# GMM, Efficient Bootstrapping, and Improved Inference*

Bryan W. Brown
Department of Economics
Rice University

Whitney K. Newey
Department of Economics
M.I.T.

August, 2001

**Abstract**

GMM has been an important innovation in econometrics. Its usefulness has motivated a search for good inference procedures based on GMM. This paper presents a novel method of bootstrapping for GMM, based on resampling from the empirical likelihood distribution, that imposes the moment restrictions. We show that this approach yields a large sample improvement and is efficient, and give examples. We also discuss the development of GMM and other recent work on improved inference.

**JEL Classification:** C14, C30
**Keywords:** Bootstrapping, GMM, Panel Data, Empirical Likelihood.

*The work for this paper was partially completed while Newey was a fellow at the Center for Advanced Study in the Behavioral Sciences. The NSF also provided financial support. Excellent research assistance was provided by M. Busse and Sean May. Helpful comments were given by G. Chamberlain, J. Hahn, J. Hausman, J. Horowitz, G. Imbens and seminar participants at Harvard-MIT, Monash, Montreal, New York, and Rice Universities. The approach to efficient bootstrapping for GMM was presented at Monash in 1992 and the January 1994 meeting of the Econometric Society.

# 1 Introduction

Generalized method of moments estimation (GMM, Hansen, 1982) has been an important innovation in econometrics. It has provided a useful approach to solving estimation problems in many settings, including rational expectations models (Hansen and Singleton, 1982), models for panel data (Hausman and Taylor, 1981, Anderson and Hsiao, 1982, Holtz-Eakin, Newey, and Rosen, 1988, Abowd and Card, 1989; see Arellano and Honore, 2001, for further references), continuous time models (Hansen and Scheinkman, 1995), and semiparametric models (e.g. Powell, 1986). In terms of the analogy principle discussed by Manski (1988), it has proven to be a particularly powerful principle for solving problems. One strand of this literature has focused on finding moment conditions for improved efficiency (Cragg, 1983, and MaCurdy 2000), as well as conditions for asymptotic efficiency with many moments (Chamberlain, 1987, Newey, 1988, 1993, Newey and Chipty, 2000). GMM has also been used as a framework for developing tests about parameters of interest (Burguette, Gallant, and Sousa, 1982, Newey and West, 1987, Newey and McFadden, 1994) as well as specification tests (Newey, 1985a, 1985b, Tauchen, 1985, Eichenbaum and Hansen, 1990). Because so many estimators and tests can be viewed as being GMM, this framework provides a powerful unifying principle for econometrics (Burguette, Gallant, and Sousa, 1982, Newey and McFadden, 1994).

GMM was developed in several steps over a long time period. Each step in its development has been important, leading to many new results. The basic idea of choosing estimators to minimize a quadratic form in deviations between sample moments and population moments can be traced at least as far back as Pearson's (1900) minimum chi-square estimator. Chiang (1956) and Ferguson (1958) considered estimators that are obtained by minimizing a quadratic form in the difference of population and sample moments. These papers also allow for more general quadratic forms, involving functions of sample moments, that have led to important work in simultaneous equations (Malinvaud, 1970, Rothenberg, 1973), panel data (Chamberlain, 1982), and dynamic models (see also Gourieroux and Monfort, 1995). In a parallel development Sargan (1958, 1959) considered both linear and nonlinear instrumental variable (IV) estimators, for models with a residual that is linear in the variables. Amemiya (1974) developed the nonlinear two-stage least squares estimator for models with a nonlinear residual, covering many important additional cases, such as intertemporal capital asset pricing models (Hansen and Singleton, 1982). A third parallel development is that of Huber (1967), who considered the case with as many nonseparable moment conditions as parameters.

The full development of GMM in Hansen (1982) allowed for both nonseparable moment conditions and more moments than parameters. The GMM set up allows moment conditions that include a wide variety of functions, such as likelihood scores, products of residuals with instruments, differences between random variables and their expectations, and any combination of these. One paper that exemplifies the full power of GMM is Imbens (1992) GMM estimator for discrete choice models with choice based sampling, that attains Cosslett's (1981) efficiency bound. The moment conditions here are conditional scores and other kinds of functions.

Another important development brought about by Hansen (1982) and White (1982) is the use of weighting matrices for IV that are optimal (asymptotically efficient) under heteroskedasticity and/or autocorrelation. The use of such optimal weighting matrices in practice has now become very common.

The usefulness of GMM makes it imperative to have accurate inference procedures for GMM estimators. Even before GMM was introduced, well known results for its precursors, especially IV estimators for linear models, indicated that asymptotic approximations could be poor, e.g. see Anderson and Sawa (1979). More recent work, e.g. in the 1996 special issue of the Journal of Business and Economic Statistics, has shown that the asymptotic approximation for GMM can be bad. It is on the improvement of inference methods for GMM that we focus most of our discussion of econometric theory.

Bootstrapping provides one approach to improved inference. This paper presents a novel method of bootstrapping for GMM, based on resampling from the empirical likelihood (EL) distribution, that imposes the moment restrictions, rather than the empirical distribution. We show that this method yields an improvement to the usual asymptotic approximation in large samples. We also discuss computational simplicity relative to the bootstrap of Hall and Horowitz (1996) and show that the empirical likelihood bootstrap provides an estimator of the distribution of t-ratios and overidentification statistics that is asymptotically efficient.[1] We give Monte Carlo and empirical examples of this bootstrap. In addition to presenting this approach to GMM bootstrapping we also survey more recent developments on improved inference procedures for GMM.

In Section 2 we describe the EL bootstrap. Section 3 gives Edgeworth expansion arguments showing why moment restricted bootstrapping leads to improvement and the nature of the efficiency gain. Section 4 presents the examples. Section 5 reviews recent contributions to improved inference for GMM.

# 2 GMM Bootstrapping

We begin by introducing our notation for GMM. Let $z$ denote a data observation, $\beta$ a $p \times 1$ parameter vector, and $g(z, \beta)$ an $m \times 1$ vector of functions of a data observation and parameters, with $m \geq p$. GMM estimators are based on a moment restriction of the form

$$E[g(z, \beta_0)] = 0. \tag{2.1}$$

Often this moment restriction is a partial implication of some model, although it may also embody all the available information.

Although improved inference for GMM is important for time series, we only discuss bootstrapping in the i.i.d. setting. Let $z_1, ..., z_n$ denote the data observations, each satisfying equation (2.1). Let $g_i(\beta) = g(z_i, \beta)$ and $\hat{g}(\beta) = \sum_{i=1}^{n} g_i(\beta)/n$. For our purposes a GMM estimator will be one obtained as

$$\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{W} \hat{g}(\beta), \tag{2.2}$$

where $\mathcal{B}$ is the set of parameter values and $\hat{W}$ a positive semidefinite matrix. The estimator obtained when $\hat{W}$ is a consistent estimator of the inverse of $\Omega = Var(g_i(\beta_0))$

[1] The GMM bootstrapping method given here was originally proposed independently of and concurrently with Hall and Horowitz (1996) and prior to Hall and Presnell (1999), in Brown and Newey (1992)

is of particular interest, because this corresponds to an efficient choice of $\hat{W}$, as shown in Hansen (1982). To that end let $\hat{\Omega}(\beta) = \sum_{i=1}^{n} g_i(\beta)g_i(\beta)'/n$ and $\tilde{\Omega} = \hat{\Omega}(\tilde{\beta})$ for some preliminary GMM estimator $\tilde{\beta}$. Then the efficient, two-step GMM estimator is

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} \hat{g}(\beta)'\tilde{\Omega}^{-1}\hat{g}(\beta). \tag{2.3}$$

As shown by Hansen (1982), this estimator has the smallest asymptotic variance in the class obtained from equation (2.2). As shown by Chamberlain (1987), it also attains the semiparametric efficiency bound in the model where all that is known is the moment restrictions of equation (2.1). In this sense, it is asymptotically efficient among all estimators based solely on those moment restrictions.

Associated with the efficient GMM estimator are consistent estimators of its asymptotic variance and a test statistic of overidentifying restrictions. Let $\hat{\Omega} = \hat{\Omega}(\hat{\beta})$ and $\hat{G} = \partial\hat{g}(\hat{\beta})/\partial\beta$. A consistent estimator of the asymptotic variance is given by

$$\hat{V} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}.$$

From this consistent estimator an asymptotic $1-\alpha$ level confidence interval for $\beta_{0j}$ can be formed as $(\hat{\beta}_j - q_{\alpha/2}\sqrt{\hat{V}_{jj}/n}, \hat{\beta}_j + q_{\alpha/2}\sqrt{\hat{V}_{jj}/n})$, where $q_{\alpha/2}$ is the $1-\alpha/2$ quantile of the normal distribution. The overidentification test statistic is

$$T = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}).$$

Under equation (2.1) and additional regularity conditions, the asymptotic distribution of $T$ is $\chi^2(m-p)$ distribution.

There are numerous examples where it is known that the asymptotic theory for these confidence intervals and tests is a poor approximation to the finite sample distribution, e.g. see the 1996 special issue of the Journal of Business and Economic Statistics on GMM. The usefulness of GMM in so many settings, as discussed above, means that it is important to improve on the existing procedures. To help in the search for this improvement we include here an original contribution on bootstrapping for GMM. Although this material appeared in working paper form some time ago, we give it here for the first time in print.

Our bootstrap improvement for GMM can be described as some simple steps. It is based on approximating the distribution of the t-statistic or overidentification statistic by a bootstrap estimator. Let $\hat{\beta}$ be an efficient GMM estimator, like the two-step estimator discussed above, $\hat{g}_i = g(z_i, \hat{\beta})$, and for any possible value $b$ of the true parameter $\beta_0$, let $S(z_1, ..., z_n, b)$ be an object we want to estimate the distribution of, such as a t-ratio or overidentification statistic. The steps are

**GMM-EL BOOTSTRAPPING:**

1) Calculate

$$\hat{\pi}_i = \frac{1}{n(1-\hat{\lambda}'\hat{g}_i)}, (i = 1, ..., n), \hat{\lambda} = \arg \max_{\lambda'\hat{g}_i < 1} \sum_{i=1}^{n} \ln(1 - \lambda'\hat{g}_i). \tag{2.4}$$

2) Draw $n$ i.i.d. observations $z_1^b, ..., z_n^b$ with replacement from the distribution with $\Pr(z = z_i) = \hat{\pi}_i$.

3) Calculate $S^b = S(z_1^b, ..., z_n^b, \hat{\beta})$.

4) Repeat 2) and 3) $B$ times, where $B$ is an integer, to obtain $S^1, ..., S^B$.

5) Let the estimator of the distribution of $S(z_1, ..., z_n, \beta_0)$ be the discrete distribution with $\Pr(S(z_1, ..., z_n, \beta_0) = S^b) = 1/B$.

This estimator of the distribution of $S$ can then be used to form critical values for test statistics and confidence intervals in the usual way. Specifically, for a symmetric confidence interval for $\beta_{0j}$, let

$$S(z_1, ..., z_n, b) = |(\hat{\beta}_j - b_j)/\sqrt{\hat{V}_{jj}/n}|.$$

Then $S^b = |(\hat{\beta}_j^b - \hat{\beta}_j)/\sqrt{\hat{V}_{jj}^b/n}|$ would be obtained as in step 3), where $\hat{\beta}_j^b$ and $\hat{V}_{jj}^b$ are the GMM parameter and asymptotic variance estimators computed entirely from the simulated data $z_1^b, ..., z_n^b$. That is, from the simulated data we would compute in sequence each of the initial weighting matrix $\hat{W}$, the estimator $\tilde{\beta}$, the estimate of the second moment matrix $\tilde{\Omega}$, the efficient GMM estimator $\hat{\beta}$, and its asymptotic variance estimate $\hat{V}$. Let $\hat{q}_\alpha^B$ be a $1 - \alpha$ quantile of the distribution estimator from step 5). A bootstrap improved, symmetric confidence interval is then

$$\hat{\beta}_j \pm \hat{q}_\alpha^B \sqrt{\hat{V}_{jj}/n}.$$

Here we focus on a symmetric confidence interval because bootstrap theory suggests it is asymptotically more accurate than equal tailed intervals.

For the overidentification test statistic, we simply let $S(z_1, ..., z_n, b) = S(z_1, ..., z_n) = T$ be the test statistic. Here $S$ does not depend on $b$. Let $\hat{q}^B_\alpha$ here be the $1 - \alpha$ quantile of the distribution estimator from step 5). A test that rejects if

$$T \geq \hat{q}^B_\alpha$$

is a bootstrap overidentification test.

This bootstrap differs from standard approaches in the use of $\hat{\pi}_i$ in step 1) rather than $1/n$ for the probability of the $i^{th}$ observation. The form of $\hat{\pi}_i$ can be explained by an equivalent formula for $\hat{\pi}_i$, as the solution to

$$\max_{\pi_1,...,\pi_n} \sum_{i=1}^{n} \ln(\pi_i), s.t. \sum_{i=1}^{n} \pi_i \hat{g}_i = 0, \sum_{i=1}^{n} \pi_i = 1, \pi_i \geq 0. \tag{2.5}$$

This is the empirical likelihood (EL) of Owen (1988), and equivalence with the previous formula is shown in that paper. Thus, step 1) consists of calculating the empirical likelihood probabilities. The respective maximization problems in equations (2.4) and (2.5) are actually duals of each other, in the sense that the $\hat{\lambda}$ of step 1) is proportional to the Lagrange multipliers for the first (moment) constraint in equation (2.5), e.g. see Newey and Smith (2001). The step 1) maximization problem is considerably easier than this one, having much smaller dimension and being a simple concave programming problem.[2] Existence of $\hat{\lambda}$ and hence $\hat{\pi}_i$ requires that the constraint set is nonempty, i.e. that the intersection of the null space of the matrix $[\hat{g}_1, ..., \hat{g}_n]$ with the unit simplex is nonempty. In applications and Monte Carlo simulations, existence of $\hat{\pi}_i$ has rarely been a problem.

Step 1) corresponds to a cumulative distribution function (CDF) estimator of the form $\hat{F}(z) = \sum_{i=1}^{n} 1(z_i \leq z)\hat{\pi}_i$. This $\hat{F}$ has several important properties. First, by virtue of the moment constraints in equation (2.5) it satisfies the moment restrictions $\int g(z, \hat{\beta})\hat{F}(dz) = 0$. Second, it is asymptotically normal, as we show below. Third, it

[2] A short Gauss procedure for the computation of $\hat{\pi}_i$ and generation of the bootstrap data $z_i^b$ is available on request from the authors.

is asymptotically efficient, attaining the semiparametric efficiency bound of Brown and Newey (1998) for estimators of the CDF under the moment restrictions of equation (2.1). Each of these properties has important implications for the GMM bootstrap that we discuss below.

Other CDF estimators share these properties. One is the estimator of Back and Brown (1993), where $\hat{\pi}_i = (1 - \bar{g}'\hat{\Omega}^{-1}\hat{g}_i)/[n(1 - \bar{g}'\hat{\Omega}^{-1}\bar{g})]$ for $\bar{g} = \hat{g}(\hat{\beta})$. This estimator has a closed form and we have not found negative $\hat{\pi}_i$ to be a problem in practice, so that it could also be used in place of empirical likelihood in step 1) of the bootstrap. Also, the empirical likelihood and Back and Brown (1993) CDF estimators are special cases of a more general class, where $\tau(r)$ is a concave function of a scalar $r$ with domain $\mathcal{V}$, $\mathcal{V}$ is an open interval containing 0, $\tau_r(r) = d\tau(r)/dr$, $\tau_r(0) \neq 0$, and

$$\hat{\pi}_i = \frac{\tau_r(\hat{\lambda}'\hat{g}_i)}{\sum_{j=1}^n \tau_r(\hat{\lambda}'\hat{g}_j)}, \hat{\lambda} = \arg\max_{\lambda'\hat{g}_i \in \mathcal{V}} \sum_{i=1}^n \tau(\lambda'\hat{g}_i). \tag{2.6}$$

Each of these estimators will also satisfy the moment conditions, by the first order condition to the maximization problem, and will be asymptotically efficient, as shown in the appendix. They include the two that we have considered for $\tau(r) = \ln(1 - r)$ and $\tau(r) = -(1 + r)^2/2$ respectively.

A comparison of this bootstrap with that based on the empirical distribution $\tilde{F}$, where $\hat{\pi}_i = 1/n$, helps explain the importance of imposing the moment conditions on the bootstrap distribution. Since, in general, $\int g(z, \hat{\beta})\tilde{F}(dz) = \hat{g}(\hat{\beta}) \neq 0$, the moment restrictions are *not* satisfied for the empirical distribution. Consequently, the empirical distribution bootstrap for the overidentification test statistic will not be consistent, because its limiting distribution has a discontinuity where the estimated moment conditions depart from zero. Indeed, it can be shown that the limit of the bootstrap distribution estimator for $T$ is the distribution of $2Y$, where $Y$ is distributed $\chi^2(m - p)$. Also, we found in Monte Carlo experiments that the test based on the quantile $\hat{q}_a^B$ from the empirical distribution never rejects. Furthermore, the asymptotic variance estimator $\hat{V}$ is based on the moment conditions, and when they are not satisfied, it will be not be a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\beta} - p\lim(\hat{\beta}))$, see Maasoumi and Phillips (1982). Consequently, the limiting distribution of the t-ratio is not standard normal over all $F$ that includes $\tilde{F}$. As a result the empirical CDF bootstrap does satisfy the asymptotic pivotal condition of Beran (1988), and hence does not yield an improvement.

Hahn (1996) did show that for $S(z_1, ..., z_n, b) = \sqrt{n}(\hat{\beta} - b)$, the bootstrap based on the empirical distribution does yield a consistent estimator of the asymptotic distribution. This fact can be used to construct confidence intervals for $\beta_0$ via the percentile method. Because the limiting distribution depends on the truth, this method does not generally yield a theoretical improvement, although it might be useful in practice. In particular, because it is centered at the estimator $\hat{\beta}$, so that each bootstrap realization has the form $S(z_1^b, ..., z_n^b, \hat{\beta}) = \sqrt{n}(\hat{\beta}^b - \hat{\beta})$, it should also provide a consistent estimator of the asymptotic distribution of $\sqrt{n}(\hat{\beta} - p\lim(\hat{\beta}))$ under misspecification, that could be used in place of the complicated variance formulae of Maasoumi and Phillips (1982).

Hall and Horowitz (1996) (HH henceforth) showed how to modify the empirical distribution bootstrap so that it yields asymptotic improvements. To describe their approach let $\mu(b, F) = \int g(z, b)F(dz)$ for a CDF $F$ and a $p \times 1$ vector $b$ that represents a possible value of the true parameter, and let $\tilde{\mu} = \mu(\hat{\beta}, \tilde{F}) = \hat{g}(\hat{\beta})$. The HH bootstrap consists of steps 2) - 5) above with the empirical CDF $\tilde{F}$ used in place of the EL $\hat{F}$ in step 2) and $g(z, \beta) - \tilde{\mu}$ used in place of $g(z, \beta)$ in step 3). This modification has the effect of "recentering" the moment conditions, avoiding the problems mentioned above. However, it does complicate step 3), requiring modification of GMM software to allow replacement of the moment functions $g(z, \beta)$ by their recentered counterparts $g(z, \beta) - \tilde{\mu}$ when doing bootstrap computations. Also, the resulting estimators of the distribution of the t-ratio and the overidentifying test statistic will be asymptotically inefficient relative to the EL bootstrap. Intuitively, the bootstrap procedure for EL can be seen to be the same as for HH, differing only in the estimator of $F_0$, because the recentering goes away for EL (by $\mu(\hat{\beta}, \hat{F}) = 0$). Then, since the bootstrap is the same except that it is based on a more efficient distribution estimator, the bootstrap should be more efficient. We make this discussion more precise in the next Section.

# 3 Bootstrap Efficiency

The bootstrap improvement can be explained in terms of Edgeworth expansions, as in Singh (1981), Hall (1986), Beran (1988). The efficiency of the EL bootstrap relative to the HH bootstrap can also be explained by a refinement of this approach. In these explanations we follow much of the literature on higher-order asymptotic properties and derive formal Edgeworth expansions without specifying all the regularity conditions for their validity. The exposition follows that of Beran (1988) and Horowitz (1996).

For ease of comparison of the EL and HH bootstrap, we set up the analysis so that both are included. Let $F$ denote the CDF of a single observation of an i.i.d. sequence $z_1, ..., z_n$ with true value $F_0$. Also let $S(z_1, ..., z_n, b, F)$ be the t-ratio or overidentification statistic when $g(z, \beta)$ is replaced by $g(z, \beta) - \mu(b, F)$, and the true parameter in the t-ratio is equal to the $j^{th}$ component of $b$. The precise form of $S(z_1, ..., z_n, b, F)$ can be simply described as the calculated value of the t-ratio or overidentifying test statistic when the moment functions are recentered by subtracting off $\mu(b, F)$. Also, let

$$H_n(b, F) = \Pr(S(z_1, \ldots, z_n; b, F) \leq t|F), \tag{3.1}$$

where the $t$ is suppressed for notational convenience. Note that $\mu(\beta_0, F_0) = 0$ at the true $F_0$ and $\beta_0$, so that $S(z_1, ..., z_n, \beta_0, F_0)$ is equal to the actual t-ratio or overidentification statistic, and $H_n(\beta_0, F_0)$ to its true distribution. The HH bootstrap estimator of $H_n(\beta_0, F_0)$ is $H_n(\hat{\beta}, \tilde{F})$, where $\tilde{F}$ is the empirical distribution. More precisely, the HH bootstrap is a simulation estimator of $H_n(\hat{\beta}, \tilde{F})$. For the purposes of this discussion we ignore the error from simulation; see Andrews and Buchinskii (2000) for results on the size of that error.

Consider a formal Edgeworth expansion

$$H_n(b, F) = H_\infty + n^{-\alpha} R_1(b, F) + n^{-\gamma} R_2(b, F) + o(n^{-\gamma}) \tag{3.2}$$

where $\gamma \geq \alpha + 1/2$ and $R_1$ and $R_2$ are remainder coefficients. In this equation $H_\infty$ is the limit of $H_n(b, F)$, assumed to exist and to not depend on $(b, F)$. Because the moment conditions have been normalized to have mean zero when $b$ and $F$ are true, the limiting distribution for $H_n(b, F)$ has this property. In the case of the absolute value of the t-statistic, $H_\infty$ is the CDF of the absolute value of a standard normal CDF at $t$, and for the overidentification test statistic it is the CDF of $\chi^2(m-p)$. "Asymptotically pivotal" refers to the fact that the limiting value $H_\infty$ does not depend on $b$ or $F$.

The asymptotic pivotal condition is well known to lead to bootstrap improvements (e.g. see the nice discussion in Horowitz, 1996). Evaluating equation (3.2) at $\tilde{F}$ and $F_0$ and differencing gives the bootstrap error

$$H_n(\hat{\beta}, \tilde{F}) - H_n(\beta_0, F_0) = n^{-\alpha}[R_1(\hat{\beta}, \tilde{F}) - R_1(\beta_0, F_0)] + o_p(n^{-\alpha}) = o_p(n^{-\alpha}), \qquad (3.3)$$

for $R_1(\hat{\beta}, \tilde{F}) \xrightarrow{p} R_1(\beta_0, F_0)$ and $R_2(\beta, F) = O_p(n^{-\alpha})$. This error is of smaller order than the asymptotic distribution error $H_\infty - H_n(\beta_0, F_0) = O(n^{-\alpha})$, showing a large sample improvement for the HH bootstrap.

Exactly the same analysis shows that the EL bootstrap gives an improvement. Note that for any $b$ and $F$ satisfying $\mu(b, F) = 0$, the function $S(z_1, ..., z_n, b, F)$ coincides with the original GMM t-ratio or test statistic, *without* subtracting off $\mu(b, F)$. In particular, since $\mu(\hat{\beta}, \hat{F}) = 0$ for EL, it follows that $S(z_1^b, ..., z_n^b, \hat{\beta}, \hat{F})$ is just the original t-ratio or overidentification test statistic, which is computed exactly as described in step 3) of the EL bootstrap. Thus, $H_n(\hat{\beta}, \hat{F})$ is simply the EL bootstrap estimator of the distribution of $S$. Then, just as for HH, the EL bootstrap error will satisfy equation (3.3), and so be a large sample improvement.

A refinement of this approach can be used to show that EL bootstrap is efficient in a certain sense. To explain the efficiency property considered here, evaluate equation (3.2) at both $(\hat{\beta}, \hat{F})$ and $(\beta_0, F_0)$, subtract the second from the first, and multiply through by $n^{a+1/2}$ to obtain

$$\begin{aligned} n^{a+1/2}[H_n(\hat{\beta}, \hat{F}) - H_n(\beta_0, F_0)] &= \sqrt{n}[R_1(\hat{\beta}, \hat{F}) - R_1(\beta_0, F_0)] \\ &\quad + O(1)[R_2(\hat{\beta}, \hat{F}) - R_2(\beta_0, F_0)] + o_p(1) \\ &= \sqrt{n}[R_1(\hat{\beta}, \hat{F}) - R_1(\beta_0, F_0)] + o_p(1). \end{aligned}$$

where the second equality assumes that $R_2(\hat{\beta}, \hat{F}) \xrightarrow{p} R_2(\beta_0, F_0)$. Then, as long as

$\sqrt{n}[R_1(\hat{\beta}, \hat{F}) - R_1(\beta_0, F_0)]$ is asymptotically normal, the bootstrap approximation error will also be asymptotically normal, at the rate $n^{\alpha+1/2}$. Furthermore, the asymptotic variance of the bootstrap remainder is directly related to the asymptotic variance of the coefficient $R_1(\hat{\beta}, \hat{F})$. The more efficient $R_1(\hat{\beta}, \hat{F})$ is as an estimator of $R_1(\beta_0, F_0)$, the closer will the bootstrap remainder be to zero in large samples. In essence this notion of bootstrap efficiency concerns the efficiency with which the bootstrap estimates the coefficients in the bootstrap error, rather than the rate of convergence of the error. The importance of this efficiency relative to convergence rate improvements depends on the size of these coefficients in particular models.

It is straightforward to show that the EL bootstrap is asymptotically efficient relative to the bootstrap based on any other estimator of $\beta$ and $F$, including the HH bootstrap. In the GMM setting $R_1(b, F)$ will depend on expectations, over $F$, of functions of $g(z, b)$ and its derivatives with respect to $b$. That is, it will take the form

$$R_1(b, F) = r_1(\int m(z, b)F(dz));$$

for some vector $m(z, b)$ of functions of the data and parameter, see Proposition 1 of HH. It follows from Theorem 4 in the Appendix and Brown and Newey (1998) that $\int m(z, \hat{\beta})\hat{F}(dz)$ is an efficient semiparametric estimator of $\int m(z, \beta_0)F_0(dz)$. Then by the delta method it follows that $R_1(\hat{\beta}, \hat{F})$ is an efficient semiparametric estimator of $R_1(\beta_0, F_0)$, having asymptotic variance smaller than any other estimator that only uses the moment condition. In particular, it will have asymptotic variance smaller than $R_1(\hat{\beta}, \tilde{F})$. In this sense the EL bootstrap is efficient relative to HH (or that based on any other estimator of $\beta$ and $F$), that is, the EL bootstrap error has a smaller asymptotic variance than the HH bootstrap error.

One interesting issue concerns the size of the efficiency gain of the EL bootstrap over the HH. To provide some information about this gain, consider a simple example. Let $z$ be a scalar and consider the hypothesis $E[z] = 0$. A simple, standard test statistic is $S = (\sum_{i=1}^{n} z_i)/\sqrt{\sum_{i=1}^{n} z_i^2}$. Here $S^2$ corresponds to the overidentification statistic for moment conditions where $\beta$ is not present, and $g(z, \beta) = z$. The standard approach to bootstrapping here would be to draw the observations from the empirical distribution and center them at the sample mean, by simulating the distribution of $\sum_{i=1}^{n}(z_i^b - \bar{z})/\sqrt{\sum_{i=1}^{n}(z_i^b - \bar{z})^2}$ where $\bar{z} = \sum_{i=1}^{n} z_i/n$. This bootstrap corresponds to that of HH. The EL bootstrap would instead draw $z_i^b$ from the distribution with probability $\hat{\pi}_i$ for each $i$, where $\hat{\pi}_i$ are formed as in equation (2.5) with $g(z,\beta) = z$, i.e. as the solution to

$$\hat{\pi}_i = \frac{1}{n(1-\hat{\lambda}z_i)}, (i = 1, ..., n), \hat{\lambda} = \arg\max_{\lambda z_i < 1} \sum_{i=1}^{n} \ln(1 - \lambda z_i).$$

Then for each draw $\sum_{i=1}^{n} z_i^b/\sqrt{\sum_{i=1}^{n}(z_i^b)^2}$ would be calculated, and the distribution of the t-ratio under the null hypothesis estimated from the realizations of these bootstrap simulations.

The well known form of the Edgeworth expansion of $S$ can be used to compare the asymptotic efficiency of the EL and HH bootstraps. Let $\phi$ denote the normal density and $\kappa(F)$ and $\gamma(F)$ denote the skewness and kurtosis, respectively, of the distribution $F$. Then applying the delta method to the Edgeworth expansion in Hall (1992, p. 73) gives

$$\begin{aligned} n^{3/2}\left[P(S \geq k|\hat{F}) - P(S \geq k|F_0)\right] &= \sqrt{n}\{c_\kappa[\kappa(\hat{F}) - \kappa(F_0)] + c_\gamma[\gamma(\hat{F}) - \gamma(F_0)]\} + o_p(1), \\ c_\kappa &= k(k^2-3)\phi(k)/6, c_\gamma = -2k(k^4 + 2k^2 - 3)\phi(k)\gamma(F_0)/9. \end{aligned}$$

Thus, the efficiency of the bootstrap will be determined by the efficiency of a linear combination of skewness and kurtosis. In calculations not reported here, it was found there is no efficiency gain if the distribution of $z$ is symmetric about its mean or if it is a member of the chi-square family of distributions, but there is some gain in other cases.

Often the purpose of the bootstrap is improved inference rather than estimation of distributions. This is a more complicated problem, where implications of bootstrapping from an efficient distribution estimator are more involved. To describe the inference problem we focus on the overidentification statistic. Let $\hat{q}_n(b, F, \alpha)$ be the critical point for a level $\alpha$ test when $b$ and $F$ are true, obtained by solving for $q$ from

$$\Pr(S_n(z_1, ..., z_n, b, F) \geq q|F) = \alpha.$$

The EL bootstrap estimator of this quantile is $\hat{q}_n = \hat{q}_n(\hat{\beta}, \hat{F}, \alpha)$. By the delta method it will follow similarly to the previous discussion that the difference of $\hat{q}_n$ and the true critical value will have smaller asymptotic variance for the EL bootstrap than for HH. However, this efficiency comparison need not carry over to the level of the test. The problem is that $S_n$ and $\hat{q}_n$ may be correlated, so that a more efficient $\hat{q}_n$ need not lead to a test level that is closer to its desired value. Among $\hat{q}_n$ that are independent of $S$ improved efficiency will improve the accuracy of test level, but otherwise need not.

An improvement in test level accuracy from the EL bootstrap does occur for some cases. As shown in Theorem 4 in the appendix, $\hat{g}(\hat{\beta})$ and $\hat{F}$ are asymptotically independent, while $\hat{g}(\hat{\beta})$ and $\tilde{F}$ are not. Consequently, by Davidson and MacKinnon (1999), for one-sided tests based on $\hat{g}(\hat{\beta})$ (i.e. tests of the moment conditions against alternatives in one direction) the level of the EL bootstrap test converges to its nominal level at a faster rate than does the level of the HH bootstrap test. The previous example of a one-sided test of the null hypothesis $E[z] = 0$ is one such case. The result of Davidson and MacKinnon (1999), when combined with the asymptotic independence of $\hat{F}$ and $S$, implies that the difference between the level of the EL bootstrap and the actual level is $O_p(n^{-3/2})$, rather than $O_p(n^{-1})$ for the HH bootstrap. Hall and Presnell (1999) give some theoretical results on the power of the EL test in this setting, and in a Monte Carlo study find mixed results on one-sided tests but find that for two-sided tests the EL bootstrap is more accurate than the HH bootstrap.

# 4 Dynamic Panel Examples

To show how well the EL bootstrap can work in practice we consider two dynamic panel data examples. We refer the interested reader to Bond and Windmeijer (2001) for a more extensive study. We first consider a Monte Carlo experiment for an autoregressive panel data model with individual effect of the form:

$$y_{it} = \rho_0 y_{i,t-1} + \alpha_i + \varepsilon_{it}, y_{i0} = \frac{\alpha_i}{1-\rho_0} + v_i, (t = 1, \ldots, 4, i = 1, \ldots, n). \qquad (4.1)$$

We assume that $\varepsilon_{it}$ and $\alpha_i$ have standard normal distributions, $v_i \sim N(0, 1/(1-\rho_0^2))$, and $\varepsilon_{i1}, \ldots, \varepsilon_{i4}, \alpha_i, \nu_i$ are mutually independent. In this model, $y_{it}$ is stationary over time because of the specification of the equation describing $y_{i0}$. Moment conditions can be obtained by time differencing and using lagged levels as instruments, see Anderson and Hsiao (1982), and Holtz-Eakin, Newey, and Rosen (1988). Specifically, equation (4.1) implies that

$$\Delta y_{it} = \rho_0 \Delta y_{i,t-1} + \Delta\varepsilon_{it}, E(y_{i,t-j}\Delta\varepsilon_{it}) = 0, (j \geq 2).$$

With four time periods, these conditions lead to a three dimensional moment vector where $\beta = \rho$ and

$$g(z_i, \beta) = [y_{i1}(\Delta y_{i3} - \beta\Delta y_{i2}), y_{i1}(\Delta y_{i4} - \beta\Delta y_{i3}), y_{i2}(\Delta y_{i4} - \beta\Delta y_{i3})]'. \tag{4.2}$$

We take the initial weighting matrix $\hat{W}$ to be block diagonal, with the first block being the inverse sample second moment of $y_{i1}$ and the second block being the inverse sample second moment of $(y_{i1}, y_{i2})$. We use a value of $\rho = 0.5$ and two sample sizes, $n = 50$ and $n = 100$. Table 1 reports the results of this experiment, listing the actual levels of asymptotic and bootstrap tests of nominal levels .10, .05, and .01 and the coverage probabilities for nominal 90 percent confidence intervals.

Table 1: Dynamic Panel Model

| | $n = 50$ | | | | $n = 100$ | | |
|---|---|---|---|---|---|---|---|
| | Nom | Asym | Boot | | Nom | Asym | Boot |
| Cov Prob | .90 | .80 | .88 | | .90 | .85 | .90 |
| Test Level | .10 | .12 | .11 | | .10 | .126 | .114 |
| Test Level | .05 | .066 | .058 | | .05 | .065 | .0575 |
| Test Level | .01 | .012 | .013 | | .01 | .014 | .015 |

Here we find some improvement in accuracy for the overidentifying test, and substantial improvement for the coverage probability of the confidence interval. These results show that bootstrapping for GMM can provide large improvements in panel data, in contrast to the results for the simultaneous equations models in Hall and Horowtiz (1996).

An experiment was also carried out using an additional moment condition that was suggested by Ahn and Schmidt (1994). Uncorrelatedness of $\alpha_i$ and $\varepsilon_{it}$ for each time period implies that $E[(y_{i4} - \rho_0 y_{i3})(\Delta y_{i3} - \rho_0 \Delta y_{i2})] = 0$. This moment condition is nonlinear in $\rho$, but it is straightforward to use an initial estimator to form a linearized version. Linearizing this moment condition around an initial estimator $\tilde{\rho}$ gives

$$y_{i4}\Delta y_{i3} - \tilde{\rho}^2 y_{i3}\Delta y_{i2} - (y_{i3}\Delta y_{i3} + y_{i4}\Delta y_{i2} - 2\tilde{\rho} y_{i3}\Delta y_{i2})\,\rho.$$

We add this fourth moment condition to the three from equation (4.2) and re-evaluate the performance of asymptotic and bootstrapped critical values. Table 2 reports the results of this experiment.

Table 2: Dyanmic Panel with Extra Moment

| | $n = 50$ | | | | $n = 100$ | | |
|---|---|---|---|---|---|---|---|
| | Nom | Asym | Boot | | Nom | Asym | Boot |
| Cov Prob | .90 | .70 | .89 | | .90 | .79 | .90 |
| Test Level | .10 | .23 | .099 | | .10 | .11 | .083 |
| Test Level | .05 | .13 | .045 | | .05 | .57 | .040 |
| Test Level | .01 | .037 | .009 | | .01 | .014 | .007 |

Here the bootstrap gives an even bigger improvement than in the previous case.

For an empirical example we consider the application of Blundell et al. (1992). The statistical model takes the form:

$$(I/K)_{it} = \beta Q_{it} + \gamma_t + \alpha_i + \nu_{it} \tag{4.3}$$

for firms $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$. Here $Q_{it}$ is an empirical measure of the ratio of the shadow value of capital and the unit price of investment goods, and $(I/K)_{it}$ is the investment rate. We use a subset of the data from Blundell et al. (1992) choosing a balanced panel of 532 firms over 10 years with different initial years ranging from 1971 to 1977. As such, the coefficients $\gamma_t$ in equation (4.3) are specified by calendar year, rather than time elapsed from the initial year in the sample.

To estimate the investment coefficient, $\beta$, consider instrumental variables estimators that use lagged values of $Q$ as instruments to estimate the differenced equation. These instruments allow for contemporaneous correlation of $Q_{it}$ and $v_{it}$, and for correlation with the fixed effect. This specification leaves different numbers of instruments for different time periods because more lags are available as instruments for the later time periods.

After differencing and using a year of data to generate lagged variables we were left with eight years of data. We also dropped the last year of data for each firm because of a singularity problem with the bootstrap, and because the $\gamma_t$ coefficient for the last year seemed to be quite different than the others, suggesting that it might be an outlier. This leaves seven years of data, and a total of $1+2+3+4+5+6+7 = 28$ moment conditions.

In describing the results we focus on the estimator of $\beta$, the coefficient of $Q$, and on the overidentifying test statistic. The estimate of $\beta$ was 0.00758 with a standard error of 0.00376. The point estimate is similar to that of Blundell et al. (1992), although it is smaller and much less precisely estimated. According to the asymptotic standard errors, it is significant at the 0.05 significance level. The overidentifying test statistic is 32.14 and has 27 degrees of freedom; it is below the asymptotic critical value of 40.11.

We found that the bootstrap confidence intervals for the coefficient are wider than the asymptotic confidence intervals. The bootstrap p-value for the t-statistic was 0.103, compared to the asymptotic value which was less than 0.05. Thus, the coefficient is less statistically significant when the bootstrap approximation is used rather than the asymptotic critical values. Also, the bootstrap distribution of the overidentifying test statistic has more weight in the tail than the asymptotic distribution; the bootstrap 0.05 critical value is 42.79, compared to the asymptotic critical value of 40.11.

The bootstrap is a Monte Carlo experiment for a design based on the data, so that the shape of the distribution is informative about small sample properties. Figure 1 gives a kernel estimator of the bootstrap density of the t-statistic. It is left skewed, and centered at a point below zero, suggesting the presence of some downward bias. The density of the coefficient has a similar shape. This shape is much like the small sample distribution of an autoregressive coefficient in time series, being skewed to the left and downward biased. Figure 2 displays the bootstrap density of the overidentifying test statistic. It is roughly similar to that of a chi-squared, but is shifted to the right. It is also multimodal near the peak, atlthough it is not clear whether this is an important feature of the density. These shapes are consistent with our Monte Carlo findings, where there is some tendency towards over rejection for the asymptotic approximation of the overidentification statistic.

# 5 Other Approaches to Improved Inference

In addition to the bootstrap their are several other approaches to improved inference for GMM. One potential reason for the poor performance of asymptotic theory for GMM is that parameters of interest are not well identified. Hansen, Heaton, and Yaron (1996) and Stock and Wright (2000) find evidence of this problem for intertemporal capital asset pricing models. The bootstrap may not help in this situation, because it is based on higher-order approximations for well identified models. Stock and Staiger (1997) and Stock and Wright (2000) address this issue by deriving asymptotic distribution theory for IV and GMM under weak identification, where the degree of identification shrinks as the sample size grows. This kind of asymptotic theory is meant to lead to better approximations for low identification cases. They show that the resulting distributional approximations are better than those obtained from the usual asymptotics in a wide variety of circumstances. They also give inference methods based on the overidentification test statistic that are asymptotically correct under weak instrument asymptotics.

There are at least two concerns with using the overidentifying test statistic for inferences concerning parameters of interest. First, it has too many degrees of freedom in general, leading to relatively low power. For instance, consider a test that one parameter is equal to some value. One would normally want a test with one degree of freedom to test such a null. If instead one uses the overidentifying test statistic under the null hypothesis, one has a many degree of freedom test, which will have lower power. The second problem is that misspecification of the model has the strange consequence of making inferences seem more accurate. As the moment conditions depart from the truth the overidentification test statistic will tend to be larger, leading to smaller confidence intervals. In our view it is unsatisfactory to base inferences on a procedure that seems to be more accurate when the model is misspecified.

Recently, Kleibergen (2001a, b) and Moreira (2001) have developed tests that have the right degrees of freedom and are asymptotically correct under weak identification asymptotics. The Kleibergen (2001a, b) tests are based on the derivative of the overidentification statistic with respect to the parameters. The Moreira (2001) tests are based on the theory of similar tests. Both provide promising approaches to inference under weak instruments.

For cross-section data it may be that using too many instruments is the primary problem for IV asymptotic approximations. Bekker (1994) addresses this problem by deriving asymptotic distribution for IV and limited information maximum likelihood estimators when the number of instruments grow at the same rate as the sample size. Donald and Newey (2001) show that this problem can be entirely avoided by choosing the number of instruments to minimize higher-order mean-square error. In a leading example, the Angrist and Krueger (1991) study of returns to education, this criteria leads to only using quarter of birth as instruments, avoiding the problem with too many instruments that was shown by Bound, Jaeger, and Baker (1996). Hall and Inoue (2001) give a cannonical correlations approach to the choice of moment conditions.

Detecting departures from the asymptotic approximation is an important problem. For IV, Hahn and Hausman (2001) have proposed a specification test based on comparing forward and reverse IV estimators. They find that often this test accurately detects when the asymptotic approximation is poor.

A quite different way of approaching inference problems for GMM is to change the estimator itself rather than just the approximation to its distribution. The idea is that if the inference problems are caused by specific features of the estimator, such as bias, then it is better to first fix the estimator, and then apply the bootstrap to obtain further improvements. To this end, a number of alternative estimators have been proposed. These include the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), and Imbens (1997), the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996), and the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). As shown by Smith (1997) and Newey and Smith (2001), these estimators share a common structure, being members of a class of generalized empirical likelihood (GEL) estimators. Estimators in this class are obtained as the solution to a saddle point problem of the form

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} \max_{\lambda' g_i(\beta) \in \mathcal{V}} \sum_{i=1}^{n} \tau(\lambda' g_i(\beta)),$$

where $\tau(r)$ is a concave function with open domain $\mathcal{V}$ as considered for the EL bootstrap. Newey and Smith (2001) derived asymptotic expansions for these estimators and for GMM. They showed that each GEL estimator removes the term in the higher-order bias for GMM corresponding to estimation of derivatives, and that EL (where $\tau(r) = \ln(1-r)$) also removes the bias term for estimating the second moment matrix. They derived bias corrected GMM and GEL estimators, and showed that among these EL is second-order efficient. Thus, to the extent inference problems are the result of bias, the use of EL will help improve GMM inference.

An interpretation of EL helps explain its low bias properties. As shown by Newey and Smith (2001) the first-order conditions for EL take the form

$$\begin{aligned} 0 &= [\sum_{i=1}^{n} \hat{\pi}_i \partial g(z_i, \hat{\beta})/\partial \beta]' [\sum_{i=1}^{n} \hat{\pi}_i g(z_i, \hat{\beta}) g(z_i, \hat{\beta})]^{-1} \hat{g}(\hat{\beta}), \\ \hat{\pi}_i &= n^{-1}[1 - \hat{\lambda}' g(z_i, \hat{\beta})]^{-1}. \end{aligned}$$

In comparison with the GMM first-order condition $[n^{-1} \sum_{i=1}^{n} \partial g(z_i, \hat{\beta})/\partial \beta] \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\hat{\beta}) = 0$, the EL estimator uses averages over the EL probabilities, rather than sample averages, to form the linear combination coefficients that appear in the first-order conditions. These averages are efficient estimators of the corresponding expectations under the moment conditions. As discussed in Brown and Newey (2001), using such efficient estimators removes bias due to their estimation, explaining the higher-order bias results for EL. Brown and Newey (2001) also extend these results to any semiparametric estimator where the nuisance parameters are estimated efficiently.

# 6 Appendix: Asymptotic Theory

Here we give a brief account of the asymptotic properties of GMM and show that the distribution estimator of equation (2.6) has the properties claimed in the text. Theorems

1 - 3 are given in Newey and McFadden (1994), but we have collected them here in the hope that a brief, coherent account might be useful. Pakes and Pollard (1989) give more general asymptotic normality results, that allow for $g(z,\beta)$ to be nondifferentiable. The conclusions remain the same, except that $G = \partial E[g(z,\beta)]/\partial\beta|_{\beta_0}$ throughout. We assume throughout this discussion that equation (2.1) is satisfied.

We first give conditions for consistency, that allow for $g(z,\beta)$ to be discontinuous in $\beta$.

**Assumption 1:** $\hat{W} \xrightarrow{p} W$ and i) $W$ is positive semidefinite and $WE[g(z,\beta)] = 0$ only if $\beta = \beta_0$; ii) $\beta_0 \in \mathcal{B}$, which is compact; iii) $g(z,\beta)$ is continuous at each $\beta \in \mathcal{B}$, with probability one; iv) $E[\sup_{\beta\in\mathcal{B}} \|g(z,\beta)\|] < \infty$.

**Theorem 1:** *If Assumption 1 is satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$.

This result is similar to Theorem 2.1 of Hansen (1982). Strictly speaking, the presence of the discontinuity means that the minimum of $\hat{g}(\beta)'\hat{W}\hat{g}(\beta)$ may not exist, but it suffices for consistency that $\hat{g}(\hat{\beta})'\hat{W}\hat{g}(\hat{\beta}) \leq \inf_{\beta\in\mathcal{B}} \hat{g}(\beta)'\hat{W}\hat{g}(\beta) + o_p(1)$, as in Pakes and Pollard (1989).

With additional regularity conditions we also establish asymptotic normality.

**Assumption 2:** i) $\beta_0 \in int(\mathcal{B})$; ii) $g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $\beta_0$ with probability one; iii) $E[\|g(z,\beta_0)\|^2] < \infty$; iv) $E[\sup_{\beta\in\mathcal{N}} \|\partial g(z,\beta)/\partial\beta\|] < \infty$; v) $G'WG$ is nonsingular for $G = E[\partial g(z,\beta_0)/\partial\beta]$.

**Theorem 2:** *If Assumptions 1 and 2 are satisfied then*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}).$$

For nonsingular $\Omega$ the asymptotic efficiency of a GMM estimator with $W = \Omega$ follows from the matrix equality

$$(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1} - (G'\Omega^{-1}G)^{-1} = A[I - B(B'B)^{-1}B']A',$$

where $A = (G'WG)^{-1}G'WC'$, $B = (C')^{-1}G$, where $C$ is a square root of $\Omega$, satisfying $C'C = \Omega$. Because the matrix in square brackets is idempotent, and hence positive semidefinite, it follows that the difference of the asymptotic variance matrix for general $W$ and that for $W = \Omega^{-1}$ is positive semidefinite.

The following is a result for the two-step efficient GMM estimator and overidentification test statistic.

**Theorem 3:** *If Assumptions 1 and 2 are satisfied, $\Omega$ is nonsingular, and $E[\sup_{\beta\in\mathcal{N}} \|g(z,\beta)\|^2] < \infty$ then*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \hat{V} \xrightarrow{p} V, V = (G'\Omega^{-1}G)^{-1},$$

*and $T \xrightarrow{d} \chi^2(m-p)$.*

Finally, we show efficiency for the distribution estimator. Let $1_i = 1(z_i \leq z)$, $g_i = g(z_i, \beta_0)$, $H = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}$ and $\psi_i = 1_i - E[1_i] - E[1_i g_i']\Omega^{-1}(I+GH)g_i$. Abbreviate "with probability approaching one" as w.p.a.1.

**Theorem 4:** *If Assumptions 1 and 2 are satisfied, $\Omega$ is nonsingular, $\tau(r)$ is twice continuously differentiable in a neighborhood of $r = 0$, $E[\sup_{\beta\in\mathcal{N}} \|g(z,\beta)\|^\zeta] < \infty$ for $\zeta > 2$, and $\hat{\beta}$ is any estimator with $\sqrt{n}(\hat{\beta} - \beta_0) = H\sqrt{n}\hat{g}(\beta_0) + o_p(1)$, then $\hat{\pi}_i$ from equation (2.6) exist and are nonnegative w.p.a.1, $\hat{F}(z) = \sum_{i=1}^n 1_i\hat{\pi}_i$ satisfies*

$$\sqrt{n}(\hat{F}(z) - F_0(z)) \xrightarrow{d} N(0, E[\psi_i^2]),$$

*$\hat{F}(z)$ is asymptotically efficient, and $\hat{F}$ and $\hat{g}(\hat{\beta})$ are asymptotically independent.*

Proof: By hypothesis, $\hat{\beta} \xrightarrow{p} \beta_0$, and by a standard mean value expansion

$$\hat{g}(\hat{\beta}) = (I + GH)\hat{g}(\beta_0) + o_p(1/\sqrt{n}) = O_p(1/\sqrt{n}), \tag{6.1}$$

so by Lemma A2 of Newey and Smith (2001) it follows that $\hat{\lambda}$ exists with probability approaching one (w.p.a.1) and $\hat{\lambda} = O_p(1/\sqrt{n})$. Then by Lemma A1 of Newey and Smith (2001), $\max_{i\leq n} |\hat{\lambda}'\hat{g}_i| \xrightarrow{p} 0$. The first conclusion then follows by continuity of $\tau_r(r)$ at $r = 0$. Now, normalize $\tau(r)$ as in Newey and Smith (2001), so that $\tau_r(0) = \tau_{rr}(0) =$

$-1$. A mean-value expansion in $\lambda$ of the first-order condition $\sum_{i=1}^{n} \tau_r(\hat{\lambda}'\hat{g}_i)\hat{g}_i/n = 0$ for $\hat{\lambda}$ then gives $0 = -\hat{g}(\hat{\beta}) + [\sum_{i=1}^{n} \tau_{rr}(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]\hat{\lambda}$. By continuity of $\tau_{rr}(r)$ at $r = 0$, $\max_{i \le n} |\tau_{rr}(\dot{\lambda}'\hat{g}_i) - 1| \xrightarrow{p} 0$. Then by Lemma 2.4 of Newey and McFadden (1994), $\sum_{i=1}^{n} \tau_{rr}(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n \xrightarrow{p} -\Omega$, so that

$$\hat{\lambda} = -\Omega^{-1}\hat{g}(\hat{\beta}) + o_p(1/\sqrt{n}). \tag{6.2}$$

Next, by a second-order mean-value expansion and the first-order conditions for $\hat{\lambda}$, $\sum_{i=1}^{n} \tau_r(\hat{\lambda}'\hat{g}_i)/n = -1 + \hat{\lambda}'[\sum_{i=1}^{n} \tau_{rr}(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]\hat{\lambda} = -1 + o_p(1/\sqrt{n})$. Also, by another mean-value expansion and Lemma 2.4 of Newey and McFadden (1994),

$$\begin{aligned} \sum_{i=1}^{n} 1_i\tau_r(\hat{\lambda}'\hat{g}_i)/n &= -\sum_{i=1}^{n} 1_i/n + [\sum_{i=1}^{n} 1_i\tau_{rr}(\dot{\lambda}'\hat{g}_i)\hat{g}_i/n]\hat{\lambda} \\ &= -\sum_{i=1}^{n} 1_i/n - E[1_i g_i]\hat{\lambda} + o_p(1/\sqrt{n}). \end{aligned}$$

It then follows by the delta method and eqs. (6.1) and (6.2) that

$$\begin{aligned} \hat{F}(z) - F_0(z) &= [\sum_{i=1}^{n} 1_i\tau_r(\hat{\lambda}'\hat{g}_i)/n]/[\sum_{i=1}^{n} \tau_r(\hat{\lambda}'\hat{g}_i)/n] - E[1_i] \\ &= \sum_{i=1}^{n} 1_i/n - E[1_i] + E[1_i g_i]\hat{\lambda} + o_p(1/\sqrt{n}) \\ &= \sum_{i=1}^{n} 1_i/n - E[1_i] - E[1_i g_i]\Omega^{-1}\hat{g}(\hat{\beta}) + o_p(1/\sqrt{n}) = \sum_{i=1}^{n} \psi_i/n + o_p(1/\sqrt{n}). \end{aligned}$$

The second conclusion then follows by the Lindbergh-Levy central limit theorem. The third conclusion follows by Brown and Newey (1998). For the last conclusion, note that $\hat{F}$ and $\hat{g}(\hat{\beta})$ are joint asymptotically normal by the last equation and eq. (6.1), with covariance

$$E[\psi_i g_i'](I + H'G') = E[1_i g_i']\{(I + H'G') - \Omega^{-1}(I + GH)\Omega(I + H'G')\} = 0.$$

Q.E.D..

## References

# Bibliography


[1] Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica* 57, 411-445.

[2] Ahn, S.C. and P. Schmidt (1994): "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68.

[3] Altonji, J. and L.M. Segal (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Economic and Business Statistics* 14, 353-366.

[4] Amemiya, T. (1974): "The Nonlinear Two-Stage Least Squares Estimator," *Journal of Econometrics* 2, 105-110.

[5] Anderson, T.W. and C. Hsaio (1982): "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics* 18.

[6] Anderson, T.W. and T. Sawa (1979): "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," *Econometrica* 47, 163-182.

[7] Andrews, D.W.K. and M. Buchinsky (2000): "A Three-Step Method for Choosing the Number of Bootstrap Replications," *Econometrica* 68, 23-51.

[8] Angrist, J. and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics, 106*, 979–1014.

[9] Arellano, M. and B. Honore (2000): "Panel Data Models: Some Recent Developments," forthcoming in *Handbook of Econometrics, Volume 5.*

[10] Bound, J., D. Jaeger, and R. Baker (1996): "Problems with Instrumental Variables Estimation when the Correlation Between Instruments and the Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association, 90*, 443-450.

[11] Back, K. and D.P. Brown (1993): "Implied Probabilities in GMM Estimators." *Econometrica* 61, 971-976.

[12] Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica 62, 657-681.*

[13] Beran, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687-697.

[14] Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli (1992): "Investment and Tobin's Q: Evidence from Company Panel Data." *Journal of Econometrics* 51, 233-257.

[15] Bond, S. and F. Windjmeier (2001): "Finite Sample Inference for GMM Estimators in Linear Dynamic Panel Data Models," Institute for Fiscal Studies Working Paper.

[16] Brown, B.W., and W.K. Newey (2001): "Efficient Estimation of Nuisance Parameters in Semiparametric Models," mimeo, Rice University.

[17] Brown, B.W., and W.K. Newey (1992): "Bootstrapping for GMM," notes for seminar at Monash University.

[18] Brown, B., and W. Newey (l998): "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

[19] Burguette, J., A.R. Gallant, and G. Souza. (1982): "On the Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews* 1, 151-190.

[20] Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics 18, 5-42.*

[21] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics* 34, 305-334.

[22] Chiang, C.L. (1956): "On Regular Best Asymptotically Normal Estimates," *Annals of Mathematical Statistics* 27, 336-351.

[23] Cosslett, S.R. (1981): "Maximum Likelihood Estimation for Choice-Based Samples," *Econometrica* 49, 1289-1316.

[24] Cragg, J.G. (1983): "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica* 51, 751-763.

[25] Davidson, R. and J.G. MacKinnon (1999): "The Size Distortion of Bootstrap Tests," *Econometric Theory* 15, 361-376.

[26] Eichenbaum, M. and L.P. Hansen (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data." *Journal of Business and Economic Statistics* 8, 53-69.

[27] Ferguson, T.S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics* 29, 1046-1062.

[28] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, Cambridge, England: Cambridge University Press.

[29] Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators." *Econometric Theory* 12, 187-196.

[30] Hahn, J. and J.A. Hausman (2001): "A New Specification Test for the Validity of Instrumental Variables," working paper, Department of Economics, MIT.

[31] Hall, A.R. and A. Inoue (2001): "A Canonical Correlations Interpretation of Generalized Method of Moments Estimation with Applications to Moment Selection," mimeo, North Carolina State University.

[32] Hall, P. (1986): "On the Bootstrap and Confidence Intervals," *Annals of Statistics* 14, 1431-1452.

[33] Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

[34] Hall, P, and J. Horowitz (1996): “Bootstrap Critical Values for Tests Based on Generalized Method of Moments,” *Econometrica*, 64,891-916.

[35] Hall, P. and B. Presnell (1999): ”Intentionally Biased Bootstrap Methods,” *Journal of the Royal Statistical Society, Series B*, 61 143-158.

[36] Hansen, L.P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50, 1029-1054.

[37] Hansen, L.P. and K.J. Singleton. (1982): ”Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica* 50, 1269-1286.

[38] Hansen, L.P., J. Heaton and A. Yaron (1996): “Finite-Sample Properties of Some Alternative GMM Estimators”, *Journal of Business and Economic Statistics* 14, 262-280.

[39] Hansen, L.P. and J.A. Scheinkman (1995): ”Back to the Future: Generating Moment Implications for Continuous Time Markov Proceses,” *Econometrica* 63, 767-804.

[40] Hausman, J.A. and W. Taylor (1981): ”Panel Data and Unobservable Individual Effects,” *Econometrica* 49, 1377-1398.

[41] Holtz-Eakin, D., W.K. Newey, and H.S. Rosen (1988), ”Estimating Vector Autoregressions With Panel Data,” *Econometrica* 56, 1371-1396.

[42] Horowitz, J. (1996): ”Bootstrap Methods in Econometrics: Theory and Numerical Performance,” in D. M. Kreps and K. F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications Seventh World Congress*, Cambridge England, Cambridge University Press.

[43] Huber, P.J. (1967): "The Behavior of Maximum Likelihood Estimators Under Nonstandard Conditions," in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-233, Berkeley, CA: University of California.

[44] Imbens, G.W. (1992): "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice Based Sampling," *Econometrica* 60, 1187-1214.

[45] Imbens, G.W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies* 64, 359-383.

[46] Imbens, G., R. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66 333-357.

[47] Kleibergen, F. (2000): "Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression", Revised Version of Tinbergen Institute Discussion Paper 2000-055/4.

[48] Kleibergen, F. (2001): "Testing parameters in GMM without assuming that they are identified", Tinbergen Institute Discussion Paper 2001-067/4.

[49] Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica* 65, 861-874.

[50] Maasoumi, E. and P.C.B. Phillips (1982): "On the Behavior of Inconsistent Instrumental Variables Estimates," *Journal of Econometrics* 19, 183-203.

[51] MaCurdy, T.E. (2000): "Using Information on the Moments of Disturbances to Increase the Efficiency of Estimation," in Hsiao, C., K. Morimune, and J.L. Powell, eds., *Nonlinear Statistical Modeling*, Cambridge: Cambridge University Press.

[52] Malinvaud, E. (1970): *Statistical Methods of Econometrics,* First Edition, Amsterdam: North-Holland.

[53] Manski, C. (1988): *Analog Estimation Methods in Econometrics*, New York: Chapman and Hall.

[54] Moreira, M.J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," mimeo, UC Berkeley.

[55] Nagar, A.L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", *Econometrica* 27, 573-595.

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

[57] Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica* 53, 1047-1070.

[58] Newey, W.K. (1988): "Adaptive Estimation of Regression Models Via Moment Restrictions," *Journal of Econometrics* 38, 301-339.

[59] Newey, W.K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics, Volume 11: Econometrics.* Amsterdam: North-Holland.

[60] Newey, W.K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics, Volume 11: Econometrics.* Amsterdam: North-Holland.

[61] Newey, W.K. and K.D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review* 28, 777-787.

[62] Newey, W., and D. McFadden (1994): "Estimation and Inference in Large Samples," in *Handbook of Econometrics, Vol. 4,* ed. by R. Engle and D. McFadden, New York: North Holland, 2111-2445

[63] Newey, W.K. and T. Chipty (2000): "Efficient Estimation of Semiparametric Models Via Moment Restrictions," with T. Chipty, October 2000.

[64] Newey, W.K. and R.J. Smith (2001): "Higher-Order Properties of GMM and Generalized Empirical Likelihood Estimators," working paper, MIT Department of Economics.

[65] Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika* 75, 237-249.

[66] Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica 57, 1027-1057.*

[67] Pearson, K.A. (1900): "On the criterion that a given system of deviations from the probable in the case of a correlated system of varibles is such that it can be reasonably supposed to have arisen from random sampling, Pilosophical Magazine Series 50, 157-175.

[68] Powell, J.L. (1986): "Symetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435-1460.

[69] Qin, J. and Lawless, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics* 22, 300-325.

[70] Rothenberg, T.J. (1973): *Efficient Estimation with a Priori Information,* Cowles Foundation Monograph 23, New Haven: Yale University Press.

[71] Sargan, J.D. (1958): "On the Estimation of Economic Relationships by Means of Instrumental Variables," *Econometrica* 26, 393-415.

[72] Sargan, J.D. (1959): "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society, Series B* 21, 91-105.

[73] Singh, K. (1981): "On the Asymptotic Accuracy of Efron's Bootstrap," *Annals of Statistics* 9, 1187-1195.

[74] Smith, R. J. (1997): “Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation”, *Economic Journal* 107, 503-519.

[75] Staiger, D. and J.H. Stock (1997): ”Instrumental Variable Regression with Weak Instruments,” *Econometrica* 65, 557-586.

[76] Stock, J.H., and J.H. Wright (2000): ”GMM with Weak Identification”, *Econometrica* 68, 1055-1096.

[77] White, H. (1982): ”Instrumental Variable Regression with Independent Observations,” *Econometrica* 50, 483-499.

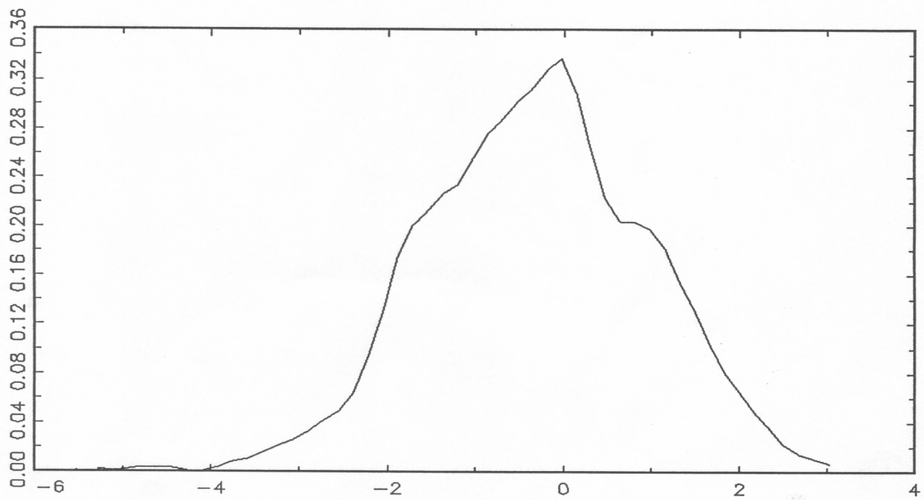


Fig. 1. Distribution of T-Statistic

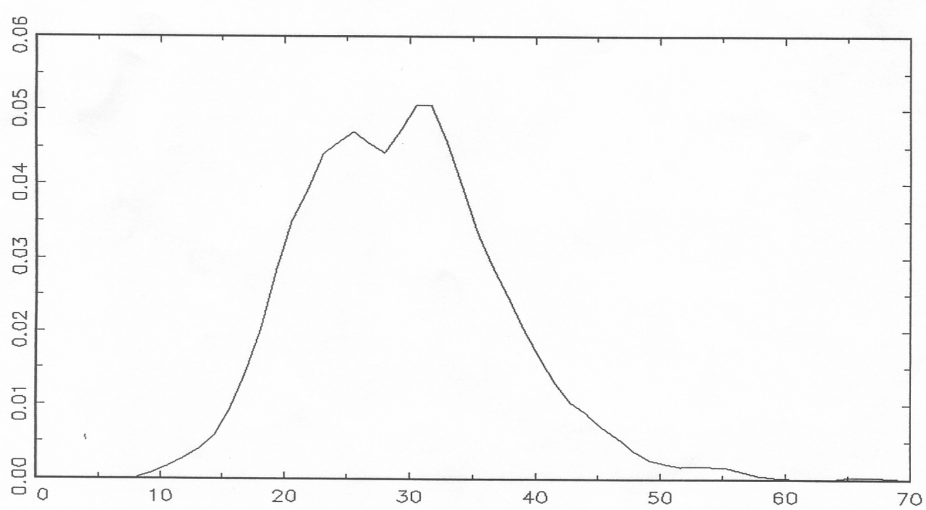


Fig. 2. Distribution of Overidentification Statistic

Figure 1: