dramatic improvement in efficiency is possible when modified estimators based on relatively small values of $h$ are used in place of the LIMLE and that the risk of severe bias often connected with the TSLS and SDL estimators can be averted by selecting smaller values of $h$. Thus, selecting a modified LIMLE with $h$ in the lower portion of the range 1 to $L$ provides a preferred alternative to the LIML and TSLS estimators over a wide range of the parameter space. For instance, we find for medium size models (e.g. $L = 9$) that the SD2 estimator ($h = 2$) provides a good compromise between the LIML and TSLS estimators in the sense that, unlike the LIMLE, it is typically near the level of efficiency implied by the normal approximation and avoids the serious level of bias and slow approach to normality that AKS associate with the TSLE. In smaller models (e.g. $L = 3$) the SD1 estimator appears to be a reasonable alternative.
In general, as the number of overidentification restrictions increases, the scaled down LIMLE becomes less efficient. For example, when $L = 18$ the sampling distributions of the SD1 and SD2 estimators can be appreciably wider than what would be suggested by the normal approximation. As a result, somewhat larger values of $h$ might be considered if $L$ is large.

The results just summarized hold for models in which the variables represented by matrix $X$ are strictly exogenous. One important situation in which this is not the case is when one or more of the structural equations contain lagged endogenous variables. Since dynamic models of this type frequently occur in application, we performed an additional set of Monte Carlo experiments in order to determine whether the observations made above concerning the properties of the $k$-class estimators for models with strictly exogenous variables also hold for dynamic models.

The simulation model employed in the experiments discussed below is similar to the one used above. This enabled us to employ previous simulation results to strategically select particular combinations of the critical parameters upon which to focus our investigation of the distributions of the $k$ class estimators for the coefficients in dynamic models. Specifically, letting $y_L$ represent a $(T \times 1)$ vector which contains the observations on $y$ lagged one period, we considered a system of three simultaneous equations in which the first equation is the following modification of Eq. (6)

$$y = \beta_1 Y_{11} + \beta_2 Y_{12} + \gamma y_L + X_1 \gamma + \mu_1$$  (7)

The remaining two equations in the system resemble Eq. (7) in that they also contain first-order lags of their ‘dependent’ variables $Y_{11}$ and $Y_{12}$. In the simulation the coefficients of the lagged endogenous variables were chosen so that the moduli of the eigenvalues of the dynamic system were all less than one insuring the stability of the model; and, for the sake of simplicity, we set all three coefficients equal to the same value. With regards to the other model parameters we let $\Omega$, the variance–covariance matrix of the reduced form errors, equal a diagonal matrix. Doing this allowed us to use the $\beta$ vectors employed in the earlier simulation experiments to generate cases in which the parameter $TR_{21}$ takes on assorted values. Finally, to facilitate the generation of variables $y$, $Y_{11}$, and $Y_{12}$ and their corresponding first-order lags, we specified that each of the $x$-variables followed a first-order autoregressive process, $X_{it} = \delta_0 + \delta_1 X_{it-1} + W_{it}$, in which $\delta_1$ ranged from 0.5 to 0.8.

Given the general specification described above, we used Monte Carlo methods to generate 30,000 vectors (each of length 500) for variables $y$, $Y_{11}$, and $Y_{12}$; and used the last $T$ observations to estimate the model in Eq. (7). Summary data similar to those reported in Tables 1–3 are provided in Table 4 for four cases labeled (1)–(4). Cases (1) and (4) correspond to models in which the critical parameter $TR_{21}$ takes on larger and smaller values, while cases (2) and (3) illustrate how the presence of a lagged endogenous variable can affect the bias of
the $k$-class estimator. In each of the four cases reported in Table 4, $L = 5$ and $T - K = 20$ and 100.

Because the exogenous variables include first-order lags of $y$, $Y_{11}$, and $Y_{12}$, we cannot, as in the previous simulations, control the exact level of $TR_{21}$ nor the exact value of $\Phi = \Omega_{22} \Pi'_{22} X'_2 M_1 X_2 \Pi_{22}$ at the outset of each experiment. However, by setting $\beta = (1, 1)'$, $(\sqrt{2}, 0)'$, and $(1, -1)'$ we were able to insure that the average value of $TR_{21}$ for each experiment could be classified as large, moderate, or small. For instance, for the first case reported in Table 4 we set $\beta$ equal to $(1, -1)'$ which resulted in an average value of $TR_{21} = 0.72$ for the 30,000 samples used to compile the data for this case, while for case (4) $\beta$ was set equal to $(1, 1)'$ which yielded an average $TR_{21} = 17.60$. In addition, by adjusting the scale of the $X$ variables we were able to exert some control over the average values of $\Phi_{11}$ and $\Phi_{22}$. This was done for the cases corresponding to $T - K = 20$ to obtain average values of $\Phi_{11}$ and $\Phi_{22}$ on the order of 50. The average values of $\Phi_{11}$, $\Phi_{22}$, and $\Phi_{12}$ for the 30,000 samples used in each of the four cases reported in Table 4 are recorded immediately above the data provided for each case. Note that the values of the off diagonal elements correspond to a value of $\rho$ equal to approximately 0.80. In order to investigate how the distributions of the estimators change as the noncentrality parameters become larger, we repeated each experiment after increasing $T - K$ from 20 to 100. In these cases the average values of $\Phi_{11}$ and $\Phi_{22}$ increased to values on the order of 200–300.

The main differences observed between the results of the simulations based on models with fixed exogenous variables and the results on models with lagged endogenous variables relate to the bias of the estimators. We examine these differences below by reviewing the results in Table 4. These results suggest that there are potentially two sources for the bias of the $k$-class estimators of $\beta_1$ in Eq. (7): the basic simultaneous equation bias which was observed above in the simulations based on the model in Eq. (6) and an additional source of bias due to the presence of the lagged endogenous variable in structural Eq. (7). In some cases the two sources of bias reinforce one and another; and as a result, there is a tendency for the $k$-class estimators for $\beta_1$ to be more biased in this situation, as is seen for case (1) in Table 4. To illustrate this fact we conducted, for the purposes of comparison, a simulation experiment in which all the exogenous variables were fixed and for which each of the critical parameters were identical to those listed for case (1). This experiment yielded 50th percentile values of $-0.03$, $-0.27$, $-0.47$, $-0.88$, and $-0.84$, respectively, for $Z_{LML}$, $Z_{SD1}$, $Z_{SD2}$, $Z_{SDL}$ and $Z_{TSL}$ when $T - K = 20$, and values of $-0.01$, $-0.11$, $-0.22$, $-0.49$, and $-0.49$ when $T - K = 100$. Note that each of these values is somewhat lower than the corresponding value observed for case (1) in Table 4.

How the presence of a lagged-dependent variable influences the degree of bias observed for the $k$-class estimators can be explored further by varying the value of the coefficients on the lagged endogenous variables. We found that, as these
Table 4
*DF* = 20, *DF* = 100

<table>
<thead>
<tr>
<th>EST</th>
<th>Beta</th>
<th>TR21</th>
<th>PHI = 51.31 51.77 38.74</th>
<th>PHI = 196.54 209.18 147.29</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z₀₂₅  Z₅₀  Z₉₇.₅</td>
<td>A  B  C  D</td>
</tr>
<tr>
<td>LML</td>
<td>1.00</td>
<td>0.72</td>
<td>−2.03 −0.30 4.67</td>
<td>0.21 0.43 0.89 0.61</td>
</tr>
<tr>
<td>SD1</td>
<td>−1.00</td>
<td></td>
<td>−2.03 −0.51 1.72</td>
<td>0.22 0.46 0.95 0.71</td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td>−2.04 −0.69 1.05</td>
<td>0.20 0.44 0.97 0.82</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>−2.10 −1.05 0.20</td>
<td>0.09 0.24 0.96 0.95</td>
</tr>
<tr>
<td>TSL</td>
<td>ALPHA = 0.90</td>
<td></td>
<td>−2.15 −1.04 0.53</td>
<td>0.11 0.27 0.95 0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EST</th>
<th>Beta</th>
<th>TR21</th>
<th>PHI = 50.68 50.59 40.78</th>
<th>PHI = 230.85 284.67 204.91</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z₀₂₅  Z₅₀  Z₉₇.₅</td>
<td>A  B  C  D</td>
</tr>
<tr>
<td>LML</td>
<td>1.00</td>
<td>0.66</td>
<td>−1.69 −0.09 5.36</td>
<td>0.23 0.46 0.88 0.53</td>
</tr>
<tr>
<td>SD1</td>
<td>−1.00</td>
<td></td>
<td>−1.74 −0.33 1.89</td>
<td>0.25 0.51 0.96 0.65</td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td>−1.80 −0.53 1.11</td>
<td>0.25 0.52 0.98 0.76</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>−1.95 −0.94 0.20</td>
<td>0.11 0.30 0.98 0.95</td>
</tr>
<tr>
<td>TSL</td>
<td>ALPHA = 0.45</td>
<td></td>
<td>−2.00 −0.90 0.73</td>
<td>0.14 0.33 0.97 0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EST</th>
<th>Beta</th>
<th>TR21</th>
<th>PHI = 51.28 50.22 40.20</th>
<th>PHI = 199.77 191.193 154.50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z₀₂₅  Z₅₀  Z₉₇.₅</td>
<td>A  B  C  D</td>
</tr>
<tr>
<td>LML</td>
<td>−1.00</td>
<td>0.68</td>
<td>−9.73 −0.31 2.18</td>
<td>0.21 0.42 0.79 0.60</td>
</tr>
<tr>
<td>SD1</td>
<td>1.00</td>
<td></td>
<td>−2.29 0.08 2.03</td>
<td>0.27 0.53 0.93 0.47</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td>−1.37 0.35 2.00</td>
<td>0.29 0.56 0.97 0.32</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>−0.34 0.85 2.05</td>
<td>0.14 0.37 0.97 0.07</td>
</tr>
<tr>
<td>TSL</td>
<td>ALPHA = 0.90</td>
<td></td>
<td>−0.95 0.81 2.10</td>
<td>0.17 0.38 0.96 0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EST</th>
<th>Beta</th>
<th>TR21</th>
<th>PHI = 50.05 49.39 40.92</th>
<th>PHI = 271.41 289.81 233.79</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z₀₂₅  Z₅₀  Z₉₇.₅</td>
<td>A  B  C  D</td>
</tr>
<tr>
<td>LML</td>
<td>1.00</td>
<td>17.60</td>
<td>−9.90 −0.01 9.71</td>
<td>0.16 0.33 0.70 0.50</td>
</tr>
<tr>
<td>SD1</td>
<td>1.00</td>
<td></td>
<td>−2.86 −0.04 2.85</td>
<td>0.20 0.40 0.84 0.51</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td>−2.25 −0.07 2.16</td>
<td>0.23 0.46 0.92 0.52</td>
</tr>
<tr>
<td>SDL</td>
<td></td>
<td></td>
<td>−1.68 −0.15 1.41</td>
<td>0.32 0.60 0.99 0.57</td>
</tr>
<tr>
<td>TSL</td>
<td>ALPHA = 0.90</td>
<td></td>
<td>−1.93 −0.14 1.78</td>
<td>0.29 0.56 0.96 0.56</td>
</tr>
</tbody>
</table>
coefficients were reduced in value, the overall level of bias generally declined. For example, in the experiment labeled as case (2) in Table 4 the coefficients of the lagged endogenous variables (ALPHA) were reduced to 0.45 from the value of 0.90 in case (1). The impact of this change was to reduce the largest eigenvalue of the dynamic system from 0.90 in case (1) to 0.45 for case (2) so that in one case we have a relatively stable system and in the other we have a system closer to the nonstationary region. Note that the overall level of bias associated with the $k$-class estimators for the model in case (2) is smaller than what is observed for the model in case (1).

Finally, in some situations we found that the two sources of bias offset one and another leading to a result that is somewhat different from what was presented in Tables 1–3 and cases (1) and (2) in Table 4. For example, refer to case (3) in Table 4. Observe in this case that changing the sign of $\beta$ reverses the direction of the simultaneous equation bias from the direction observed in case (1); and as a result, the simultaneous equation bias offsets the bias due to the presence of the lagged endogenous variable. Notice that in this case the SD1 estimator is preferred to the LIML estimator in terms of both location and spread whereas in the previous set of experiments based on Eq. (6) the distribution of the LIMLE was always the most centrally located of the distribution.

With the exception of the minor differences noted above, the results presented in Table 4 resemble the results presented in Tables 1–3. For instance, we found, as before, that the overall level of bias of the $k$-class estimators tended to decline as the value of $TR_{21}$ increased in value so that for models with larger values of $TR_{21}$ the overall level of bias associated with $b_1(k)$ tended to be relatively low (e.g. see case (4) in Table 4). Moreover, the results in Table 4 imply that in dynamic models setting $h = 1$ provides a scaled down estimator which is more efficient than the LIMLE and that for $h > 1$, the bias of the scaled down LIML estimator increases with $h$ so that the tradeoff involved with the selection of $h$ identified in the previous simulations is maintained. Overall, then, the results suggest that the advantages derived from using a scaled down LIML estimator with small $h$ in lieu of the LIML or TSLS estimators extend to dynamic models.

5. Conclusion

This paper investigates the general properties of a $k$-class estimator in which $k = (1 - h/(L + T - K)) \lambda_0$ with $h > 0$. Because for any value $h > 0$ the term $k$ is smaller than the LIML root $\lambda_0$, these estimators are referred to as scaled down LIML estimators. An examination of the small $\sigma$ properties of these modified estimators demonstrates that they dominate the LIMLE and Fuller modified LIML estimators in terms of asymptotic mean squared error and concentration of probability.
The empirical relevance of the suggested modification is demonstrated through a Monte Carlo simulation in which it is shown that using a scaled down LIMLE with \( h = 1 \) provides an estimator that is more efficient than the LIMLE in finite samples and that this improvement in efficiency can be dramatic for practical values of the model’s critical parameters when \( G_1 \geq 2 \). This leads us to recommend that a scaled down LIMLE with \( h = 1 \) be used in place of the LIMLE.

In general, we find that choosing among values of \( h \) greater than one involves a tradeoff between bias and variance. Larger values of \( h \) are associated with higher levels of bias, and the properties of scaled down LIML estimators with \( h \) on the order of \( L \) resemble those of the TSLSE. Thus, the risk of severe bias and the slowness in approach to normality that led Anderson et al. (1982) to recommend against the application of the TSLSE leads us to recommend against the application of scaled down LIML estimators with values of the shrinkage parameter \( h \) falling outside of the lower portion of the range \( 1-L \). An extensive Monte Carlo study demonstrates that restricting \( h \) to the lower portion of the range \( 1-L \) yields a scaled down LIML estimator that is preferred over the LIML and TSLS estimators for a wide range of the parameter space.

Appendix A

The results in Anderson et al. (1986) can be used to establish Theorems 3.1 and 3.2.

Proof of Theorem 3.1. Consider \( b(k_1) \) with \( k_1 = (1 - h/(L + T - K))\lambda_0 \) and \( b(k_2) \) with \( k_2 = \lambda_0 - h/(T - K) \) where \( h < 2(L + T - K) \). Refer to Eq. (A.33) and (A.34) in Anderson et al. (1986) and take \( \hat{e}_1 = e_{k1} \) and \( \hat{e}_2 = e_{k2} \). Referring to the expression for AMSE in Lemma 3.2 and Eq. (A.34) it can be seen that

\[
d_1 = 0, \quad d_2 = \frac{-h^2L}{(L + T - K)(T - K)}
\]

and

\[
d_3 = \frac{h^2L}{(L + T - K)(T - K)(T - K - 2)} - \frac{2hL}{(T - K)(T - K - 2)}
\]

Since \( d_2 < 0 \) and \( d_3 < 0 \) if \( h < 2(L + T - K) \) it follows that \( APC_\sigma(b(k_1)) > APC_\sigma(b(k_2)) \) and \( AMSE(b(k_1)) < AMSE(b(k_2)) \).

Proof of Theorem 3.2. Consider \( b(k_1) \) with \( k_1 = (1 - \chi_1/(L + T - K))\lambda_0 \) and \( b(k_2) \) with \( k_2 = (1 - h/(L + T - K))\lambda_0 \) where \( 0 \leq \chi_1 \leq h \). Refer to Eqs. (A.33) and (A.34) in Anderson et al. (1986) and take \( \hat{e}_1 = e_{k1} \) and \( \hat{e}_2 = e_{k2} \). Referring to
the expression for AMSE in Lemma 3.2 and Eq. (A.34) it can be seen that
\[ d_1 = \frac{1}{2} (\chi_1 - h), \]
\[ 2d_2 = (L + T - K + 2) (\chi_1^2 - h^2) - 2(\chi_1 - h) \]
and
\[ 2(T - K - 2)d_3 = -2(L + T - K - 2)(\chi_1 - h) \]
\[ + (L + T - K - 2) (\chi_1^2 - h^2). \]

Note that \( d_1 \leq 0 \) and \( 2(T - K - 2)d_3 = (x_1 - h)[ -2(L + T - K - 2) + ((L + T - K - 2)/L + T - K)(x_1 + h)] \) implies that \( d_3 \geq 0 \) as long as \( T - K > 2 \) and \( h \leq L + T - K \). Since \( 2d_2 = (\chi_1 - h)[((L + T - K + 2)/L + T - K)(\chi_1 + h) - 2] \) implies that \( d_2 \geq 0 \) if \( h \leq (L + T - K)/(L + T - K + 2) \), it follows that \( \text{APC}_a(b(k_1)) < \text{APC}_a(b(k_2)) \) when \( h < L + T - K/L + T - K + 2 = h' \).

The result concerning AMSE follows directly from the expression for \( \hat{h} \) in Section 3.

**References**


Kadane, J., 1971. Comparison of \( k \)-class estimators when the disturbances are small. Econometrica 39, 723–737.


