OPTIMAL ENDOGENOUS INSTRUMENTAL VARIABLES ESTIMATION IN NONLINEAR SYSTEMS

BRYAN W. BROWN[1]
RICE UNIVERSITY

REVISED, SEPTEMBER 1997

ABSTRACT

The instrumental variables (IV) approach has been widely applied in situations where the structural model is parameterized but the disturbance distribution is nonparametric. In this semiparametric setting, a great deal of attention has been given to the selection of an optimal set of instruments. By and large, however, the suggested instruments have been asymptotically endogenous (or predetermined) in the sense that they are functions only of the exogenous (predetermined) variables and not the disturbances. In this paper, I extend previous work to study endogenous instrumental variables estimation of nonlinear systems under semiparametric assumptions. For the cases of disturbances independent of the exogenous variables, conditional symmetry, and a combination of both, a simple and feasible endogenous instrumental variables estimator is developed. The asymptotic behavior of the estimators is investigated and compared to previous estimators of the model. The new estimators can be combined with previously proposed exogenous instrumental variables estimators to obtain estimators that globally dominate the latter, in terms of asymptotic quadratic loss. A Sampling experiment conducted on a simple nonlinear model suggests that the new estimators can perform exceedingly well when the underlying stochastic assumption is satisfied.

KEYWORDS: Instrumental variables estimation, semiparametric estimation, nonlinear simultaneous systems, semiparametric efficiency bounds.

PRELIMINARY: Please do not quote without author's permission.

OPTIMAL ENDOGENOUS INSTRUMENTAL VARIABLES ESTIMATION IN NONLINEAR SYSTEMS

BRYAN W. BROWN[1]

1. INTRODUCTION

The instrumental variables (IV) approach has, rightfully, enjoyed a prominent position in the econometrics literature. In particular, IV has been widely applied in situations where the structural model is parameterized but the disturbance distribution is nonparametric. In this semiparametric setting, a great deal of attention has been given to the selection of an optimal set of instruments. By and large, however, the suggested instruments have been asymptotically predetermined in the sense that they are functions only of the exogenous or lagged endogenous variables and not the current endogenous variables or disturbances. Under various stochastic assumptions, including the disturbances being independent of the predetermined variables or having conditional mean zero given the predetermined variables, such instruments will be assured of being uncorrelated with the disturbances. Of course, beyond uncorrelatedness with the disturbances, we seek instruments that will maximize the efficiency (minimize the variances) of the resulting estimators.

For several leading cases, the optimal instrumental variables estimator turns out to have predetermined instruments and, moreover, attains the semiparametric efficiency bound. In static linear models, under an independence assumption, the three-stage least squares estimator, which uses a linear transformation of the exogenous variables as instruments, is the efficient instrumental variables estimator and, at least locally, attains the semiparametric efficiency bound. In static nonlinear models, with disturbances unrestricted except to have conditional mean zero, Chamberlain [1987] has shown that the estimator which achieves the semiparametric efficiency bound is an IV estimator with exogenous instruments. Similarly, for (possibly) dynamic models, with disturbances only restricted to have mean zero, the optimal GMM estimator, which attains the semiparametric efficiency bound, is an IV estimator with nonstochastic and hence predetermined instruments (see Hansen [1984]).

More specifically, for the nonlinear model under more restrictive stochastic assumptions, such as independence or conditional symmetry, there is no reason why functions of endogenous variables cannot be uncorrelated with the disturbances and hence improved estimators obtained through the use of endogenous instruments. In a parametric context, Amemiya [1977] has shown that the maximum likelihood estimator for the nonlinear simultaneous equation model with normally distributed disturbances has an instrumental variables interpretation with instruments, as it turns out, that are functions of endogenous variables. In a semiparametric context, Hausman, Newey, and Taylor [1988] have demonstrated that the maximum likelihood estimator under normality for the linear model with covariance restrictions, which is really a nonlinear model, is an endogenous instrumental variables estimator and remains consistent under nonnormality and is hence semiparametric. Similarly, Brown and Mariano [1989] have argued that the maximum likelihood estimator for a specific triangular nonlinear simultaneous equations model is a simple endogenous instrumental variables estimator and is consistent under a variety of misspecifications of the disturbance distribution.[2]

[1] I am endebted to Bobby Mariano, Xing Ming, Whitney Newey, and Ingmar Prucha for helpful comments at various stages of the development of this paper. I am, of course, solely responsible for any remaining shortcomings.

[2] Breusch, Mizon, and Schmidt [1989], in a panel data model where the individual effects may be correlated with a

The objective of this paper is to systematically examine the possibilities for obtaining improved instrumental variables estimators for nonlinear systems through the use of endogenous instrumental variables. The basic approach taken is to purge the regressors that would be used in the absence of simultaneity of their linear relationship with the disturbances but preserve the remaining nonlinear relationship. The formal model is introduced under a variety of stochastic assumptions, in the next section, together with some basic asymptotic properties of instrumental variables estimators, including endogenous IV estimators. In the third section, a simple class of endogenous IV estimators is introduced and analyzed, for each stochastic assumption, based on efficient estimates of the residual of the regressors projected onto the disturbances. The semiparametric efficient estimators for each stochastic model are shown to have an endogenous IV interpretation, in the fourth section and compared to the simple endogenous IV estimators. A sampling experiment is conducted on a simple nonlinear model, in the fifth section, to verify the validity of the asymptotic results in moderate sized samples.

The major findings of this paper are the following. For the cases of disturbances independent of the exogenous variables, conditional symmetry, and combined conditional symmetry and independence, simple and feasible endogenous instrumental variables estimators are developed. The asymptotic behavior of these estimators is established and compared to previous estimators of the model. The new estimators are found to be consistent and asymptotically normal when the stochastic assumption which underlies their form is satisfied. The new estimators are generally indefinite relative to the familiar best nonlinear three-stage least squares estimator but may be combined to obtain an estimator that globally dominates either. For some special cases, the new estimators are found to locally attain their respective semiparametric efficiency bound when the true distribution of the disturbances is normal. In the sampling experiment, which was an example of the special case, the performance of the new estimators, particularly the combined independence and symmetry estimator, was particularly promising.

2. Model and Basic Concepts

Consider, as a point of departure, a static nonlinear system, which can be formally represented

$$\epsilon = \rho(z, \beta), \tag{2.1}$$

where z is a $(g+k)\times 1$ vector of the observable variables, ϵ is a $g \times 1$ disturbance vector, $\rho(y, x, \beta)$ is a $g \times 1$ vector of known functions, and β is a $p \times 1$ vector of unknown parameters. If (2.1) are the structural equations of a simultaneous equation system, it is useful to to utilize the more explicit representation

$$\epsilon = \rho(y, x, \beta), \tag{2.2}$$

where y is a $g \times 1$ vector of endogenous variables and x is a $k \times 1$ vector of exogenous variables. Since (2.2) may be viewed as the underlying structure which generates the joint behavior of the

subset of the explanatory variables, propose instruments that are not strictly exogenous in the sense that they include linear transformations of the correlated variables. However, the linear transformations, as proper instruments, are uncorrelated with the individual effects so the correlated variables can be rewritten as the simple sum of a component which is linear in the proposed instruments and in a real sense is exogenous and another that is correlated. In fact, they point out that we can write formal reduced form equations with the uncorrelated (transformed) variables as explanatory variables.

observable variables, in the latter case, we assume the existence of the unique inverse relationship

$$y = \pi(\epsilon, x, \beta). \tag{2.3}$$

In practice, the inverse form given by (2.3), may not be available in closed form but can be obtained by numerical solution techniques.

In this paper, we only consider the case where the vector of driving variables $(\epsilon', x')'$ are jointly $i.i.d.$ and hence the combined vector of observable variables $z = (y', x')'$ is also jointly $i.i.d.$ The joint distribution of $(\epsilon', x')'$ will remain unspecified and unrestricted other than additional assumptions on the degree and nature of the dependence between ϵ and x and some smoothness restrictions, which makes this a semiparametric problem. The alternative additional semiparametric assumptions, in rough order of increasing restrictiveness, are $E[\epsilon] = 0$, $E[\epsilon|x] = 0$, ϵ symmetrically distributed given x, ϵ independent of x, and ϵ symmetric and independent of x. The vector β will include intercept parameters and ϵ will have mean zero except for the case where ϵ is only assumed to be independent of x. The existence of an inverse relationship, as given by (2.3), is not required for the zero unconditional mean and zero conditional mean cases. In any case, ϵ is assumed to have finite unconditional and conditional covariance matrices.

The instrumental variables approach is a straightforward way to transform the semiparametric restrictions into a simple m-estimator. In a nonlinear model, the $g \times p$ matrix of derivatives

$$R(z, \beta) = \partial\rho(z, \beta)/\partial\beta' \tag{2.4}$$

is analogous to the regressor matrix in a linear model. Proper instruments would be any similarly dimensioned matrix function of the observable data, say $Q(z)$, such that the instruments are, under the semiparametric restriction, uncorrelated with the disturbances

$$0 = E[Q(z)'\epsilon] = E[Q(z)'\rho(z_t, \beta_0)]. \tag{2.5}$$

Based on the uncorrelated condition, we obtain the instrumental variables (IV) estimator as the solution to the (possibly) nonlinear set of p equations

$$0 = \sum_{t=1}^{n} Q(z_t)'\rho(z_t, \hat{\beta}). \tag{2.6}$$

For the moment we treat $Q(z)$ as known and ignore the possibility that it might require estimates of unknown parameters including β.

Under fairly general conditions, the asymptotic behavior of the estimator defined in (2.6) is well known. Specifically, we have

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N[0, P^{-1}M \cdot P^{-1\prime}] \tag{2.7}$$

where

$$\begin{aligned} P =& \mathrm{E}[Q(z)'R(z, \beta_0)] \text{ finite and nonsingular} \\ M =& \mathrm{E}[Q(z)'\rho(z, \beta_0) \cdot \rho(z, \beta_0)'Q\,(z)] \text{ finite.} \end{aligned} \tag{2.8}$$

Implicit in (2.8) is the notion that the instruments should be sufficiently well correlated with the regressors. In fact, it is widely argued that instruments should be as highly correlated with the regressors as possible, subject to the condition of orthogonality relative to the disturbances. This argument will be formalized below.

In the absence of simultaneity, the most highly correlated instrument matrix is obviously the regression matrix suitably transformed. Suppose the unconditional covariance matrix

$$\Sigma_0 = E[\rho(z,\beta_0)\cdot\rho(z,\beta_0)'] - E[\rho(z,\beta_0)]\cdot E[\rho(z,\beta_0)]' \tag{2.9}$$

is nonsingular, then the instruments

$$Q(z) = \Sigma_0^{-1} R(z,\beta_0) \tag{2.10}$$

define the nonlinear seemingly unrelated regression (NLSUR) as the solution to the equations

$$0 = \sum_{t=1}^{n} R(z_t,\widetilde{\beta})'\widetilde{\Sigma}^{-1}\rho(z_t,\hat{\beta}). \tag{2.11}$$

where $\widetilde{\beta}$ and $\widetilde{\Sigma}$ are preliminary estimators. The problem, in general, is that

$$E[R(z,\beta_0)'\Sigma_0^{-1}\rho(z,\beta_0)] = d \neq 0, \tag{2.12}$$

due to simultaneity, whereupon the NLSUR estimator will be inconsistent.

The first line of defense against simultaneity has typically been to utilize predetermined or, in the current static model, exogenous instruments. By exogenous instruments, I mean that the instruments are, at least asymptotically, not a function of the disturbances. Prominent examples of the use of exogenous instruments are GMM for the zero unconditional mean case and Chamberlain's [1987] estimator for the zero conditional mean case. In addition, for the cases where $E[\epsilon|x] = 0$ and $E[\epsilon\epsilon'|x] = \Sigma_0$, Amemiya [1977] has proposed the best nonlinear three-stage least squares instruments

$$Q(z) = \Sigma_0^{-1}\overline{R}(x,\theta_0) \tag{2.13}$$

where Σ_0 is given above and

$$\overline{R}(x,\theta_0) = E[R(z,\beta)|x]. \tag{2.14}$$

In fact, these are the optimal instruments that are functions only of x. Various authors have shown how asymptotically appropriate estimators of these instruments can be utilized to obtain a feasible BNL3S estimator which is asymptotically equivalent to the estimator that uses the true values in (2.13) as instruments.[3]

The basic idea behind the exogenous instruments is to purge the regressors completely of any relationship with ϵ, leaving only possible dependence on x. In fact, according to the criteria introduced above, all we want to do is purge the regressors of their linear relationship with the disturbances. In a linear model, where we can decompose the solution of the model into the sum of an exogenous component depending only on x and a endogenous component depending only on ϵ, the two approaches are equivalent. In nonlinear models, however, there is a real distinction between purging the regressors of only the linear relationship, which would leave endogenous instruments, and purging the regressors completely of any relationship, which would leave exogenous instruments. It is conceivable that the endogenous instruments will be more highly correlated with the

[3] See Brown [1990], Robinson [1991], Newey [1990a], and Rilstone [1989] for various approaches to feasible BNL3S.

regressors and hence yield improved precision relative to a exogenous approach. Consequently, in the sequel, we consider the possibility that $Q(z)$ is a nontrivial function of ϵ.

If the instruments involve unknown parameters, the usual procedure has been to use preliminary estimates and proceed as when the instruments are known. Formally, we have

$$0 = n^{-1/2} \sum_{t=1}^{n} Q(z_t, \widetilde{\theta})' \rho(z_t, \hat{\beta}) \tag{2.15}$$

where $\widetilde{\theta}$ is a preliminary estimator of $\theta = (\beta', \eta')'$ and η is a vector of nuisance parameters. Unfortunately, this approach is (potentially) problematical when the instruments are endogenous. This is best seen through the following asymptotic expansion, which is valid under fairly general conditions that rule out complications from $\widetilde{\eta}$,

$$\begin{aligned} 0 \;=\;& n^{-1/2} \textstyle\sum_{t=1}^{n} Q(z_t, \theta_0)' \rho(z_t, \beta_0) + n^{-1} \sum_{t=1}^{n} Q(z_t, \theta_0)' R(z_t, \beta_0) \cdot n^{1/2}(\hat{\beta} - \beta_0) \\ & + \textstyle\sum_{i=1}^{p} n^{-1} \sum_{t=1}^{n} \{\partial Q(z_t, \theta_0)/\partial \beta_i\}' \rho(z_t, \beta_0) \cdot n^{1/2}(\widetilde{\beta}_i - \beta_{i0}) + o_p(1). \end{aligned} \tag{2.16}$$

Solving for $n^{1/2}(\hat{\beta} - \beta_0)$ we find the asymptotic distribution of $\hat{\beta}$ depends, in general, on the distribution of $n^{1/2}(\widetilde{\beta} - \beta_0)$ unless $E[\{\partial Q(z, \theta_0)/\partial \beta_i\}' \rho(z, \beta_0)] = 0$. Consequently, we must iterate on the preliminary estimator of β as well, which means we are effectively solving

$$0 = n^{-1/2} \sum_{t=1}^{n} Q(z_t, \hat{\beta}, \widetilde{\eta})' \rho(z_t, \hat{\beta}). \tag{2.17}$$

The appeal of the exogenous instruments is now evident, since $E[\epsilon | x] = 0$ implies the condition will likely be met if $Q(z, \theta_0)$ is a function only of x.

The asymptotic behavior of the estimator yielded by (2.17) is somewhat different when $E[\{\partial Q(z, \theta_0)/\partial \beta_i\}'$ 0. Assuming that estimation of η introduces no additional problems, the asymptotic behavior of the estimator is still given by

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N[0, P^{-1} M \cdot P^{-1\prime}], \tag{2.18}$$

as in (2.6), but now

$$\begin{aligned} & P = E[\partial \{Q(z, \beta_0, \eta_0)' \rho(z, \beta_0)\}/\beta'] \text{ finite and nonsingular} \\ & M = E[Q(z, \theta_0)' \rho(z, \beta_0) \cdot \rho(z, \beta_0)' Q(z, \theta_0)] \text{ finite.} \end{aligned} \tag{2.19}$$

Under fairly general conditions, these components can be consistently estimated for inference purposes by

$$\begin{aligned} & \hat{P} = n^{-1} \textstyle\sum_{t=1}^{n} [\partial Q\{(z_t, \hat{\beta}, \widetilde{\eta})' \rho(z_t, \hat{\beta})\}/\beta'] \\ & \hat{M} = n^{-1} \textstyle\sum_{t=1}^{n} [Q\{(z_t, \hat{\beta}, \widetilde{\eta})' \rho(z_t, \hat{\beta}) \cdot \rho(z_t, \hat{\beta})' Q(z_t, \hat{\beta}, \widetilde{\eta})]. \end{aligned} \tag{2.20}$$

Strictly speaking, we do not have to iterate to convergence to obtain this asymptotic behavior, but may stop after one step of a Newton-Raphson procedure.

3. A Simple Class of IV Estimators

In this section, we obtain a class of estimators by directly applying the criteria for instruments laid out in the previous section. Specifically, we will attempt to utilize instruments that are uncorrelated with the disturbances but are as highly correlated with the regressors as possible. This will be accomplished by purging the instruments only of the component that is correlated with the disturbances. We will first introduce the general approach and point out some possible difficulties and then apply the approach to each of the stochastic cases mentioned in the previous section. In many of these cases, the new approach yields endogenous instruments with a rather natural interpretation.

Consider the i^{th} column of the regressor matrix

$$r_i(z, \beta_0) = R(z, \beta_0) \cdot e_i \tag{3.1}$$

where e_i is a $p \times 1$ vector which is zero except for element i, which is unity. The corresponding instrument which is uncorrelated with $\epsilon = \rho(z, \beta_0)$ but most highly correlated with $r_i(z, \beta_0)$ is given as the residual of the least squares projection

$$r_i^*(z, \theta_0) = r_i(z, \beta_0) - C_i \cdot D^{-1} \rho(z, \beta_0) \tag{3.2}$$

where $C_i = E[r_i(z, \beta_0) \cdot \rho(z, \beta_0)']$ and $D = E[\rho(z, \beta_0) \cdot \rho(z, \beta_0)']$. If we first transform as with the SUR estimator, then the optimal instruments by the criteria set out above, are

$$Q(z) = \Sigma_0^{-1} R^*(x, \beta_0, \eta_0) \tag{3.3}$$

where $r_i^*(z, \beta_0) = R^*(z, \beta_0) \cdot e_i$ and η consists of D and the various C_i. Note that $D \neq \Sigma_0$ is possible for cases, such as independence, where $E[\rho(z, \beta_0)] = 0$ is not guaranteed.

Now $E[\{\partial Q(z, \theta_0)/\partial \beta_i\}' \rho(z, \beta_0)] \neq 0$, in general, for this choice of $Q(z)$, so we will not use a preliminary estimator of β in the instruments. Accordingly, a feasible IV estimator based on these instruments is given by

$$0 = \sum_{t=1}^{n} R^*(z_t, \hat{\beta}, \widetilde{\eta})' \widetilde{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \tag{3.4}$$

where $\widetilde{\eta} = (\widetilde{D}, \widetilde{C}_i)$ and $\widetilde{\Sigma}$ are preliminary estimators. The problem is, of course, the appropriate choice of estimators for D, C_i and Σ_0. Unfortunately, if, based on a preliminary estimator $\widetilde{\beta}$, we use the rather natural sample analogs as estimators of these expectations, namely

$$\begin{aligned} \widetilde{D} &= n^{-1} \textstyle\sum_{t=1}^{n} \rho(z_t, \widetilde{\beta}) \cdot \rho(z_t, \widetilde{\beta})' \\ \widetilde{C}_i &= n^{-1} \textstyle\sum_{t=1}^{n} r_i(z_t, \widetilde{\beta}) \cdot \rho(z_t, \widetilde{\beta})' \\ \widetilde{\Sigma} &= \widetilde{D} - [n^{-1} \textstyle\sum_{t=1}^{n} \rho(z_t, \widetilde{\beta})] \cdot [n^{-1} \textstyle\sum_{t=1}^{n} \rho(z_t, \widetilde{\beta})]', \end{aligned} \tag{3.5}$$

then the estimator proposed here becomes trivial, as will be seen below, since $\hat{\beta} = \widetilde{\beta}$ will satisfy (3.4). The solution proposed in this section is to utilize the efficient semiparametric (distribution free) estimators of the expectation functions proposed in Brown and Newey [1992]. Of course, as was shown in that work, the form of the estimators depends on the stochastic restriction employed.

Considerable insight can be gained into this approach by placing (3.4) in an alternative representation. Let $\hat{\eta} = (\hat{D}, \hat{C}_i)$ and $\hat{\Sigma}$ be our estimators, then the $i-th$ estimating equation becomes:

$$\begin{aligned} 0 &= \textstyle\sum_{t=1}^n r_i^*(z_t, \hat{\beta}, \hat{\eta})' \hat{\Sigma}^{-1} \rho(z_t, \hat{\beta} \\ &= \textstyle\sum_{t=1}^n [r_i(z_t, \hat{\beta})' \hat{\Sigma}^{-1} \rho(z_t, \hat{\beta}) - \rho(z_t, \hat{\beta})' \hat{D}^{-1} \hat{C}_i' \hat{\Sigma}^{-1} \rho(z_t, \hat{\beta})] \\ &= \textstyle\sum_{t=1}^n [tr\{\hat{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \cdot r_i(z_t, \hat{\beta})'\} - tr\{\hat{D}^{-1} \hat{C}_i' \hat{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \rho(z_t, \hat{\beta})'\}] \\ &= tr[\hat{\Sigma}^{-1} \widetilde{C}_i'] - tr[\hat{\Sigma}^{-1} \widetilde{D} \cdot \hat{D}^{-1} \hat{C}_i'] \end{aligned} \tag{3.6}$$

where the last line follows from division by n and a redifinition of the expressions in (3.6) using $\hat{\beta}$. Obviously, for $\hat{D} = \widetilde{D}$ and $\hat{C}_i = \widetilde{C}_i$, the equation is satisfied trivially for whatever value of β was used in forming the estimators. More generally, for $\hat{C}_i \neq \widetilde{C}_i$ but $\hat{D} = \widetilde{D}$ or $\hat{D} - \widetilde{D} = o_p(n^{-1/2})$, as turns out to be the case for the conditions of interest considered below, then

$$0 = tr[\hat{\Sigma}^{-1} \widetilde{C}_i'] - tr[\hat{\Sigma}^{-1} \hat{C}_i'] + o_p(n^{-1/2}). \tag{3.7}$$

The first term can be seen to be the NLSUR estimating equation while the second term introduces a correction for the fact that the first term converges to a nonzero value, in general, when evaluated at the true parameter point β_0.

The general approach will now be applied to the case where the disturbances are unrestricted except to have zero mean. This assumption results in a model for which GMM has been suggested as the optimal approach. Define $c_i = vec(C_i)$, $d = vec(D)$, and

$$\widetilde{\Sigma} = \widetilde{D} = n^{-1} \sum_{t=1}^n [\rho(z_t, \hat{\beta}) \cdot \rho(z_t, \hat{\beta})']. \tag{3.8}$$

Applying the results of Brown and Newey for this case yields

$$\begin{aligned} \hat{c}_i &= n^{-1} \textstyle\sum_{t=1}^n [(\rho(z_t, \hat{\beta}) \otimes r_i(z_t, \hat{\beta})) - \{n^{-1} \sum_{s=1}^n (\rho(z_s, \hat{\beta}) \otimes r_i(z_s, \hat{\beta})) \cdot \rho(z_s, \hat{\beta})'\} \widetilde{D}^{-1} \rho(z_t, \hat{\beta})] \\ \hat{d} &= n^{-1} \textstyle\sum_{t=1}^n [(\rho(z_t, \hat{\beta}) \otimes \rho(z_t, \hat{\beta})) - \{n^{-1} \sum_{s=1}^n (\rho(z_s, \hat{\beta}) \otimes \rho(z_s, \hat{\beta})) \cdot \rho(z_s, \hat{\beta})'\} \widetilde{D}^{-1} \rho(z_t, \hat{\beta})] \end{aligned} \tag{3.9}$$

as the efficient estimators of c_i and d. Substituting these estimators into (3.6) or (3.8) and rearranging, we obtain

$$0 = [n^{-1} \sum_{t=1}^n R(z_t, \hat{\beta})' \widetilde{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \cdot \rho(z_t, \hat{\beta})'] \widetilde{\Sigma}^{-1} [n^{-1} \sum_{t=1}^n \rho(z_t, \hat{\beta})] + O_p(\hat{D} - \widetilde{D}). \tag{3.10}$$

Under the assumptions of local behavior introduced in the next section, $\hat{D} - \widetilde{D}$ will be $o_p(n^{-1/2})$, so we define our estimator as the solution to the first term in (3.10).

This estimator has obvious similarities to the GMM estimator. Given the known optimality of GMM under the current stochastic assumption, the new estimator is not proposed as a serious alternative. Rather it is presented for completeness and also because it gives us additional insight into where and why the new estimator falls short. Under fairly general conditions this estimator will be consistent and have the following limiting behavior:

$$n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{d} N[0, (\overline{R}*'\,\Sigma_0^{-1}\overline{R})^{-1}\overline{R}*'\,\Sigma_0^{-1}\overline{R}*(\overline{R}'\Sigma_0^{-1}\overline{R}*)^{-1}] \tag{3.11}$$

where $\overline{R} = E[R(z_t, \beta_0)]$ and $\overline{R}* = E[\rho(z_t, \beta_0)\cdot\rho(z_t, \beta_0)'\Sigma_0^{-1}R(z_t, \beta_0)]'$. Obviously, this estimator will be equivalent to GMM if $\overline{R} = \overline{R}*$. More will be said on the relationship between the two in the next section. It is worth noting that, since the resultant instruments are asymptotically exogenous, the same asymptotic behavior will result if we use any preliminary estimator of β in forming $\widetilde{\Sigma}$ and the first bracketed component in (3.10).

We next turn to the related case where the disturbances are only assumed to have zero mean conditional on x. This form of the model allows for unspecified conditional heteroskedasticity and has been analyzed extensively by Chamberlain [1987], who has established the lower bound and an efficient semiparametric estimator for the problem. According to Brown and Newey, the efficient estimators of c_i and d would have the form

$$\hat{c}_i = n^{-1}\textstyle\sum_{t=1}^{n}\{(\rho(z_t,\hat{\beta})\otimes r_i(z_t,\hat{\beta})) - \hat{E}[(\rho(z,\hat{\beta})\otimes r_i(z,\hat{\beta}))\cdot\rho(z,\hat{\beta})'|x_t]\widetilde{\Sigma}_t^{-1}\rho(z_t,\hat{\beta})\}$$

$$\hat{d} = n^{-1}\textstyle\sum_{t=1}^{n}\{(\rho(z_t,\hat{\beta})\otimes \rho(z_t,\hat{\beta})) - \hat{E}[(\rho(z,\hat{\beta})\otimes \rho(z,\hat{\beta}))\cdot\rho(z,\hat{\beta})'|x_t]\widetilde{\Sigma}_t^{-1}\rho(z_t,\hat{\beta})] \tag{3.12}$$

where $\hat{E}[\cdot|x]$ indicates an appropriate conditional expectation estimator such as the kernel. If we define

$$\widetilde{\Sigma}_t = \widetilde{D}_t = \hat{E}[\rho(z,\hat{\beta})\cdot\rho(z,\hat{\beta})'|x_t] \tag{3.13}$$

as the corresponding estimator of the conditional covariance some algebra yields

$$0 = n^{-1}\sum_{t=1}^{n}\hat{E}[R(z_t,\hat{\beta})'\widetilde{\Sigma}_t^{-1}\rho(z_t,\hat{\beta})\cdot\rho(z_t,\hat{\beta})'|x_t]\widetilde{\Sigma}_t^{-1}\rho(z_t,\hat{\beta}) + O_p(\hat{D}-\widetilde{D}). \tag{3.14}$$

We again take the first term as our estimator since the last will be $o_p(n^{-1/2})$ under the local behavior of the next section.

Not too surprisingly, this estimator is closely related to an estimator based on the suggestions of Chamberlain. As in the previous case, we do not propose the new estimator as a serious alternative so much as proof that the basic approach advocated in this section generates reasonable estimators. Under apprpriate conditions, we find the estimator is consistent and has the following asymptotic behavior

$$n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{d} N[0, P^{-1}M\cdot P^{-1\prime}] \tag{3.15}$$

where now

$$\begin{aligned} P &= E[\overline{R}_t^{*\prime}\Sigma_t^{-1}R_t] \\ M &= E[\overline{R}_t^{*\prime}\Sigma_t^{-1}\overline{R}_t^{*}] \end{aligned} \tag{3.16}$$

and for convenience $R_t = R(z_t, \beta_0)$ and $\overline{R}_t^* = E[R(z,\beta_0)'\Sigma_t^{-1}\rho(z,\beta_0)\cdot\rho(z,\beta_0)'|x_t]$. This estimator will be asymptotically equivalent to that proposed by Chamberlain if $\overline{R}_t^* = E[R(z,\beta_0)|x_t]$. Since the resultant instruments are again exogenous we could use a preliminary estimator for β in the instruments.

In each of the previous two cases, the approach advocated in this section yielded interesting instruments but they were exogenous rather than endogenous. That is, purging the regressors of that component which is correlated with the disturbances also purged them completely of any stochastic relationship with the disturbances. Strengthening our stochastic assumption to require that the disturbances be independent of x does, however, yield endogenous instruments. Following Brown and Newey for this case we find

$$\begin{aligned} \hat{c}_i &= n^{-1} \textstyle\sum_{t=1}^{n} n^{-1} \sum_{s=1}^{n} \{\rho(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}) \otimes r_i(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})\} \\ \hat{d} &= n^{-1} \textstyle\sum_{t=1}^{n} \{\rho(z_t, \hat{\beta}) \otimes \rho(z_t, \hat{\beta})\} \end{aligned} \tag{3.17}$$

are the efficient estimators of c_i and d. Substituting into (3.8) and rearranging, we obtain

$$0 = n^{-1} \sum_{t=1}^{n} [R(z_t, \hat{\beta}) - n^{-1} \sum_{s=1}^{n} \{R(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})\}]' \widetilde{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \tag{3.18}$$

as our feasible estimating equation. The instruments generated by the approach of this section are clearly endogenous for this case, even in large samples.

This estimator has a close relationship, which will be revealed in the next section, to an estimator proposed by Newey [1989]. In particular, the presence of the double summation means that the estimator is an example of a v-statistic as studied by Serfling [1980]. Given regularity conditions of the type utilized there, we find that the estimator generated by (3.18) is consistent with the following asymptotic distribution:

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N[0, P^{-1} M \cdot P^{-1\prime}] \tag{3.19}$$

where

$$\begin{aligned} \psi(z, \beta) &= \{R(z, \beta) - E[R(\pi(\rho(z, \beta), x, \beta), x, \beta) | (\rho(z, \beta)]\}' \Sigma_0^{-1} \rho(z, \beta) \\ P &= E[\partial \psi(z, \beta_0) / \partial \beta'] \\ M &= E[\psi(z, \beta_0) \cdot \psi(z, \beta_0)']. \end{aligned} \tag{3.20}$$

This estimator bears an interesting relationship to the BNL3S estimator proposed by Amemiya [1977]. There the target instruments are $\Sigma_0^{-1} E[R(z, \beta_0) | x]$ whereas in the current estimator they are $\Sigma_0^{-1} \{R(z, \beta_0) - E[R(z, \beta_0) | \epsilon]\}$. Although it does not appear possible, in general, to rank these alternatives according to asymptotic efficiency, we should be able to easily combine the two to obtain an estimator that dominates both, in general.[4]

An alternative to independence, in strengthening the assumption of zero conditional mean, is to impose conditional symmetry of the distribution of the disturbances about zero, given x. According

[4] We simply form a GMM estimation problem by combining the moment condition underlying (3.19), namely $E\{(R(z, \beta_0) - E[R(z, \beta_0) | \epsilon])' \Sigma_0^{-1} \rho(z, \beta_0)\} = 0$ with the moment condition underlying BNL3S, namely, $E\{E[R(z, \beta_0) | x]' \Sigma_0^{-1} \rho(z, \beta_0)\} = 0$ to obtain an overidentified system of $2p$ momemt conditions in the p unknown β. We estimate Σ_0 and $E[R(z, \beta_0) | \epsilon]$ as in the text above and proceed with GMM using an estimated optimal weight matrix. This estimator should dominate both of the just-identified estimators except in the cases, given in Section 5, where the endogenous IV approach is locally optimal.

to Brown and Newey, the estimator of $D = \Sigma_0$ is the same as the previous case, while

$$\hat{c}_i = n^{-1}\sum_{t=1}^{n}\{\rho(z_t,\hat{\beta})\otimes r_i(z_t,\hat{\beta}))+\rho(\pi(-\rho(z_t,\hat{\beta}),x_t,\hat{\beta}),x_t,\hat{\beta})\otimes r_i(\pi(-\rho(z_t,\hat{\beta}),x_t,\hat{\beta}),x_t,\hat{\beta})\}/2 \tag{3.21}$$

is the efficient estimator of c_i. Substituting and rearranging as before yields

$$0 = n^{-1}\sum_{t=1}^{n}[R(z_t,\hat{\beta}) + \{R(-\pi(\rho(z_t,\hat{\beta}),x_t,\hat{\beta}),x_t,\hat{\beta})\}]'\widetilde{\Sigma}^{-1}\rho(z_t,\hat{\beta}) \tag{3.22}$$

as the estimating equation. The instruments are clearly endogenous in this case and have an interpretation as the even component, in terms of the disturbances, of the regressors. Under general conditions the resulting estimator is consistent with limiting distribution given by (3.19) and (3.20), except now

$$\psi(z,\beta) = [R(z,\beta) + R(-\pi(\rho(z,\beta),x,\beta),x,\beta)]'\Sigma_0^{-1}\rho(z,\beta). \tag{3.23}$$

The relationship of this estimator to an estimator proposed by Newey for this case will be explicated in the next section.

Combining the previous two cases, we now entertain the case where the disturbances are symmetric about zero and independent of x. Following the general approach proposed by Brown and Newey, the efficient estimator of c_i is

$$\begin{aligned}\hat{c}_i \;=\; & n^{-1}\textstyle\sum_{t=1}^{n}\{\rho(z_t,\hat{\beta})\otimes r_i(z_t,\hat{\beta})) \\ & -n^{-1}\textstyle\sum_{s=1}^{n}[\rho(\pi(\rho(z_t,\hat{\beta}),x_s,\hat{\beta}),x_s,\hat{\beta})\otimes r_i(\pi(\rho(z_t,\hat{\beta}),x_s,\hat{\beta}),x_s,\hat{\beta}) \\ & -\rho(\pi(-\rho(z_t,\hat{\beta}),x_s,\hat{\beta}),x_s,\hat{\beta})\otimes r_i(\pi(-\rho(z_t,\hat{\beta}),x_s,\hat{\beta}),x_s,\hat{\beta})]/2\}\end{aligned} \tag{3.24}$$

while the efficient estimator of Σ_0 is the same as the last two cases. Substituting and rearranging yields the following estimation equation

$$\begin{aligned}0 \;=\; & n^{-1}\textstyle\sum_{t=1}^{n}[R(z_t,\hat{\beta}) - n^{-1}\sum_{s=1}^{n}\{R(\pi(\rho(z_s,\hat{\beta}),x_s,\hat{\beta}),x_s,\hat{\beta}) \\ & -R(-\pi(\rho(z_s,\hat{\beta}),x_s,\hat{\beta}),x,\hat{\beta})\}/2]'\widetilde{\Sigma}^{-1}\rho(z,\hat{\beta}).\end{aligned} \tag{3.25}$$

Under general conditions, this estimator will be consistent with asymptotic distribution also given by (3.19) and (3.20) except now

$$\begin{aligned}\psi(z,\beta) \;=\; & \{R(z,\beta) - E[R(\pi(\rho(z,\beta),x,\beta),x,\beta) \\ & -R(\pi(-\rho(z,\beta),x,\beta),x,\beta)|(\rho(z,\beta)]/2\}'\Sigma_0^{-1}\rho(z,\beta).\end{aligned} \tag{3.26}$$

If the assumed stochastic conditions of conditional symmetry and independence hold, this estimator should be efficient relative to both of the last two alternatives. Of course, this increased efficiency comes at the cost of increased risk of misspecification of the distribution and the accompanying inconsistency.

4. LOCALLY EFFICIENT IV ESTIMATION

The semiparametric estimators proposed in the previous section were designed to be optimal in terms of being as highly correlated with the regressors as possible. It is of interest to compare the new estimators with existing estimators that are known to be globally or locally semiparametric efficient. A globally semiparametric efficient estimator is a regular semiparametric estimator which attains the semiparametric efficiency bound regardless of the true distribution, provided the stochastic restriction is met. By locally efficient, I mean that the estimator is regular semiparametric and hence consistent against misspecification of the distribution provided the stochastic restriction is satisfied but will attain the semiparametric efficiency bound if the disturbances have a specified distribution such as the normal. In cases where globally efficient estimators are not feasible, such estimators are desirable since they are the best estimator under the specified distribution that remains consistent in the event the distribution is misspecified.

The semiparametric efficiency bound can be defined as the inverse covariance matrix of the efficient score. For a particular choice of distribution, the unrestricted score can be defined as the derivative of the log-likelihood with respect to the parameter vector β. For the problem at hand it has the form

$$s_\beta(z) = J_\beta(z) + R(z, \beta_0)' s_\epsilon(z) \tag{4.1}$$

where $J_\beta(z) = \partial |\det(\partial\rho(z, \beta_0)/\partial y')|/\partial\beta'$ and $s_\epsilon(z) = \partial ln(f(\epsilon, x))/\partial\epsilon$. The nonparametric tangent set $\mathcal{T}$ is defined as the space of p-dimensioned linear combinations of the derivatives of the loglikehood with respect to the nuisance parameters for all models satisfying the semiparametric assumption. The efficient score is obtained as the residual from the projection of the unrestricted score onto this space

$$s(z) = s_\beta(z) - Proj(s_\beta(z)|\mathcal{T}) \tag{4.2}$$

and the semiparametric efficiency bound is given by

$$V_\beta^* = (E[s(z) \cdot s(z)'])^{-1}. \tag{4.3}$$

See the survey by Newey [1990b] or Brown and Newey [1992] for details.

For the case where the disturbances are unrestricted except to have zero mean the efficient score can be shown to be

$$\begin{aligned} s(z) &= E[s_\beta(z) \cdot \rho(z, \beta_0)'] \cdot \Sigma_0^{-1} \rho(z, \beta_0), \\ &= -\overline{R}' \Sigma_0^{-1} \rho(z, \beta_0), \end{aligned} \tag{4.4}$$

since

$$\overline{R} = E[R(z, \beta_0)] = -E[\rho(z, \beta_0) \cdot s_\beta(z)'] \tag{4.5}$$

by the generalized information matrix equality. Thus the semiparametric efficiency bound for the problem is given by

$$V_\beta^* = (\overline{R}' \Sigma_0^{-1} \overline{R})^{-1}. \tag{4.6}$$

An IV estimator which globally attains this bound, under general conditions, regardless of the true distribution, is given by the solution to

$$0 = [n^{-1} \sum_{t=1}^{n} R(z_t, \widetilde{\beta})]' \widetilde{\Sigma}^{-1} [n^{-1} \sum_{t=1}^{n} \rho(z_t, \hat{\beta})] \tag{4.7}$$

where $\widetilde{\beta}$ and $\widetilde{\Sigma}$ are preliminary consistent estimates. The solution to these estimating equations is, of course, the GMM estimator and is the globally efficient IV estimator.

Under certain conditions, the IV estimator proposed in the previous section for this case is locally efficient. Suppose

$$E[\rho(z, \beta_0) \cdot J_\beta(z)'] = 0 \tag{4.8}$$

for all distributions satisfying the semiparametric assumption. Due to the unrestricted nature of these distributions, this condition will, in general, only be met if $J_\beta(z)$ is constant. Substituting (4.1) into (4.5) and imposing (4.8) yields

$$\overline{R} = -E[\rho(z, \beta_0) \cdot s_\epsilon(z)' R(z, \beta_0)]. \tag{4.9}$$

If $\epsilon \sim N(0, \Sigma)$ independent of x is the true distribution, then we further have

$$\overline{R} = -E[\rho(z, \beta_0) \cdot \rho(z, \beta_0)' \Sigma_0^{-1} R(z, \beta_0)] = -\overline{R}* \tag{4.10}$$

which, from (3.10), means that the simple IV estimator is asymptotically equivalent to GMM under normality and the condition on the Jacobian derivative. But this means it is locally semiparametric efficient. Note that since ϵ is symmetrically distributed under the normality assumption, the condition $\hat{D} - \widetilde{D} = o_p(n^{-1/2})$ will be met.

A similar analysis applies for the case where the disturbances are only restricted to have conditional mean zero, given x. This problem has been well studied by Chamberlain and the efficient score is known to be

$$\begin{aligned} s(z) &= E[s_\beta(z) \cdot \rho(z, \beta_0)' | x_t] \Sigma_t^{-1} \rho(z, \beta_0) \\ &= -\overline{R}_t' \Sigma_t^{-1} \rho(z, \beta_0), \end{aligned} \tag{4.11}$$

since

$$\overline{R}_t = E[R(z, \beta_0) | x_t] = -E[\rho(z, \beta_0) \cdot s_\beta(z)' | x_t]. \tag{4.12}$$

by the generalized information matrix inequality. The semiparametric efficiency bound is therefore given by

$$V_\beta^* = (E[\overline{R}_t' \Sigma_t^{-1} \overline{R}_t | x_t])^{-1}. \tag{4.13}$$

An IV estimator which globally attains this bound, under sufficiently restrictive conditions, is given as the solution to

$$0 = n^{-1} \sum_{t=1}^{n} \hat{E}[R(z_t, \widetilde{\beta}) | x_t]' \widetilde{\Sigma}_t^{-1} \rho(z_t, \hat{\beta}), \tag{4.14}$$

where $\hat{E}[R(z_t, \widetilde{\beta}) | x_t]$ and $\widetilde{\Sigma}_t$ are estimators of the target conditional expectations. An example of such an estimator is Robinson's [1987] estimator for the linear regression model with unspecified heteroskedasticity in which $\hat{E}[R(z_t, \widetilde{\beta})|x_t]$ is given directly by the regressors x_t and $\widetilde{\Sigma}_t$ is a scalar and is estimated using nearest neighbor techniques.

Again, the simple IV estimator proposed for this case can be shown to attain the semiparametric efficiency bound under certain conditions. Suppose

$$E[\rho(z, \beta_0) \cdot J_\beta(z)'|x_t] = 0, \tag{4.15}$$

which for the distribution of $\rho(z, \beta_0)$ unrestricted except to have mean zero implies $J_\beta(z)$ is a function only of x. Substituting (4.1) into (4.12) and imposing (4.15) yields

$$\overline{R}_t = -E[\rho(z, \beta_0) \cdot s_\epsilon(z)' R(z, \beta_0)|x_t] \tag{4.16}$$

which becomes

$$\overline{R}_t = -E[\rho(z, \beta_0) \cdot \rho(z, \beta_0)' \Sigma_t^{-1} R(z, \beta_0)|x_t] = -\overline{R}_t^* \tag{4.17}$$

if $\epsilon|x_t \sim N(0, \Sigma_t)$ is the true distribution. But, according to (3.16) and (3.17), this means that the simple IV estimator attains the lower bound and is hence locally efficient when the disturbances are conditionally normal and the Jacobian condition (4.15) is satisfied.

The case where the disturbances are unrestricted except to be independent of x has been extensively studied by Newey [1989], who has shown that the efficient score for the problem takes the form

$$\begin{aligned} s(z) &= s_\beta(z) - E[s_\beta(z)|\rho(z, \beta_0)] \\ &= J_\beta(z) - E[J_\beta(z)|\rho(z, \beta_0)] + \{R(z, \beta_0) - E[R(z, \beta_0)|\rho(z, \beta_0)]\}' s_\epsilon(z). \end{aligned} \tag{4.18}$$

The semiparametric efficiency bound is given as the inverse covariance of this score, but is not particularly informative so is not reported here. If the true distribution of the disturbances is joint normality, he shows that the estimator given by the solution to the following equation

$$\begin{aligned} 0 &= n^{-1} \textstyle\sum_{t=1}^n [J_\beta(z_t, \hat{\beta}) - n^{-1} \sum_{s=1}^n J_\beta(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})] \\ &\quad + [R(z_t, \hat{\beta}) - n^{-1} \textstyle\sum_{s=1}^n R(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})]' \widetilde{\Sigma}^{-1} \rho(z_t, \hat{\beta}) \end{aligned} \tag{4.19}$$

attains the bound. Using the definition of $\widetilde{\Sigma}$ and rearranging terms this estimation equation can be rewritten

$$\begin{aligned} 0 &= n^{-1} \textstyle\sum_{t=1}^n [\{R(z_t, \hat{\beta}) - n^{-1} \sum_{s=1}^n R(\pi(\rho(z_t, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})\}' \\ &\quad + n^{-1} \textstyle\sum_{\tau=1}^n \{J_\beta(z_\tau, \hat{\beta}) - n^{-1} \sum_{s=1}^n J_\beta(\pi(\rho(z_\tau, \hat{\beta}), x_s, \hat{\beta}), x_s, \hat{\beta})\} \cdot \rho(z_t, \hat{\beta})'] \widetilde{\Sigma}^{-1} \rho(z_t, \hat{\beta}). \end{aligned} \tag{4.20}$$

which is an IV estimator with endogenous instruments. This estimator differs from the simple IV estimator proposed in the previous section by the presence of the Jacobian derivative terms. The two will be the same and hence the simple estimator will be semiparametric efficient if $J_\beta(z) = E[J_\beta(z)|\rho(z, \beta_0)]$, which will occur if $J_\beta(z)$ is a function only of ϵ or x.

The case of disturbances with a conditionally symmetric distribution has also been examined by Newey, who has obtained the following efficient score for the problem

$$\begin{aligned} s(z) &= s_\beta(z) - s_\beta(\pi(-\rho(z,\beta_0),x,\beta_0),x,\beta_0) \\ &= J_\beta(z) + R(z,\beta_0)' s_\epsilon(z) \\ &\quad -[J_\beta(\pi(-\rho(z,\beta_0),x,\beta_0),x) + R(\pi(-\rho(z,\beta_0),x,\beta_0)' s_\epsilon(\pi(-\rho(z,\beta_0),x)]. \end{aligned} \tag{4.21}$$

Under general conditions, Newey has shown that the estimator given by the solution to

$$\begin{aligned} 0 &= n^{-1}\textstyle\sum_{t=1}^n [J_\beta(z_t,\hat\beta) - J_\beta(\pi(-\rho(z_t,\hat\beta),x_t,\hat\beta),x_t,\hat\beta)] \\ &\quad +[R(z_t,\hat\beta) + R(-\pi(\rho(z_t,\hat\beta),x_t,\hat\beta),x_t,\hat\beta)]'\widetilde{\Sigma}^{-1}\rho(z_t,\hat\beta) \end{aligned} \tag{4.22}$$

is regular semiparametric and attains the efficiency bound when the true distribution is normal. As in the previous case, we can exploit the definition of $\widetilde{\Sigma}$ to rewrite this estimating equation in the form

$$\begin{aligned} 0 &= n^{-1}\textstyle\sum_{t=1}^n \{[R(z_t,\hat\beta) + R(-\pi(\rho(z_t,\hat\beta),x_t,\hat\beta),x_t,\hat\beta)]' \\ &\quad +n^{-1}\textstyle\sum_{\tau=1}^n [J_\beta(z_\tau,\hat\beta) - J_\beta(\pi(-\rho(z_\tau,\hat\beta),x_\tau,\hat\beta),x_\tau,\hat\beta)]\cdot\rho(z_t,\hat\beta)'\}\widetilde{\Sigma}^{-1}\rho(z_t,\hat\beta). \end{aligned} \tag{4.23}$$

which is an IV estimator with endogenous instruments. If $J_\beta(z) = J_\beta(\pi(-\rho(z,\beta_0),x,\beta_0),x,\beta_0)$ or $J_\beta(z)$ is symmetric in ϵ, this estimator reduces to the simple IV estimator proposed in the previous section. Thus under normality and the Jacobian restriction the estimator is locally semi-parametric efficient.

The findings for the previous two cases can be combined for the case where the disturbances are independent of x and symmetric around zero. The efficient score for this case is

$$s(z) = s_\beta(z) - E[s_\beta(\pi(\rho(z,\beta_0),x,\beta_0),x,\beta_0) + s_\beta(\pi(-\rho(z,\beta_0),x,\beta_0),x,\beta_0)|\rho(z,\beta_0)]/2. \tag{4.24}$$

If the true distribution of the disturbances is normal, the estimator given by the solution to the following equation

$$\begin{aligned} 0 &= n^{-1}\textstyle\sum_{t=1}^n [J_\beta(z_t,\hat\beta) - n^{-1}\sum_{s=1}^n \{J_\beta(\pi(\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta) - J_\beta(\pi(-\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta)\}] \\ &\quad +[R(z_t,\hat\beta) - n^{-1}\textstyle\sum_{s=1}^n \{R(\pi(\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta) - R(\pi(-\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta)\}]'\widetilde{\Sigma}^{-1}\rho(z_t,\hat\beta) \end{aligned} \tag{4.25}$$

attains the lower bound defined by the inverse covariance of the efficient score. This estimator can also be rearranged using the definition of $\widetilde{\Sigma}$ to obtain

$$\begin{aligned} 0 &= n^{-1}\textstyle\sum_{t=1}^n [\{R(z_t,\hat\beta) - n^{-1}\sum_{s=1}^n R(\pi(\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta) - R(\pi(-\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta)\}' \\ &\quad +n^{-1}\textstyle\sum_{\tau=1}^n \{J_\beta(z_\tau,\hat\beta) - n^{-1}\sum_{s=1}^n J_\beta(\pi(\rho(z_\tau,\hat\beta),x_s,\hat\beta),x_s,\hat\beta) \\ &\quad -J_\beta(\pi(-\rho(z_t,\hat\beta),x_s,\hat\beta),x_s,\hat\beta)\}\cdot\rho(z_t,\hat\beta)']\hat{\Sigma}^{-1}\rho(z_t,\hat\beta) \end{aligned} \tag{4.26}$$

which is an IV estimator with endogenous instruments. This estimator will reduce to the simple IV estimator of the previous section if

$$J_\beta(z) = E[J_\beta(\pi(\rho(z, \beta_0), x, \beta_0), x, \beta_0) + J_\beta(\pi(-\rho(z, \beta_0), x, \beta_0), x, \beta_0)|\rho(z, \beta_0)]/2 \quad (4.27)$$

which will occur if the Jacobian derivative is a function only of x or ϵ only and is symmetric. And, once again, we find that the simple IV estimator is locally semiparametric efficient under normality if a condition on the Jacobian is met.

The simple endogenous IV estimators proposed in the previous section were designed to obtain the instruments which are most highly correlated with the regressors. This approach did not, however, result in estimators that achieve the semiparametric efficiency bound except under special conditions on the Jacobian[5]. Nor were they the best endogenous IV estimators, since the optimal estimator in each case had an IV interpretation. This nonoptimality resulted most obviously, from studying the efficient scores for the various cases, because the impact of the Jacobian was ignored. More to the point, though, it results because our criterion for choosing endogenous instruments aimed for the wrong target. We were maximizing the correlation with the regressors, which guaranteed a reduction in the limiting covariance matrix by minimizing P^{-1}, which occurs twice, but we ignored the remaining component M. An interesting finding is the rather close relationship between IV estimation and normality. Even when the appropriate condition on the Jacobian was met, the optimal IV estimator locally attained the semiparametric efficiency bound only under normality.

5. A Sampling Experiment

The optimality claims for the estimators proposed in the previous sections are based on large-sample asymptotic analysis. It is of interest to quantify the magnitude of the improvement in precision resulting from utilizing the asymptotically efficient estimators rather than more standard approaches. It is also useful to measure the magnitude of the deterioration of the various estimators when the distribution is misspecified relative to the conditions for which it was designed. In addition, it is of interest to determine the extent to which the asymptotic results manifest themselves in smaller samples. These issues are best addressed, if only partially, by conducting a sampling experiment on specific models. Of course, the results of the experiment are subject to the usual qualifications and can be generalized to other models only with great care.

The sampling experiment is conducted on a single-equation extended Box-Cox (EBC) model. This model can be formally represented

$$(\mathrm{sgn}(y)|y|^{\beta_3} - 1)/\beta_3 - \beta_1 - \beta_2 \cdot x = \epsilon \quad (5.1)$$

where $E[\epsilon^2] = \sigma^2$. In the independent case the parameter β_1 will be subsumed into the distribution of the disturbances. This model has several salient features. The first is the fact that it is nonlinear

[5]One obvious case where the conditions on the Jacobian derivative are met for all the stochatic alternatives is, following an appropriate normalization, a triangular model. The conditions, however, can be satisfied by much more general models including the following two-equation nontriangular example

$$y_1 = \beta_1 + \beta_2 y_2 + \beta_3 x + \epsilon_1$$
$$\ln(y_2) = \beta_4 + \beta_5 y_1 + \beta_6 x + \epsilon_2.$$

The determinate of the Jacobian for this model is $1/y_2 - \beta_2\beta_5$, which has a constant derivative with respect to β.

in the variables and parameters. The second is that the solution form of the model can be written in the closed form

$$y = \text{sgn}(1 + \beta_3 \cdot (\beta_1 + \beta_2 \cdot x + \epsilon)) \cdot [1 + \beta_3 \cdot (\beta_1 + \beta_2 \cdot x + \epsilon)]^{1/\beta_3} \tag{5.2}$$

which greatly simplifies the generation of observations in the experiment. The third is that the conditions on the Jacobian derivatives discussed in the previous section are all met so the simple IV estimators are semiparametric efficient.

For each model of the distribution we conducted 1000 replications of a Monte Carlo experiment using alternative sample sizes of (25,50,100,200). The alternative models of the distribution of the disturbances used were:

Basic: $\epsilon \sim N(0, \sigma^2)$ independent of x,
Nonnormality: $\epsilon \sim t(4) \cdot (\sigma/\sqrt{2})$ independent of x
Asymmetry: $\epsilon \sim (\sqrt{6}/\pi) \cdot (0.57721 + \ln(-\ln(U(0,1))))$ independent of x,
Nonindependence: $\epsilon \sim N(0, \sigma^2) \cdot \{\gamma + (1-\gamma) \cdot \exp(x - E[x] - V[x]/2)\}^{1/2}$.

The values of the parameters used to generate the replications of the experiment were chosen to approximate the values obtained from estimating the model for the Engel curve data described in Koenker and Bassett [1982]. The specific values used were

$$(\beta_1, \beta_2, \beta_3, \sigma^2) = (-10.86, 3.73, 0.23, 0.359) \tag{5.3}$$

the values of the exogenous variables were drawn from $x \sim N(6.7, 0.191)$. For each replication, the model was estimated by maximum likelihood based on a normality assumption (MLE or QMLE), an IV estimator using $(1, (\beta_2 \cdot x)^2)$ as instruments (BIV), the residual-based feasible best nonlinear three-stage least squares (B3SR)[6], the optimal endogenous IV estimator designed for independence (BIND), the endogenous IV estimator that is optimal under symmetry (BSYM), and the endogenous IV estimator designed for both independence and symmetry (BCOM).

The results for the estimates of β_3 for the basic model are reported in Table 1. We report the median bias, median absolute error (MAE), mean bias, and root mean squared error (RMSE). For this stochastic model, all the estimators should be consistent and MLE should be efficient. The bias (median and mean) results seem to be of order $1/n$ and are roughly consistent with what we expect. The MAE and RMSE seem to be of order $1/\sqrt{n}$, which also is what we expect. Perhaps the most striking feature in the table is the strength, in terms of MAE and RMSE, of the BCOM estimator which seems to match the performance of the MLE on its home ground. There is evidence of fat-tailed behavior which could lead to nonexistence of moments for the BIND estimator, which is manifested in small samples by its relatively poor performance in terms of RMSE and mean bias but its strong performance in terms of median bias and MAE. In large samples, BIND seems to dominate both BIV and B3SR by all criteria.

The findings for the nonnormality case where the disturbances are drawn from a t distribution with 4 degrees of freedom are presented in Table 2. For this stochastic model, we would expect all the models to be consistent with the possible exception of maximum likelihood. It appears, however, the QMLE is consistent against this particular form of distributional misspecification. This is most evident in the bias numbers which exhibit much the same behavior for QMLE as in the normality case. The biggest change relative to the normality case is substantial improvement in the performance of the BIV and B3SR relative to the other estimators. This is probably a manifestation

[6] See Brown [1990] and Robinson [1989] for the form of the residual-based BNL3S estimator.

of the local nature of the optimality of the latter. They are designed to be efficient when the true distribution is normal but remain consistent in the event of nonnormality. They certainly exhibit this behavior, but this leaves room for the possibility that other estimators will do better under nonnormality. Here also there is evidence of fat-tailed behavior by the BIND estimator.

The results for the stochastic model with asymmetric disturbances are given in Table 3. For this stochastic model, the BIV, B3SR, and BIND estimators should be consistent, while the QMLE, BSYM, and BCOM estimators could conceivably be inconsistent. This expectation is certainly born out by the bias results, which clearly have the same sort of order as before for the former but are converging to positive nonzero values for the latter. In terms of MAE and RMSE the results are less clear-cut due to the usual mean-variance type trade-off. Specifically, in small samples the variances are large enough relative to the biases that the estimators that performed best under correct stochastic specification continue to do well. In large samples, however, the variances shrink and we are left with the nonzero bias for the inconsistent estimators. It is worth noting that BIND seems to exhibit better performance than B3SR and BIV for this case, particularly in large samples, which is to be expected.

The findings for the stochastic model with nonindependent or heteroskedastic disturbances are reported in Table 4. In this case we expect the BIV, B3SR, and BSYM estimators to be consistent but the QMLE, BIND, and BCOM estimators to be inconsistent. As with the previous case, the bias results clearly agree with this expectation with the bias advantage of the consistent estimators becoming apparent in moderate sized samples. Again there is the biasdispersion trade-off with QMLE and BCOM being preferred in small samples in terms of MAE and RMSE, due to their smaller variance, but BSYM, BIV and B3SR preferable in large samples, due to their consistency. The BSYM seems to enjoy an advantage over the other consistent estimators, as might be expected, but the advantage is very small.

Table 1.
EBC Model, β_3
Normal Disturbances

n	Estimator	Median Bias	MAE	Mean Bias	RMSE
25	MLE	-0.0193	0.1662	-0.0233	0.2791
	BIV	-0.0411	0.2211	-0.0523	0.3437
	B3SR	-0.0413	0.2170	-0.0832	0.3984
	BIND	-0.0631	0.1815	-0.1290	0.4627
	BSYM	-0.0473	0.2111	-0.0682	0.3572
	BCOM	-0.0191	0.1662	-0.0231	0.2798
50	MLE	-0.0105	0.1088	-0.0122	0.1764
	BIV	-0.0076	0.1420	-0.0146	0.2305
	B3SR	-0.0076	0.1403	-0.0197	0.2366
	BIND	-0.0198	0.1116	-0.0438	0.2881
	BSYM	-0.0031	0.1387	-0.0143	0.2243
	BCOM	-0.0103	0.1088	-0.0121	0.1765
100	MLE	0.0029	0.0752	-0.0001	0.1150
	BIV	0.0083	0.0970	0.0038	0.1516
	B3SR	0.0075	0.0975	0.0025	0.1498
	BIND	0.0000	0.0762	-0.0014	0.1209
	BSYM	0.0059	0.0911	0.0035	0.1431
	BCOM	0.0031	0.0747	0.0000	0.1150
200	MLE	-0.0034	0.0526	-0.0025	0.0781
	BIV	-0.0032	0.0686	0.0014	0.1035
	B3SR	-0.0033	0.0684	0.0000	0.1026
	BIND	-0.0039	0.0529	-0.0033	0.0795
	BSYM	-0.0003	0.0655	0.0020	0.1006
	BCOM	-0.0032	0.0525	-0.0025	0.0781

Table 2.
EBC Model, β_3
t(4) Disturbances

n	Estimator	Median Bias	MAE	Mean Bias	RMSE
25	QMLE	-0.0095	0.1787	-0.0086	0.3145
	BIV	-0.0241	0.2051	-0.0574	0.8336
	B3SR	-0.0277	0.2034	-0.1056	0.5597
	BIND	-0.0807	0.2101	-0.1725	0.4998
	BSYM	-0.0454	0.2085	-0.0860	0.3907
	BCOM	-0.0186	0.1774	-0.0356	0.3294
50	QMLE	-0.0049	0.1177	-0.0101	0.2065
	BIV	-0.0069	0.1353	-0.0260	0.2252
	B3SR	-0.0081	0.1359	-0.0364	0.2553
	BIND	-0.0507	0.1373	-0.1512	0.4161
	BSYM	-0.0113	0.1399	-0.0382	0.2697
	BCOM	-0.0089	0.1159	-0.0269	0.2208
100	QMLE	-0.0063	0.0922	-0.0047	0.1634
	BIV	-0.0036	0.0874	-0.0067	0.1624
	B3SR	-0.0039	0.0883	-0.0119	0.1856
	BIND	-0.0274	0.0995	-0.1212	0.3732
	BSYM	-0.0044	0.1058	-0.0155	0.2141
	BCOM	-0.0108	0.0915	-0.0190	0.2099
200	QMLE	0.0006	0.0689	0.0018	0.1208
	BIV	0.0033	0.0650	0.0009	0.1043
	B3SR	0.0030	0.0625	-0.0002	0.1064
	BIND	-0.0172	0.0699	-0.0972	0.3581
	BSYM	-0.0014	0.0871	-0.0092	0.1954
	BCOM	-0.0063	0.0691	-0.0163	0.1738

Table 3.
EBC Model, β_3
Asymmetric Disturbances

n	Estimator	Median Bias	MAE	Mean Bias	RMSE
25	QMLE	0.0577	0.1839	0.0735	0.3181
	BIV	-0.0593	0.2271	-0.0479	0.3496
	B3SR	-0.0622	0.2232	-0.0813	0.3982
	BIND	-0.0498	0.1882	-0.0511	0.3331
	BSYM	0.0244	0.2445	0.0075	0.4390
	BCOM	0.0435	0.1791	0.0376	0.3215
50	QMLE	0.0629	0.1240	0.0740	0.2053
	BIV	-0.0213	0.1460	-0.0179	0.2303
	B3SR	-0.0238	0.1438	-0.0296	0.2498
	BIND	-0.0226	0.1359	-0.0293	0.2138
	BSYM	0.1014	0.1609	0.0954	0.3038
	BCOM	0.0576	0.1226	0.0576	0.2140
100	QMLE	0.0832	0.1071	0.0883	0.1588
	BIV	-0.0092	0.0923	0.0019	0.1547
	B3SR	-0.0101	0.0911	-0.0036	0.1663
	BIND	-0.0004	0.0866	-0.0089	0.1396
	BSYM	0.1261	0.1387	0.1360	0.2283
	BCOM	0.0799	0.1061	0.0823	0.1614
200	QMLE	0.0867	0.0892	0.0900	0.1247
	BIV	0.0006	0.0634	-0.0018	0.0965
	B3SR	-0.0007	0.0625	-0.0019	0.0983
	BIND	-0.0028	0.0603	-0.0042	0.0907
	BSYM	0.1252	0.1263	0.1356	0.1716
	BCOM	0.0835	0.0871	0.0860	0.1245

Table 4.
EBC Model, β_3
Heteroskedastic Disturbances

n	Estimator	Median Bias	MAE	Mean Bias	RMSE
25	QMLE	-0.1015	0.1820	-0.1106	0.2983
	BIV	-0.0826	0.2331	-0.0843	0.3502
	B3SR	-0.0793	0.2286	-0.1217	0.4166
	BIND	-0.1324	0.1985	-0.1730	0.4084
	BSYM	-0.0644	0.2400	-0.0955	0.3833
	BCOM	-0.1016	0.1818	-0.1104	0.2993
50	QMLE	-0.0935	0.1318	-0.0936	0.2020
	BIV	-0.0114	0.1580	-0.0228	0.2560
	B3SR	-0.0104	0.1562	-0.0383	0.2887
	BIND	-0.1037	0.1360	-0.1198	0.2631
	BSYM	-0.0089	0.1478	-0.0296	0.2671
	BCOM	-0.0937	0.1322	-0.0935	0.2020
100	QMLE	-0.0950	0.1057	-0.0914	0.1529
	BIV	0.0058	0.1060	0.0115	0.1692
	B3SR	0.0055	0.1030	0.0063	0.1691
	BIND	-0.0986	0.1098	-0.0958	0.1592
	BSYM	0.0009	0.1024	0.0067	0.1623
	BCOM	-0.0951	0.1058	-0.0913	0.1529
200	QMLE	-0.0984	0.0996	-0.0949	0.1251
	BIV	-0.0092	0.0687	0.0034	0.1110
	B3SR	-0.0082	0.0684	0.0047	0.1084
	BIND	-0.0988	0.1006	-0.0990	0.1294
	BSYM	-0.0060	0.0666	0.0044	0.1049
	BCOM	-0.0984	0.0996	-0.0949	0.1251

6. CONCLUDING REMARKS

The existing literature on instrumental variables (IV) estimators has concentrated, almost exclusively, on the use of instruments that are functions only of exogenous (predetermined) variables. Under a wide variety of assumptions, this approach yields instruments that satisfy the primary objective of IV estimators, namely, that the instruments be uncorrelated with the disturbances. A secondary objective, for efficiency reasons, is to choose instruments that are as highly correlated with the regressors as possible. In linear models and certain nonlinear models, specific exogenous instruments can also satisfy the second objective. In other nonlinear models, however, it is possible that nonlinear functions of the endogenous variables are uncorrelated with the disturbances and hence improved instruments may be obtained by using instruments that are functions of both endogenous and exogenous (predetermined) variables. The purpose of this paper is to systematically examine this possibility within the context of the static nonlinear simultaneous equation model under nonparametric distributional assumptions.

The basic approach used in this paper is to use an estimate of the residual from the linear projection of the regressors onto the disturbances as the instruments. A problem with this approach is that using the obvious estimators of the components, which are expectation functions, in the projection yields a trivial estimator. The solution is to use the semiparametric efficient estimators of the expectation functions proposed in Brown and Newey [1992] in forming the instruments. The form of the estimates depends on the stochastic assumption regarding the degree of dependence between the disturbances and the exogenous variables. For an assumption of either unconditional or conditional mean zero the resulting estimator turns out to be a function only of exogenous variables. For an assumption of independence the resulting instruments are the regressors less an estimate of the expectation of the instruments conditioned on the disturbances. Under an assumption of conditional symmetry, the resulting instruments are an estimate of the even component, in terms of the disturbances, of the regressors. For combined independence and symmetry, the resulting instruments are the regressors less an estimate of the expectation of their odd component conditioned on the disturbances.

The new estimators are compared, in terms of asymptotic behavior, to various semiparametric estimators that are known to be either globally or locally semiparametric efficient. An estimator is locally semiparametric efficient under a particular distribution if it is semiparametric and hence consistent under the general stochastic assumption, which allows a variety of distributions, but achieves the semiparametric efficiency bound for the problem when the true distribution has the specified form. Obviously, an estimator that is globally semiparametric efficient is locally efficient for any distribution that satisfies the stochastic assumptions. Under an assumption that the true distribution of the disturbances is normal and a restriction on the derivative of the Jacobian determinant with respect to the coefficients, each of the new estimators is found to be locally semiparametric efficient under the stochastic assumption for which it was designed. The restrictions on the derivative of the Jacobian determinant range from being constant to being a function only of the disturbances or exogenous variables.

The finite sample performance of the new estimators is examined in a sampling experiment on a single-equation Box-Cox model. This model is chosen for its obvious simplicity and because it satisfies the conditions on the Jacobian derivative and hence the new estimators will be locally semiparametric efficient. The results of the experiment are consistent with our expectations with a couple of exceptions. First, the new estimator based on a combined assumption of independence and symmetry performed at least as well as the maximum likelihood estimator (MLE) even when the true distribution is normal, whereupon the MLE should be parametrically efficient. Second, the maximum likelihood estimator based on normality seemed to be consistent against some forms of nonnormality of the true distribution. The usual caveats regarding the generality of the findings of the sampling experiment beyond the specific model used apply.

The approach and estimators advocated in this paper can be extended in several directions. If the disturbances are assumed to be intertemporally independent, then the estimators of this paper should apply pretty much as is with predetermined variables replacing exogenous variables. Under more general assumptions regarding the degree of intertemporal dependence the general approach of using endogenous instruments should still have value but the appropriate form will likely be different. Likewise in the context of latent variables models the specific estimators proposed here will not be appropriate but the general approach should prove valuable in developing IV estimators with improved efficiency. Finally, the interpretation of the general approach as the SUR estimator with a correction term for the fact that it has nonzero expectation at the truth can be utilize to develop a general class of locally semiparametric efficient estimators. Specifically, the scores for a particular choice of distribution can be corrected, in a similar fashion, for the fact that they have nonzero expectation at the true parameters when the distribution is misspecified.

7. References

AMEMIYA, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultneous Equation Model," *Econometrica* 45, 955-968.

ANDREWS, D. (1990a), "Asymptotics for Semiparametric Econometric Models: I. Estimation," mimeo, Cowles Foundation, Yale University.

ANDREWS, D. (1990b), "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity," mimeo, Cowles Foundation, Yale University.

BREUSCH, T., G. MIZON, AND P. SCHMIDT (1989), "Efficient Estimation Using Panel Data," *Econometrica* 57, 695-700.

BROWN, B. (1990), "Simulation-Based Semiparametric Estimation and Prediction in Nonlinear Systems," mimeo, Department of Economics, Rice University.

BROWN B.W. AND R.S. MARIANO (l989b), "Predictors in Dynamic Nonlinear Models: Large Sample Behavior," *Econometric Theory* 5, 430-452.

BROWN, B. AND W. NEWEY (1992), "Efficient Semiparametric Estimation of Expectations," mimeo, Department of Economics, Rice Univeersity.

CHAMBERLAIN, G. (1986), "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics* 32, 189-218.

HANSEN, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1055.

KOENKER, R. AND G. BASSETT (1982), "Robust Tests for Heteroskedasticity Based on Regression Quantiles," *Econometrica* 50, 46-31.

KELEJIAN, H.H. (1974), "Efficient Instrumental Variables Estimation of Large-Scale Nonlinear Econometric Models," mimeo.

NEWEY, W. (1989), “Locally Efficient, Residual-Based Estimation of Nonlinear Simultaneous Equations,” mimeo, Department of Economics, Princeton University.

NEWEY, W. (1990a), “Efficient Instrumental Variables Estimation of Nonlinear Models,” *Econometrica* 58, 809-838.

NEWEY, W. (1990b), “Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics* 5, 99-135.

NEWEY, W. (1992), “The Asymptotic Variance of Semiparametric Estimators,” mimeo, Department of Economics, MIT.

NEWEY, W. AND D. MCFADDEN (1992), “Estimation in Large Samples,” mimeo, Department of Economics, MIT.

PAGAN, A.R. AND A. ULLAH (1988), “The Econometric Analysis of Models with Risk Terms,” *Journal of Applied Econometrics* 3, 87-105.

RILSTONE, P. (1989), “Semiparametric Instrumental Variables Estimation,” forthcoming, *Journal of Econometrics*.

ROBINSON, P. (1987), “Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form,” *Econometrica* 55, 875-891.

ROBINSON, P. (1988), “Semiparametric Econometric: A Survey,” *Journal of Applied Econometrics* 3, 35-51.

ROBINSON, P. (1991), “Best Nonlinear Three Stage Least Squares Estimation of Certain Econometric Models,” *Econometrica* 59, 755-786.

SERFLING, R. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.