

Finite Sample Properties of One-step, Two-step and Bootstrap Empirical Likelihood Approaches to Efficient GMM Estimation

Joachim Inkmann*

Department of Economics and Center of Finance and Econometrics (CoFE)
University of Konstanz, Box D124, 78457 Konstanz, Germany

Abstract: This paper compares conventional GMM estimators to empirical likelihood based GMM estimators which employ a semiparametric efficient estimate of the unknown distribution function of the data. One-step, two-step and bootstrap empirical likelihood and conventional GMM estimators are considered which are efficient for a given set of moment conditions. The estimators are subject to a Monte Carlo investigation using a specification which exploits sequential conditional moment restrictions for binary panel data with multiplicative latent effects. Among other findings the experiments show that the one-step and two-step estimators yield coverage rates of confidence intervals below their nominal coverage probabilities. The bootstrap methods improve upon this result.

Key words: GMM; empirical likelihood; bootstrap; sequential moment restrictions

JEL class.: C33, C35

* Joachim.Inkmann@uni-konstanz.de. Financial support by the *Deutsche Forschungsgemeinschaft* (DFG) is gratefully acknowledged. I wish to thank Michael Lechner and Winfried Pohlmeier for helpful suggestions.

1. Introduction

Economic theories frequently imply moment conditions of the form $E[\psi(Z, \theta_0)] = 0$ where $\psi(Z, \theta)$ is a $r \times 1$ vector of moment functions and θ is an unknown $q \times 1$ parameter vector of interest with true value θ_0 . Z denotes a random vector including both dependent and explanatory variables with a joint probability distribution function. Well known examples of such moment conditions can be found in the empirical literature on asset pricing theories (cf. e.g. Tauchen, 1986, and Kocherlakota, 1990). Suppose the data $\{Z_i : i = 1, \dots, n\}$ consist of a random sample of Z . Throughout the paper it is assumed that the data are independent and identically (i.i.d.) distributed according to some unknown distribution function $F_0(z)$. The interest focuses on the estimation of θ_0 using sample information and knowledge about the population moment condition. Rewrite the moment restriction as

$$R(F_0, \theta_0) = \int \psi(z, \theta_0) dF_0(z) = 0. \quad (1.1)$$

By application of the analogy principle (cf. Manski, 1988, ch. 1.2), an estimate $\hat{\theta}$ of θ_0 can be obtained by substituting the unknown distribution function $F_0(z)$ with the empirical distribution function $F_n(z)$. The latter is ‘feasible’ if it provides a solution to the equation

$$R(F_n, \hat{\theta}) = 0 \quad (1.2)$$

subject to $\theta \in \Theta$, where Θ is some space of possible parameter values. A prominent example for (1.2) is the method of moments (MM) estimator as a special case of the generalized method of moments principle (GMM; cf. Hansen, 1982) with $r = q$ moment functions. It is well known that (1.2) generally can not be solved in the presence of overidentifying restrictions, i.e. $r > q$. However, in this case a direct extension of the analogy principle (cf. Manski, 1988, ch. 1.2.2) allows solving

$$\hat{\theta} = \arg \min_{\theta \in \Theta} d(R(F_n, \theta)), \quad (1.3)$$

where $d(\cdot)$ maps values of $R(\cdot, \cdot)$ into the non-negative real half-line. The GMM estimator is the best known example for (1.3). Alternatively, Manski (1988, ch. 1.2.1) suggests solving

$$R(\pi(F_n), \hat{\theta}) = 0 \quad (1.4)$$

where the function $\pi(\cdot)$ projects the empirical distribution on the space of feasible distribution functions. This procedure has recently attracted much interest in GMM literature. In particular, an estimate $\pi(F_n) \equiv \hat{F}_n(z)$ of the distribution function $F_0(z)$ was developed by Back and Brown (1993), Qin and Lawless (1994), Brown and Newey (1995, 1998) and Imbens (1997) which is feasible in the sense of providing a solution to (1.4) and embodies the semiparametric restriction (1.1). The latter classifies $\hat{F}_n(z)$ as a semiparametric estimate of $F_0(z)$ in contrast to the nonparametric distribution estimate $F_n(z)$. The estimate $\hat{F}_n(z)$ results from different approaches and the references given above follow different routes to introduce this distribution function estimate. For example, Imbens (1997) shows that $\hat{F}_n(z)$ is implied by ML estimation of the points of finite support of a discrete multinomial distribution characterizing a sample analog of the moment condition (1.1). The finite support is not restrictive because any distribution function $F_0(z)$ can be approximated arbitrarily well by a multinomial distribution. Back and Brown (1993) show that $\hat{F}_n(z)$ is the implied distribution function estimate of efficient GMM estimators and Brown and Newey (1998) introduce $\hat{F}_n(z)$ in the context of semiparametric estimation of expectations. This paper follows Qin and Lawless (1994), Brown and Newey (1995) and Imbens (1997) and presents the empirical likelihood approach to $\hat{F}_n(z)$ which has a particularly simple interpretation: $\hat{F}_n(z)$ is the discrete multinomial distribution with n support points which has the highest probability of generating the observed sample subject to a sample counterpart of the moment condition (1.1).

The reason for considering $\hat{F}_n(z)$ in combination with (1.4) as an alternative to the usual GMM approach (1.3) is the semiparametric efficiency of $\hat{F}_n(z)$ in the class of regular estimators accounting for the moment condition (1.1). One might expect that this efficiency advantage of the distribution estimator carries over to the resulting parameter estimate. However, the semiparametric efficiency bound for estimators exploiting moment conditions of the form (1.1) as the only distributional assumption has been established by Chamberlain (1987) and it is well known that a GMM estimator using an optimal weight matrix attains this bound. Indeed, this efficient GMM estimator and the estimator solving (1.4) with $\pi(F_n) = \hat{F}_n(z)$ share the same first order asymptotic properties. Nevertheless, Brown and Newey (1998) conjecture that efficiency gains of higher order for the parameters of interest could be realized by using the efficient estimate $\pi(F_n) = \hat{F}_n(z)$. In addition, they show that any expectation

$$T(F_0, \theta_0) = \int m(z, \theta_0) dF_0(z). \quad (1.5)$$

can be efficiently estimated subject to the semiparametric restriction (1.1) by

$$T(\hat{F}_n, \hat{\theta}) \tag{1.6}$$

given the semiparametric efficiency of $\hat{\theta}$. This property will be used later with (1.5) being the optimal weight matrix for GMM estimators. The estimate (1.6) of this weight matrix is suggested by Back and Brown (1993) and Brown and Newey (1998).

This paper compares the finite sample properties of three versions of the conventional GMM and empirical likelihood based GMM (GMM_EL) estimators in the presence of overidentifying restrictions. All have in common that they reach the semiparametric efficiency bound for given moment conditions (1.1). The first pair of estimators are two-step estimators solving (1.3) in a second step with a first step estimate of the optimal weight matrix. The GMM estimator uses an estimate of this weight matrix based on the empirical distribution function $F_n(z)$, the GMM_EL estimator rests on $\hat{F}_n(z)$ using an estimate of the form (1.6). The second pair of estimators are one-step estimators solving (1.3) in the case of GMM and (1.4) in the case of GMM_EL. The GMM estimator is the continuous updating estimator introduced by Hansen, Heaton and Yaron (1996). The third pair consists of bootstrap estimators solving (1.3) where the nonparametric and semiparametric distribution estimates $F_n(z)$ and $\hat{F}_n(z)$ describe the respective resampling probabilities for the GMM and GMM_EL bootstrap estimators. These bootstrap approaches were introduced by Hall and Horowitz (1996; GMM) and Brown and Newey (1995; GMM_EL).

The one-step and bootstrap alternatives to the usual two-step GMM estimator are attractive as possible solutions to the well known small sample shortcomings of the two-step GMM estimator. Summarizing the small sample evidence obtained by Tauchen (1986), Kocherlakota (1990), Ferson and Foerster (1994), Hansen, Heaton and Yaron (1996) for asset pricing models, Altonji and Segal (1996) and Clark (1996) for covariance structures, and Arellano and Bond (1991), Ziliak (1997), and Blundell and Bond (1998) for dynamic panel data models, a number of finite sample problems appear to be very robust: The bias of the two-step GMM estimator increases with the number of overidentifying restrictions, the coverage probabilities of confidence intervals could be heavily distorted and the size of the J test may deviate from its nominal value. The one-step approaches could solve the first problem because results from Altonji and Segal (1996) suggest that the weight matrix estimate introduces a correlation between the moment functions and the weight matrix which creates finite sample

bias. The one-step approaches circumvent the weight matrix estimation step and therefore solve this source of potential bias. In addition, the one-step continuous updating GMM estimator is proven to be consistent under Bekker's (1994) large instruments asymptotics which renders the two-step GMM estimator inconsistent (in the framework of a linear simultaneous equation model with conditional homoskedasticity). The bootstrap methods may improve upon the inference in small samples because they provide asymptotic refinements for the coverage probabilities of confidence intervals and the size of the J test of overidentifying restrictions (cf. Hall and Horowitz, 1996). However, the motivation of the bootstrap is a pure asymptotic one and small sample experiments are necessary to evaluate the bootstrap based inference in comparison to inference based on conventional first order asymptotic theory for GMM estimation.

Hence, the aims of this paper are twofold: One the on hand it compares conventional and empirical likelihood approaches to efficient GMM estimation, on the other hand it provides evidence on the relative performance of one-step, two-step and bootstrap estimators. This is done by means of a Monte Carlo investigation using a specification suggested by Wooldridge (1997) which exploits sequential conditional moment restrictions for binary panel data with multiplicative latent effects. The Monte Carlo experiments suggest that the empirical likelihood based two-step GMM estimator may improve upon the reliability of the J test of overidentifying restrictions while the bootstrap methods are recommended for obtaining more reliable coverage rates of symmetric confidence intervals which are much too small if they are based on conventional asymptotic theory. The one-step continuous updating GMM estimator exhibits fat tails which prevents an useful application while the one-step empirical likelihood estimator performs similar to the conventional two-step GMM estimator.

The outline of the paper is as follows. Section 2 introduces the nonparametric distribution estimate $F_n(z)$ of $F_0(z)$ and the resulting one-step, two-step and bootstrap GMM estimators. Section 3 derives the semiparametric distribution estimate $\hat{F}_n(z)$ of $F_0(z)$ and describes the corresponding one-step, two-step and bootstrap GMM_EL estimators. Section 4 presents the Monte Carlo experiments and Section 5 concludes.

2. Conventional approaches to efficient GMM estimation

2.1 Nonparametric distribution estimation

Conventional GMM estimators apply the analogy principle (1.3) using the empirical distribution function

$$F_n(z) = \frac{1}{n} \sum_{i=1}^n 1(Z_i \leq z) \quad (2.1.1)$$

as a nonparametric estimate of $F_0(z)$, where $1(\cdot)$ denotes the indicator function. F_n is a discrete distribution function which places equal probability $1/n$ on each observation such that the sample equivalent of the population moment condition becomes a sample average.

2.2 Two-step GMM estimation

The GMM estimator $\hat{\theta}$ of the unknown parameter vector θ_0 is defined as the vector minimizing the objective function

$$\hat{J}_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta) \right)' \hat{W} \left(\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta) \right) \quad (2.2.1)$$

subject to $\theta \in \Theta$, where Θ denotes the set of possible parameter values and \hat{W} is a positive semidefinite weight matrix of dimension $r \times r$ which converges in probability to W . Note that (2.2.1) is a special case of (1.3). Under regularity conditions (cf. Newey and McFadden, 1994) the GMM estimator is consistent and the asymptotic distribution of the stabilizing transformation is normal.

The Cramér-Rao efficiency bound for estimators using (1.1) as the only substantive distributional assumption is derived by Chamberlain (1987) as

$$\Lambda = (G_0' V_0^{-1} G_0)^{-1} \quad (2.2.2)$$

$$\text{with } V_0 = E[\psi(Z, \theta_0) \psi(Z, \theta_0)'] \quad \text{and} \quad G_0 = E \left[\frac{\partial \psi(Z, \theta_0)}{\partial \theta'} \right].$$

A necessary and sufficient condition for the GMM estimator attaining the lower bound Λ is $G_0' W = F \cdot G_0' V_0^{-1}$ for any nonsingular matrix F (cf. Hansen, 1982). The necessary condition $W = V_0^{-1}$ is the origin of the usual two-step GMM estimation principle which consists of minimizing (2.2.1) using a parameter independent weight matrix such as $\hat{W} = I_r$ in a first

step, computing a weight matrix with the first step estimates which converges in probability to $W = V_0^{-1}$, and finally minimizing (2.2.1) again using this optimal weight matrix. Replacing the population moment V_0 by a sample equivalent based on (2.1.1) and noting the continuity of matrix inversion, a consistent estimate $\hat{W} = \hat{V}_1^{-1}$ of the optimal weight matrix is

$$\hat{V}_1^{-1} = \left[\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \hat{\theta}_1) \psi(Z_i, \hat{\theta}_1)' \right]^{-1}, \quad (2.2.3)$$

where $\hat{\theta}_1$ is the consistent first step GMM estimate. Denote the second step GMM estimate as $\hat{\theta}_2$. A consistent estimate $\hat{\Lambda}$ of the asymptotic variance-covariance matrix of $\hat{\theta}_2$ is obtained by replacing V_0^{-1} in (2.2.2) with \hat{V}_1^{-1} and G_0 with its corresponding sample moment evaluated at $\hat{\theta}_2$. Choosing $\hat{W} = \hat{V}_1^{-1}$, $n \cdot \hat{J}_n(\hat{\theta}_2)$ is asymptotically χ_{r-q}^2 distributed suggesting a test of the overidentifying restrictions (cf. Hansen, 1982) which has become known as the J test.

2.3 One-step GMM estimation

Hansen, Heaton and Yaron (1996) introduce the continuous updating GMM estimator which results from altering the optimal weight matrix in each iteration step to embody the restrictions of the model. The GMM objective function is modified to

$$\hat{J}_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta) \right)' \left[\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta) \psi(Z_i, \theta)' \right]^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \theta) \right) \quad (2.3.1)$$

which follows immediately from combining (2.2.1) with (2.2.3) evaluated at θ instead of $\hat{\theta}_1$. The objective function is again a special case of (1.3). The authors point out that the asymptotic distribution of the GMM estimator remains unchanged by this modification although the first order conditions for a minimum of (2.3.1) contain an additional derivative term for the parameter dependent weight matrix.

For systems of linear simultaneous equations under conditional homoskedasticity the continuous updating estimator becomes the limited information maximum likelihood (LIML) estimator with its known advantages over the two-stage least squares (2SLS) estimator which is the linear simultaneous equations model counterpart of the two-step GMM estimator. In particular, the LIML estimator remains consistent under Bekker's (1994) large instruments asymptotic theory which renders 2SLS inconsistent. Therefore one might argue that the continuous updating estimator could outperform the two-step GMM estimator if the number of

overidentifying restrictions is large. Note that by definition of (2.3.1) the corresponding J test is more conservative than the J test based on the usual two-step GMM estimator. Hansen, Heaton and Yaron (1996) report results from Monte Carlo experiments which indicate an improved size performance of the continuous updating J test over the conventional J test.

2.4 Bootstrap GMM estimation

The bootstrap is a resampling method for estimating the distribution of an estimator or statistic. The bootstrap method treats the sample data as if they were the population and estimates the distribution of interest using the empirical distribution of the relevant estimator or test statistic generated by randomly resampling the sample data. The reasons for using the bootstrap are twofold: On the one hand the bootstrap offers a simple way to compute the distribution of estimators or test statistics in those cases in which an analytical derivation or approximation is difficult. On the other hand the bootstrap often provides a more accurate approximation of the distribution of interest than the usual approximation obtained from first order asymptotic theory. The latter argument is particularly well documented in Horowitz (1997) who presents some examples in which the bootstrap yields asymptotic refinements.

The application of the bootstrap to overidentified GMM estimators is affected by one serious problem: The GMM principle rests on the main assumption that the estimation data $Z_d = \{Z_i : i = 1, \dots, n\}$ is a random sample of the population distribution of the random vector Z which satisfies the orthogonality condition $E[\psi(Z, \theta_0)] = 0$. The bootstrap treats Z_d as if it were the population and draws random samples $Z_b = \{Z_i^b : i = 1, \dots, n\}$ from Z_d with replacement by placing probability $1/n$ on each observation. Thus, the bootstrap does not implement a semiparametric restriction on Z_d which corresponds to the orthogonality condition under bootstrap sampling. In other words, the bootstrap would impose a moment condition which does not hold in the population from which the bootstrap samples. As a consequence, the bootstrap either does not improve upon conventional first order asymptotic approximations or does even worse. As far as the estimation of confidence intervals is concerned, the bootstrap produces the same approximation error of the coverage probability as the asymptotic theory as shown by Brown and Newey (1995) for the bootstrap-t method and by Hahn (1996) for the percentile method. Regarding the bootstrap estimate of the critical value of the J test of overidentifying restrictions, the bootstrap produces the wrong size, even asymptotically (cf. Brown and Newey, 1995). These problems would be solved if the bootstrap imposed

a moment condition on the original sample which corresponds to the population orthogonality condition. Using the two-step GMM estimate $\hat{\theta}_2$ as the sample counterpart of the population parameter θ_0 , the bootstrap counterpart of the orthogonality condition can be written as

$$E_B[\psi(Z, \hat{\theta}_2)] = 0 \quad \Leftrightarrow \quad \bar{\psi}_2 \equiv \frac{1}{n} \sum_{i=1}^n \psi(Z_i, \hat{\theta}_2) = 0, \quad (2.4.1)$$

where $E_B[\cdot]$ denotes the expectation under bootstrap sampling. Obviously, (2.4.1) generally does not hold in the presence of overidentifying restrictions. However, (2.4.1) suggests re-centering the original moment functions around their sample mean $\bar{\psi}_2$ to implement a sample orthogonality condition. This procedure was proposed by Hall and Horowitz (1996) and implies the following recentered moment functions

$$\psi^*(Z_i, \theta) \equiv \psi(Z_i, \theta) - \bar{\psi}_2 \quad (2.4.2)$$

which, evaluated at $\hat{\theta}_2$, satisfy the bootstrap counterpart

$$E_B[\psi^*(Z, \hat{\theta}_2)] = 0 \quad (2.4.3)$$

of the population orthogonality condition $E[\psi(Z, \theta_0)] = 0$. Hence, for any bootstrap sample Z_b the bootstrap version of the two-step GMM estimator solves

$$\hat{\theta}_{1b} = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n \psi^*(Z_i^b, \theta) \right)' \hat{W} \left(\frac{1}{n} \sum_{i=1}^n \psi^*(Z_i^b, \theta) \right) \quad (2.4.4)$$

in the first estimation step using some parameter independent weight matrix \hat{W} . For the second estimation step the optimal weight matrix is computed according to (2.2.3) as

$$\hat{V}_b^{-1} = \left[\frac{1}{n} \sum_{i=1}^n \psi^*(Z_i^b, \hat{\theta}_{1b}) \psi^*(Z_i^b, \hat{\theta}_{1b})' \right]^{-1}. \quad (2.4.5)$$

Finally, the bootstrap version of the second step GMM estimator minimizes

$$\hat{J}_b(\theta) = \left(\frac{1}{n} \sum_{i=1}^n \psi^*(Z_i^b, \theta) \right)' \hat{V}_b^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi^*(Z_i^b, \theta) \right) \quad (2.4.6)$$

and yields the bootstrap GMM estimate $\hat{\theta}_b$ in replication b. Denote the $r \times q$ Jacobian matrix

of the recentered moments by $G^*(Z, \theta) = [\partial \psi^*(Z, \theta) / \partial \theta]$. Then the asymptotic variance-covariance matrix of the stabilizing transformation of $\hat{\theta}_b$ is consistently estimated by

$$\hat{\Lambda}_b = \left[\left(\frac{1}{n} \sum_{i=1}^n G^*(Z_i^b, \hat{\theta}_b) \right)' \hat{V}_b^{-1} \left(\frac{1}{n} \sum_{i=1}^n G^*(Z_i^b, \hat{\theta}_b) \right) \right]^{-1}. \quad (2.4.7)$$

Replicating the estimation steps (2.4.4) – (2.4.7) B times generates an empirical distribution function of the relevant statistics from which the bootstrap estimates can be derived. The following paragraphs discuss the bootstrap estimate of the bias, the bootstrap estimate of symmetric confidence intervals and the bootstrap estimate of the size of the J test of overidentifying restrictions. In all cases the bootstrap treats the estimation data as if it were the population and therefore replaces the population parameter vector θ_0 with the sample estimate $\hat{\theta}_2$ and the latter with $\hat{\theta}_b$. Hence, the bootstrap estimate of the bias is defined as

$$\hat{b}_B = \bar{\hat{\theta}}_B - \hat{\theta}_2, \quad \text{where} \quad \bar{\hat{\theta}}_B = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b \quad (2.4.8)$$

is the expected parameter estimate under bootstrap sampling because the empirical distribution of the $\hat{\theta}_b$ places probability $1/B$ on each estimate. Using (2.4.8) a bootstrap estimated bias corrected estimate $\hat{\theta}_2^c \equiv \hat{\theta}_2 - \hat{b}_B = 2\hat{\theta}_2 - \bar{\hat{\theta}}_B$ is readily available (cf. Horowitz, 1998).

For the bootstrap estimation of confidence intervals two different approaches can be distinguished by their respective representation of the coverage probability. Let $\hat{t} = (\hat{\theta}_2^k - \theta_0^k) / \hat{s}^k$ denote the t statistic based on the k th element of the two-step GMM estimator with \hat{s}^k , the k th diagonal element of $(\hat{\Lambda}/n)^{1/2}$. The optimal situation for statistical inference would be described by the knowledge of the quantile t^α such that $\Pr(\hat{t} \leq t^\alpha) = 1 - \alpha$ holds exactly in small samples. However, t^α is not known and is therefore usually replaced with the $1 - \alpha/2$ quantile $z^{\alpha/2}$ of the standard normal distribution which is the limiting distribution of the t statistic using asymptotic theory. The bootstrap provides alternative estimates of t^α . Rewrite the coverage probability of the exact confidence interval as

$$\Pr(-t^\alpha \leq (\hat{\theta}_2^k - \theta_0^k) / \hat{s}^k \leq t^\alpha) = 1 - \alpha \quad (2.4.9)$$

$$\Leftrightarrow \Pr(-t^\alpha \cdot \hat{s}^k \leq \hat{\theta}_2^k - \theta_0^k \leq t^\alpha \cdot \hat{s}^k) = 1 - \alpha,$$

$$\Leftrightarrow \Pr(-\theta^\alpha \leq \hat{\theta}_2^k - \theta_0^k \leq \theta^\alpha) = 1 - \alpha, \quad (2.4.10)$$

with $\theta^\alpha = t^\alpha \cdot \hat{s}^k$. The first bootstrap method for the estimation of confidence intervals is based on (2.4.10) and replaces the unknown distribution of $\hat{\theta}_2^k - \theta_0^k$ with the empirical distribution of the corresponding bootstrap statistic $\hat{\theta}_b^k - \hat{\theta}_2^k$. Thus, the bootstrap estimate of θ^α can be derived as the $1 - \alpha$ quantile

$$\hat{\theta}_B^\alpha = F_n^{-1}(1 - \alpha), \quad \text{where} \quad F_n(\theta) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(|\hat{\theta}_b^k - \hat{\theta}_2^k| \leq \theta) \quad (2.4.11)$$

is the relevant empirical distribution function generated by the bootstrap and $F_n^{-1}(\cdot)$ its inverse. The corresponding bootstrap estimate I_B of the confidence interval for θ_0^k with nominal coverage probability $1 - \alpha$ results from (2.4.10) as

$$I_B = (\hat{\theta}_2^k - \hat{\theta}_B^\alpha, \hat{\theta}_2^k + \hat{\theta}_B^\alpha). \quad (2.4.12)$$

This procedure is known as the percentile approach to bootstrap confidence intervals (cf. e.g. Efron and Tibshirani, 1993, Section 13.3). The coverage error of this confidence interval for θ_0^k defined as the difference between the true and nominal coverage probability, $\Pr(\theta_0 \in I_B) - (1 - \alpha)$, has the same size as the coverage error of the confidence interval based on first order asymptotic theory as point out by Hall and Horowitz (1996) and Horowitz (1998). Thus, the bootstrap does not yield an asymptotic refinement in this case. Hahn (1996) proves that this result holds as well for an uncentered version of the GMM bootstrap which uses the uncentered moment functions $\psi^*(Z, \theta) \equiv \psi(Z, \theta)$ throughout the estimation steps (2.4.4) – (2.4.7).

However, recentering the moment functions becomes necessary in order for the second bootstrap approach to the estimation of confidence intervals to achieve asymptotic refinements upon asymptotic theory. This method is known as the bootstrap-t (or percentile-t) method (cf. e.g. Efron and Tibshirani, 1993, Section 12.5) and is based on an approximation to (2.4.9). The bootstrap-t method replaces the unknown distribution of $\hat{t} = (\hat{\theta}_2^k - \theta_0^k) / \hat{s}^k$ with the empirical distribution of the corresponding bootstrap statistic $\hat{t}_b = (\hat{\theta}_b^k - \hat{\theta}_2^k) / \hat{s}_b^k$ where \hat{s}_b^k is the k th diagonal element of $(\hat{\Lambda}_b / n)^{1/2}$. Thus, the bootstrap estimate of the exact critical value t^α can be derived as the $1 - \alpha$ quantile

$$\hat{t}_B^\alpha = F_n^{-1}(1 - \alpha), \quad \text{where} \quad F_n(t) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(|\hat{t}_b| \leq t) \quad (2.4.13)$$

is the relevant empirical distribution function generated by the bootstrap. The corresponding bootstrap estimate I_B of the confidence interval for θ_0^k with nominal coverage probability $1 - \alpha$ results from (2.4.9) as

$$I_B = \left(\hat{\theta}_2^k - \hat{t}_B^\alpha \cdot \hat{s}^k, \hat{\theta}_2^k + \hat{t}_B^\alpha \cdot \hat{s}^k \right). \quad (2.4.14)$$

Hall and Horowitz (1996, Theorem 3) show that the coverage error of the symmetric bootstrap-t confidence interval is $o(n^{-1})$ and therefore smaller than the size $O(n^{-1})$ of the confidence interval which uses the asymptotic approximation $z^{\alpha/2}$ of t^α . Hall (1992, chap. 3.6) and Horowitz (1997, 1998) point out that the coverage error of the bootstrap-t confidence interval is usually of the order $O(n^{-2})$. Hall shows that this result depends on the symmetry of the bootstrap-t confidence interval. A two-sided equal-tailed bootstrap-t confidence interval does not improve upon the asymptotic approximation of the coverage probability.

Horowitz (1997, 1998) explains the superiority of the bootstrap-t method over the percentile method in the sense of providing an asymptotic refinement by the fact that the former method samples the asymptotically pivotal statistic $(\hat{\theta}_2^k - \theta_0^k)/\hat{s}^k$ while the latter method samples the statistic $\hat{\theta}_2^k - \theta_0^k$ which converges to a limiting distribution which depends on unknown population parameters.

The J test statistic $n \cdot \hat{J}_n(\hat{\theta}_2)$ is asymptotically pivotal as well. Hence, it is not surprising that the bootstrap improves upon the accuracy of the asymptotic approximation of the exact rejection probability $\Pr(n \cdot \hat{J}_n(\hat{\theta}_2) > J_\alpha) = \alpha$. The conventional J test replaces the unknown critical value J_α with the $1 - \alpha$ quantile of the χ_{r-q}^2 distribution which is the limiting distribution of the test statistic $n \cdot \hat{J}_n(\hat{\theta}_2)$ using first order asymptotic theory. The bootstrap-J method replaces J_α with the $1 - \alpha$ quantile

$$\hat{J}_B^{1-\alpha} = F_n^{-1}(1 - \alpha), \quad \text{where} \quad F_n(J) = \frac{1}{B} \sum_{b=1}^B 1(n \cdot \hat{J}_b(\hat{\theta}_b) \leq J) \quad (2.4.15)$$

is the relevant empirical distribution function. Hall and Horowitz (1996, Theorem 3) prove that the size approximation error of the bootstrap is of order $o(n^{-1})$ and therefore converges faster to zero than the size approximation error using the critical value implied by asymptotic theory which is of order $O(n^{-1})$.

Hall and Horowitz (1996) report the results of some Monte Carlo experiments using a data generating process which resembles an asset pricing model with a single overidentifying

restriction. For sample sizes of 50 and 100 the empirical levels of the conventional t test and J test turn out to be much larger than their nominal values. The bootstrap-t and bootstrap-J methods usually reduce these approximation errors without completely eliminating the small sample size distortions. Further small sample evidence on the performance of the Hall and Horowitz GMM bootstrap appears to be very limited. One exception is the work by Bergström, Dahlberg and Johansson (1997) who also find an improved size performance of the Hall and Horowitz bootstrap-J test over the conventional J test based on asymptotic theory. Horowitz (1998) presents Monte Carlo experiments for the finite sample performance of confidence intervals obtained by the bootstrap-t method. Using the data generating process analyzed before by Altonji and Segal (1996) for the estimation of covariance structures he finds a substantial improvement of the empirical coverage probability over conventional confidence intervals.

One question which has not been addressed in this section concerns the choice of the number of bootstrap replications B . Horowitz (1998) recommends increasing B until a further increase has no further impact on the bootstrap statistics of interest. This principle requires repeated computation of these statistics and a stopping rule which defines ‘no impact’ in a mathematical sense. Andrews and Buchinsky (1997) suggest a three-step method that yields such a stopping rule in terms of an approximate percentage deviation of the bootstrap estimate from the ideal bootstrap estimate with an infinite number of replications. They provide some Monte Carlo evidence that points in favor of their method. However, in most applications of the bootstrap the number of replications is chosen ad hoc. The above mentioned applications of the Hall and Horowitz GMM bootstrap method rely on numbers between 100 and 500 replications.

3. Empirical likelihood approaches to efficient GMM estimation

3.1 Semiparametric distribution estimation

The empirical likelihood principle introduced by Owen (1988, 1990) and applied to GMM by Qin and Lawless (1994) and Imbens (1997) is based on the optimization program

$$\max_{\pi, \theta} \sum_{i=1}^n \ln \pi_i \quad \text{s.t.} \quad \sum_{i=1}^n \pi_i = 1, \quad \pi_i \geq 0, \quad (3.1.1)$$

$$\sum_{i=1}^n \pi_i \psi(Z_i, \theta) = 0, \quad (3.1.2)$$

with $\pi = (\pi_1, \dots, \pi_n)'$. Solving this $(n + q)$ dimensional optimization problem implies searching for a discrete probability distribution function which places probability π_i on observation i and guarantees that the sample version (3.1.2) of the moment condition (1.1) is satisfied. Note that solving (3.1.1) without noting (3.1.2) yields the estimate $\hat{\pi}_i = \frac{1}{n}$ of π_i and implies the empirical distribution function (2.1.1). The optimization problem can be solved by using a Lagrange approach. Let γ denote a scalar Lagrange parameter associated with the first restriction in (3.1.1) and λ be a $r \times 1$ vector of Lagrange multipliers associated with restriction (3.1.2). Then the Lagrange function to be maximized over $\pi, \theta, \gamma, \lambda$ can be written as

$$L = \sum_{i=1}^n \ln \pi_i + \gamma \left(1 - \sum_{i=1}^n \pi_i \right) - n \lambda' \sum_{i=1}^n \pi_i \psi(Z_i, \theta) \quad (3.1.3)$$

and implies the following first order conditions for the empirical likelihood estimates

$$\frac{\partial L}{\partial \gamma} = 0 \quad \Rightarrow \quad \sum_{i=1}^n \hat{\pi}_i^{\text{el}} = 1,$$

$$\frac{\partial L}{\partial \lambda} = 0 \quad \Rightarrow \quad \sum_{i=1}^n \hat{\pi}_i^{\text{el}} \psi(Z_i, \hat{\theta}_{\text{el}}) = 0, \quad (3.1.4)$$

$$\frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \quad \hat{\lambda}'_{\text{el}} \sum_{i=1}^n \hat{\pi}_i^{\text{el}} \frac{\partial \psi(Z_i, \hat{\theta}_{\text{el}})}{\partial \theta} = 0, \quad (3.1.5)$$

$$\sum_{i=1}^n \pi_i \frac{\partial L}{\partial \pi_i} = 0 \quad \Rightarrow \quad \hat{\gamma}_{\text{el}} = n,$$

$$\frac{\partial L}{\partial \pi_i} = 0 \quad \Rightarrow \quad \hat{\pi}_i^{\text{el}} = \frac{1}{n(1 + \hat{\lambda}'_{\text{el}} \psi(Z_i, \hat{\theta}_{\text{el}}))}. \quad (3.1.6)$$

The resulting semiparametric distribution estimate places probability $\hat{\pi}_i^{\text{el}}$ on each observation

$$\hat{F}_n^{\text{el}}(z) = \sum_{i=1}^n \hat{\pi}_i^{\text{el}} 1(Z_i \leq z). \quad (3.1.7)$$

The probabilities $\hat{\pi}_i^{\text{el}}$ have a simple interpretation. From (3.1.6) it is obvious that the weights decrease with an increasing estimated Lagrange parameter $\hat{\lambda}_{\text{el}}$ or alternatively, with an increasing departure of the sample moment condition from zero. Substitution of (3.1.6) into the first order conditions (3.1.4) and (3.1.5) of the Lagrange approach suggest an alternative way of obtaining empirical likelihood estimates of the parameters of interest θ and λ by a just identified moment estimator

$$\sum_{i=1}^n \psi_{el}(\mathbf{z}_i, \hat{\theta}_{el}, \hat{\lambda}_{el}) = 0 \quad (3.1.8)$$

$$\text{with } \psi_{el}(\mathbf{z}, \theta, \lambda) = \left(\begin{array}{c} \psi(\mathbf{z}, \theta)/(1 + \lambda' \psi(\mathbf{z}, \theta)) \\ \lambda' (\partial \psi(\mathbf{z}, \theta)/\partial \theta)/(1 + \lambda' \psi(\mathbf{z}, \theta)) \end{array} \right).$$

This allows reducing the number of unknown parameters from $(n + q)$ in the original optimization program to $(r + q)$. Imbens (1997) shows that the estimated Lagrange parameters converge in probability to zero and $\hat{\theta}_{el}$ shares the first order asymptotic properties of the conventional two-step GMM estimator $\hat{\theta}_2$ and is therefore semiparametric efficient for given moment restrictions (1.1).

The latter results suggest an alternative formulation of the empirical likelihood approach by replacing the unknown θ in (3.1.2) with the two-step GMM estimate $\hat{\theta}_2$ and optimizing (3.1.1) only with respect to π . This approach was suggested by Brown and Newey (1995) and will be referred to as modified empirical likelihood (subscript elm) in the following. The Lagrange function is altered correspondingly which eliminates (3.1.5) and simplifies (3.1.8) to

$$\sum_{i=1}^n \psi_{elm}(\mathbf{z}_i, \hat{\theta}_2, \hat{\lambda}_{elm}) = 0$$

$$\text{with } \psi_{elm}(\mathbf{z}, \hat{\theta}_2, \lambda) = \psi(\mathbf{z}, \hat{\theta}_2)/(1 + \lambda' \psi(\mathbf{z}, \hat{\theta}_2)). \quad (3.1.9)$$

The corresponding semiparametric distribution estimate relies on probabilities $\hat{\pi}_i^{elm}$ of the form (3.1.6) with $(\hat{\theta}_{el}, \hat{\lambda}_{el})$ replaced by $(\hat{\theta}_2, \hat{\lambda}_{elm})$. The moment function (3.1.9) can be thought of as being the first order condition to the optimization problem

$$\hat{\lambda}_{elm} = \arg \max_{\lambda} \sum_{i=1}^n \ln(1 + \lambda' \psi(\mathbf{z}_i, \hat{\theta}_2)), \quad (3.1.10)$$

s.t. $1 + \lambda' \psi(\mathbf{z}_i, \hat{\theta}_2) > 0$ which was proposed by Brown and Newey. They show that the modified empirical likelihood estimates in (3.1.9) and (3.1.10) describe just one special case of a general class of semiparametric distribution estimates of the form

$$\hat{\pi}_i = \frac{\nabla T(\hat{\lambda}' \psi(\mathbf{z}_i, \hat{\theta}^*))}{\sum_{j=1}^n \nabla T(\hat{\lambda}' \psi(\mathbf{z}_j, \hat{\theta}^*))} \quad \text{with} \quad \hat{\lambda} = \arg \max_{\lambda} \sum_{i=1}^n T(\lambda' \psi(\mathbf{z}_i, \hat{\theta}^*)), \quad (3.1.11)$$

where $T(v)$ is a differentiable concave function with scalar argument v and with domain that is an open interval containing zero, $\nabla T(v) = dT(v)/dv$, and $\hat{\theta}^*$ is any semiparametric effi-

cient parameter estimate. (3.1.11) includes the (modified) empirical likelihood estimator with $(\hat{\theta}^* = \hat{\theta}_2)$ $\hat{\theta}^* = \hat{\theta}_{el}$ and $T(v) = \ln(1 + v)$. A second example given by Brown and Newey which results from $T(v) = -\exp(v)$ yields the exponential tilting estimator considered by Imbens (1997), Imbens, Spady and Johnson (1998), and Kitamura and Stutzer (1997). A third example based on the choice $T(v) = -(1 + v)^2$ is particularly convenient because it leads to a closed form solution for $\hat{\lambda}$. In this case, with $\hat{\theta}^* = \hat{\theta}_2$, the first order conditions for λ from (3.1.11) imply the estimated Lagrange parameters¹

$$\hat{\lambda} = -\hat{V}_2^{-1}\bar{\psi}_2 \quad (3.1.12)$$

$$\text{with } \hat{V}_2^{-1} = \left[\frac{1}{n} \sum_{i=1}^n \psi(Z_i, \hat{\theta}_2) \psi(Z_i, \hat{\theta}_2)' \right]^{-1}$$

and $\bar{\psi}_2$ defined in (2.4.1). The associated probabilities of the semiparametric distribution estimate follow from (3.1.11) as

$$\hat{\pi}_i = \frac{1 - \bar{\psi}_2' \hat{V}_2^{-1} \psi(Z_i, \hat{\theta}_2)}{n(1 - \bar{\psi}_2' \hat{V}_2^{-1} \bar{\psi}_2)} \quad (3.1.13)$$

and were in similar form (ignoring the term in parentheses in the denominator) also obtained from different approaches by Back and Brown (1993) and Brown and Newey (1998). These authors prove that under regularity conditions the resulting distribution function estimate is semiparametric efficient for given moment restrictions (1.1) which holds as well for any other probability estimates derived from (3.1.11) as shown by Brown and Newey (1995).

3.2 Two-step GMM_EL estimation

Back and Brown (1993) and Brown and Newey (1998) recommend using a semiparametric efficient estimate of the optimal weight matrix V_0^{-1} instead of the estimate (2.2.3) for the usual two-step GMM estimation procedure. Following (1.6) such an estimate requires a semiparametric efficient distribution estimate $\hat{F}_n(z)$ and an initial parameter estimate attaining the lower bound (2.2.2). Back and Brown (1993) and Brown and Newey (1998) use the semiparametric distribution estimate resulting from (3.1.13) and the two-step GMM estimate $\hat{\theta}_2$ as an initial estimate. Brown and Newey prove that the resulting estimate of the optimal weight

¹ These Lagrange parameter estimates are also obtained by the log Euclidean likelihood estimator considered by Owen (1991), Qin and Lawless (1994), and Imbens, Spady and Johnson (1998).

matrix is asymptotically efficient relative to the usual estimate (2.2.3) if the third moments of $\psi(Z, \theta_0)$ are not zero. Here, the modified empirical likelihood approach (3.1.9) is chosen to obtain the estimated probabilities $\hat{\pi}_i^{\text{elm}}$ and the two-step GMM estimates $\hat{\theta}_2$ as initial estimates. Thus, the semiparametric efficient estimate (1.6) of V_0^{-1} becomes

$$\hat{V}_3^{-1} = \left[\sum_{i=1}^n \hat{\pi}_i^{\text{elm}} \cdot \psi(Z_i, \hat{\theta}_2) \psi(Z_i, \hat{\theta}_2)' \right]^{-1} \quad (3.2.1)$$

and is used as a replacement for \hat{W} in (2.2.1). The asymptotic properties of the resulting parameter estimates as well as the J test remain unchanged from using \hat{V}_3^{-1} instead of \hat{V}_1^{-1} .

3.3 One-step GMM_EL estimation

The one-step GMM_EL estimator was already introduced in Section 3.1 and results from solving (3.1.8) which is a special case of the analogy principle (1.4) with $\pi(F_n) \equiv \hat{F}_n(z)$. The usual J test of overidentifying restrictions is not available in this case because the minimized objective function always attains zero. However, by definition of the Lagrange function (3.1.3), a test of the null hypothesis $H_0 : \lambda = 0$ provides a test of the overidentifying restrictions and is therefore an alternative to the J test in the conventional GMM framework. Contrary to the J test, this test procedure also allows testing a subset of the overidentifying restrictions. Imbens, Spady and Johnson (1998) suggest three different Lagrange multiplier (LM) tests of $H_0 : \lambda = 0$ which can be written as $LM = \hat{\lambda}' R \hat{\lambda}$ and only differ by the respective choice of R. These LM tests share the asymptotic χ_{r-q}^2 distribution of the J test. They compare these test statistics in some Monte Carlo experiments with the J tests based on the two-step and continuous updating GMM estimators and find that the LM test using

$$R = \left(\sum_{i=1}^n \psi(Z_i, \hat{\theta}) \psi(Z_i, \hat{\theta})' \hat{\pi}_i \right) \left[\sum_{i=1}^n \psi(Z_i, \hat{\theta}) \psi(Z_i, \hat{\theta})' \hat{\pi}_i^2 \right]^{-1} \left(\sum_{i=1}^n \psi(Z_i, \hat{\theta}) \psi(Z_i, \hat{\theta})' \hat{\pi}_i \right) \quad (3.3.1)$$

outperforms all other tests with respect to the empirical size. While the authors base their evidence on the exponential tilting estimator, a corresponding LM test is also available for the empirical likelihood estimator with $(\hat{\theta}, \hat{\lambda}, \hat{\pi}_i) = (\hat{\theta}_{\text{el}}, \hat{\lambda}_{\text{el}}, \hat{\pi}_i^{\text{el}})$. Further evidence on the small sample performance of this estimator is limited. One exception is the Monte Carlo experiment conducted by Imbens (1997) who compares the two-step and the iterated optimally weighed GMM estimators with the empirical likelihood estimator using a data generating process for a

linear model of covariance structures similar to the one analyzed by Abowd and Card (1989). In these experiments the empirical likelihood estimator exhibits about half of the bias created by the two GMM estimators. Similar small sample experiments conducted by Qian and Schmidt (1999), who focus on the efficiency gains of additional, parameter independent moment functions, do not reveal any systematic differences between the two-step GMM and empirical likelihood estimators regarding bias and mean squared error performance.

3.4 Bootstrap GMM_EL estimation

The empirical likelihood approach to GMM estimation implements a moment condition on the sample data which corresponds to the population orthogonality condition. Therefore the GMM bootstrap methods described in detail in Section 2.4 can be applied to the moment functions

$$\psi^*(Z_i, \theta) \equiv \hat{\pi}_i^{\text{elm}} \cdot \psi(Z_i, \theta) \quad (3.4.1)$$

which serve as an alternative to the recentered moment functions (2.4.2) suggested by Hall and Horowitz (1996). This empirical likelihood based GMM bootstrap was suggested by Brown and Newey (1995). By definition of the modified empirical likelihood (3.1.3) in connection with (3.1.9) the moment functions (3.4.1) satisfy the sample moment condition (2.4.3). Thus, the GMM bootstrap methods documented above can be used without any modification. Brown and Newey expect that using the moment functions (3.4.1) instead of the recentered moment functions suggested by Hall and Horowitz should translate into an improved large sample accuracy of the GMM bootstrap.

Brown and Newey provide some Monte Carlo evidence on the small sample performance of the moment restricted bootstrap for a dynamic linear panel data model with fixed effects. For sample sizes of 50 and 100 they show that the bootstrap-t confidence intervals achieve a better approximation to the nominal coverage probability than the confidence intervals based on first order asymptotic theory. Other applications of the Brown and Newey GMM bootstrap include Ziliak (1997) who replaces the modified empirical likelihood probabilities $\hat{\pi}_i^{\text{elm}}$ in (3.4.1) with the probabilities $\hat{\pi}_i$ given in (3.1.13) and uses the bootstrap as a Monte Carlo experiment for a particular data set. He compares different GMM and instrumental variable estimators for panel data models with weakly exogenous instruments. Bergström, Dahlberg and Johansson (1997) seem to provide the only currently existing comparison of the Hall and

Horowitz (1996) and Brown and Newey (1995) bootstrap approaches. They conduct a Monte Carlo experiment with 100 observations and focus on the small sample size properties of the bootstrap-J tests of overidentifying restrictions in the dynamic linear panel data model. The authors conclude that both methods provide an improvement over the conventional J test whereby the Brown and Newey bootstrap clearly dominates the Hall and Horowitz method.

4. Monte Carlo Investigation

4.1 Experimental Setup

This section tries to shed some light on the small sample performance of the one-step, two-step and bootstrap GMM and GMM_EL estimators. This is done by a number of Monte Carlo experiments using a data generating process for binary panel data with multiplicative unobserved time-constant effects and weakly exogenous instruments suggested by Wooldridge (1997). Following Chamberlain (1992), GMM estimators are considered which are based on a set of sequential conditional moment restrictions of the form

$$E[\rho_t(Z, \theta_0) | X_{1t}] = 0, \quad t = 1, \dots, T, \quad (4.1.1)$$

where $\rho_t(Z, \theta)$ denotes a scalar conditional moment function and $X_{1t} = (X'_1, \dots, X'_t)'$ a set of conditioning variables with $K \times 1$ elements X_s , $s = 1, \dots, t$, which expands with increasing t . Wooldridge (1997) considers a class of conditional moment functions $E[Y_t | \phi, X_{1t}] \equiv \tau_t(X, \theta_0) = \mu_t(X, \theta_0) \phi_0$ involving a nonlinear conditional mean function $\mu_t(X, \theta)$ and a multiplicative latent effect ϕ which may be correlated with the explanatory variables in $\mu_t(X, \theta)$. A special case in this class of conditional moment functions is defined by $\mu_t(X, \theta) = \exp(X'_t \theta)$ and was suggested before by Chamberlain (1992) and Pohlmeier (1994) for count data. In order to apply (4.1.1), both authors eliminate ϕ by a quasi-differencing method which leads to the transformed conditional moment function

$$\rho_t(Z, \theta) = Y_t - Y_{t+1} \frac{\tau_t(X, \theta)}{\tau_{t+1}(X, \theta)} = Y_t - Y_{t+1} \frac{\mu_t(X, \theta)}{\mu_{t+1}(X, \theta)}, \quad t = 1, \dots, T-1. \quad (4.1.2)$$

The GMM and GMM_EL estimators presented in Section 2 and 3 can be applied to the $\frac{1}{2}T(T-1)K \times 1$ vector of unconditional moment functions $\psi(Z, \theta) = A(X)p(Z, \theta)$ with $\rho(Z, \theta) = (\rho_1(Z, \theta), \dots, \rho_{T-1}(Z, \theta))'$ and

$$A(\mathbf{X}) = \begin{pmatrix} \mathbf{X}_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{X}_{1T-1} \end{pmatrix} \quad (4.1.3)$$

which satisfies $E[\psi(\mathbf{Z}, \theta_0)] = 0$ by the law of iterated expectations.²

The standard normal cumulative distribution function, $\Phi(\cdot)$, is an obvious candidate for the specification of the conditional mean function for binary data, $\mu_t(\mathbf{X}, \theta) = \Phi(\mathbf{X}'_t \theta)$, assuming that the variance of the error term of the underlying latent equation is one in all periods.³ A corresponding data generating process which is used throughout the subsequent Monte Carlo experiments is defined as follows:

$$\begin{aligned} Y_{it}^* &= \alpha_0 C_{it} + \beta_0 D_{it} + \varepsilon_{it}, & \varepsilon_{it} &\sim \text{iid } N(0,1), \quad i=1, \dots, n, \quad t=1, \dots, T, \\ A_{it}^* &= A_i + \xi_{it}, & A_i &\sim \text{iid } N(2,2), \quad \xi_{it} \sim \text{iid } N(0,1), \\ Y_{it} &= 1(Y_{it}^* > 0, A_{it}^* > 0), & & \\ C_{it} &= 0.5C_{it-1} + 0.15(A_i + \varepsilon_{it-1}) + \eta_{it}, & \eta_{it} &\sim \text{iid } U[-1,1], \\ D_{it} &= 1(D_{it}^* > 0), & D_{it}^* &\sim \text{iid } N(0,1), \end{aligned} \quad (4.1.4)$$

with $\theta = (\alpha, \beta)'$, $\theta_0 = (-1, 1)'$ and $\mathbf{X}_t = (C_t, D_t)'$ in the notation used before. The data generating process starts at $t = -10$ with $C_{it-1} = \varepsilon_{it-1} = 0$. The observability rule (4.1.4) for binary panel data was suggested by Wooldridge (1997) and implies the conditional mean function $E[Y_t | A, \mathbf{X}_t] = \Phi(\mathbf{X}'_t \theta_0) \Phi(A) \equiv \mu_t(\mathbf{X}, \theta_0) \phi_0$ which is of the multiplicative form that initiated the quasi-differencing approach employed in (4.1.2). Four experiments are distinguished by the magnitude of n and T as shown in Table 1.

Table 1.
Description of the Monte Carlo Experiments

	MC1	MC2	MC3	MC4
number of individuals (n)	100	200	100	200
number of periods (T)	3	3	4	4
number of orthogonality conditions (r)	6	6	12	12

² This choice of instruments is suboptimal. GMM estimation with optimal instruments for given sequential conditional moment functions is discussed in Chamberlain (1992) and Hahn (1997). In this section the conditional moment approach is just seen as a mean to generate unconditional moment functions for which all estimators under consideration are efficient.

³ This assumption can be weakened as shown in Inkermann (1999) who considers conditional moment estimators for the panel probit model with heteroskedasticity over time.

The estimators under consideration are those described in Sections 2.2 – 2.4 and 3.2 – 3.4. The number of Monte Carlo replications is 1,000. The dummy and continuous regressors are regenerated in each replication. The bootstrap estimators are based on 400 bootstrap samples in each replication of the Monte Carlo experiments. All calculations were performed with Gauss using the optimization package with user supplied first and second (except for the one-step estimators) analytical derivatives of the respective criterion functions.

4.2 Results

Tables A1 – A4 in the appendix contain summary statistics of the four different Monte Carlo experiments. For the two bootstrap estimators the summary statistics refer to the bias corrected parameter estimates using the correction described below (2.4.8). The COVER rows contain the empirical coverage of the 95% confidence interval around the true parameter value using the asymptotic critical values for the one-step and two-step estimators and the bootstrap-t critical values for the bootstrap estimators given in the T-CRIT row. Similar, LEVEL denotes the empirical rejection probability for the J test (or LM test for the empirical likelihood estimator) of overidentifying restrictions using the asymptotic 95% critical value for the one-step and two-step estimators and the bootstrap-J critical values for the bootstrap estimators given in the J-CRIT row. The content of the remaining rows is obvious.

All estimators exhibit a considerable amount of bias in the MC1 and MC2 experiments which imply 6 orthogonality conditions and 4 overidentifying restrictions. Increasing the number of observations from 100 in MC1 to 200 in MC2 reduces the bias of the dummy regressor coefficient α but does not improve upon the bias of the coefficient β of the continuous regressor. Doubling the number of orthogonality conditions from 6 in MC1 and MC2 to 12 in MC3 and MC4 increases the small sample bias of the one-step and two-step estimators. The bootstrap bias corrections for the two-step GMM estimates work in the wrong direction in all experiments and amplify the bias. This holds for both the Hall/Horowitz GMM bootstrap and the Brown/Newey GMM_EL bootstrap whereby the latter always performs worse. The harmful impact of the bias correction is much less severe in MC2 and MC4 which suggests that the bias correction may become effective in larger sample sizes.

Efron and Tibshirani (1993, p. 138) point out that bias correction can be dangerous in practice because of the high variability of the estimated correction term. This is obviously the case in the experiments considered here as can be seen from the standard errors of the bias

corrected estimates which always exceed the standard error of the underlying two-step GMM estimates. The continuous updating estimator exhibits the largest variation with standard errors around two times of the magnitude of the conventional two-step GMM estimates. Similar findings were reported before by Hansen, Heaton and Yaron (1996) and Imbens, Spady and Johnson (1998) who attribute this problem to flat sections of the objective function. In accordance to previous results obtained by the first group of authors, the continuous updating estimator leads to the smallest median bias of all estimators. The empirical likelihood estimator produces standard errors in the magnitude of the two-step GMM estimator in the experiments involving the larger sample size but performs worse on the smaller samples. The two-step GMM_EL estimator creates smaller standard errors than the two-step GMM estimator for $T = 3$ but larger standard errors for $T = 4$. This pattern is reflected in terms of RMSE performance but the differences between the two-step estimators are always small.

The empirical coverage rates of the symmetric confidence intervals with nominal coverage probability 0.95 are much too small for all estimators which rely on the percentiles of the asymptotic distribution of the t statistic for the construction of the confidence interval. The empirical likelihood estimator and the two two-step estimators lead to coverage rates around 0.85 while the continuous updating estimator performs worse and only reaches about 0.70 in MC3. Using the bootstrap- t method for the construction of the confidence intervals improves upon these findings and produces empirical coverage rates up to 0.90 whereby the Brown/Newey bootstrap method has a minor advantage over the Hall/Horowitz bootstrap. An explanation for the remaining coverage error could be an underestimation of the asymptotic standard errors as reported by Inkmann (1999) for the two-step GMM estimator of the random effects panel probit model. The coverage rates of the confidence interval around the true coefficient of the continuous regressor are always less distorted than the corresponding rates for the dummy regressor coefficient. The underlying average bootstrap- t critical value for the coefficient of the dummy variable is in the magnitude of 2.9 while it is around 3.4 for the coefficient of the continuous regressor.

While the bootstrap- t method improves upon the conventional t statistic, the bootstrap- J method turns out to be inferior to the conventional J test of overidentifying restrictions using the asymptotic distribution in all experiments whereby the Brown/Newey bootstrap performs slightly worse than the Hall/Horowitz bootstrap. The conventional J test for the two-step GMM estimator underrejects the null hypothesis in all experiments and the bootstrap- J meth-

ods do not yield an increase in the empirical size. The continuous updating estimator amplifies this underrejection by definition of its criterion function. The best size performance is obtained by the J test using the two-step GMM_EL estimator for $T = 3$. However, for $T = 4$ this J test overrejects. The LM test of overidentifying restrictions employed in combination with the empirical likelihood estimator performs best in MC4 where it reaches a very accurate empirical size of 0.051 but underrejects in MC1/MC2 and overrejects in MC3.

Summarizing these results, the two two-step estimators and the one-step empirical likelihood estimator show a similar overall performance. The continuous updating estimator can not be recommended because of the fat tails of its Monte Carlo distribution. Bootstrapping is useful to obtain more reliable empirical coverage probabilities but does not completely eliminate the coverage distortion of the conventional GMM approach. The bootstrap bias correction and the bootstrap-J method do not reveal the asymptotic refinements of these methods over the conventional approaches in small samples. As always, these results have to be seen conditional on the experimental setup employed in this Monte Carlo investigation.

5. Conclusion

This paper compares GMM estimators which rely on the empirical likelihood approach to the semiparametric efficient estimation of the unknown distribution of the data to conventional GMM estimators which are based on the empirical distribution function as a nonparametric estimate. One-step, two-step and bootstrap empirical likelihood and conventional approaches to efficient GMM estimation are distinguished. The estimators are subject to a Monte Carlo investigation using a specification which exploits sequential conditional moment restrictions for binary panel data with multiplicative latent effects. The Monte Carlo experiments suggest that the empirical likelihood based two-step GMM estimator may improve upon the reliability of the J test of overidentifying restrictions whereas the bootstrap-J method does not lead to a small sample size improvement. The bootstrap-t method is recommended for obtaining more reliable coverage rates of confidence intervals which are much too small if they are computed using the percentiles of the asymptotic distribution of the t statistic. The one-step continuous updating GMM estimator exhibits fat tails which prevents an useful application while the one-step empirical likelihood estimator performs similar to the conventional two-step GMM estimator.

Appendix: Tables

Table A1.
Results from 1,000 Monte Carlo Replications of the MC1 Experiment ($T = 3, n = 100$)

Estimators:		Two-step		One-step		Bootstrap	
Probabilities:		1/n	$\hat{\pi}^{\text{elm}}$	1/n	$\hat{\pi}^{\text{el}}$	1/n	$\hat{\pi}^{\text{elm}}$
α	MEAN	-0.9360	-0.9355	-1.1437	-0.9480	-0.8606	-0.8409
	BIAS	0.0640	0.0645	-0.1437	0.0520	0.1394	0.1591
	MEDIAN	-0.8570	-0.8660	-0.9389	-0.8883	-0.7275	-0.7266
	SE	0.3959	0.3768	1.0069	0.4272	0.5773	0.5609
	RMSE	0.4011	0.3823	1.0171	0.4304	0.5939	0.5830
	COVER	0.8610	0.8660	0.7880	0.8670	0.8700	0.8720
	T-CRIT	1.9600	1.9600	1.9600	1.9600	2.8436	2.8365
β	MEAN	1.0128	1.0088	1.1926	1.0245	0.7078	0.6983
	BIAS	0.0128	0.0088	0.1926	0.0245	-0.2922	-0.3017
	MEDIAN	0.8177	0.8144	0.8845	0.8307	0.5220	0.5207
	SE	0.8323	0.7972	1.3105	0.8643	1.2314	1.2023
	RMSE	0.8324	0.7973	1.3246	0.8646	1.2655	1.2396
	COVER	0.8650	0.8620	0.7690	0.8460	0.9000	0.9020
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.6851	3.7732
LEVEL	0.0250	0.0450	0.0090	0.0320	0.0260	0.0030	
J-CRIT	9.4877	9.4877	9.4877	9.4877	9.7374	10.9426	

Note: The probabilities given in the second row of the table refer to the weight which is placed on a single observation using either the nonparametric (GMM) or semiparametric (GMM_EL) distribution function estimators. The summary statistics given in the Bootstrap columns refer to the bias corrected parameter estimates. COVER denotes the empirical coverage rate of a symmetric confidence interval with nominal coverage probability 0.95. LEVEL denotes the empirical rejection rate of the test of overidentifying restrictions with nominal size 0.05. T-CRIT and J-CRIT refer to the corresponding percentiles of the asymptotic and bootstrap distributions of the t and J test statistics.

Table A2.
Results from 1,000 Monte Carlo Replications of the MC2 Experiment (T = 3, n = 200)

Estimators:		Two-step		One-step		Bootstrap	
Probabilities:		1/n	$\hat{\tau}^{\text{elm}}$	1/n	$\hat{\tau}^{\text{el}}$	1/n	$\hat{\tau}^{\text{elm}}$
α	MEAN	-0.9557	-0.9564	-1.1253	-0.9682	-0.9292	-0.9123
	BIAS	0.0443	0.0454	-0.1253	0.0318	0.0708	0.0877
	MEDIAN	-0.8929	-0.8939	-0.9572	-0.9119	-0.8218	-0.8079
	SE	0.3203	0.3182	0.6388	0.3202	0.4838	0.4731
	RMSE	0.3233	0.3214	0.6509	0.3217	0.4889	0.4811
	COVER	0.8530	0.8510	0.7850	0.8580	0.8740	0.8710
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.1340	3.1076
β	MEAN	0.9367	0.9374	1.1379	0.9353	0.8082	0.7895
	BIAS	-0.0633	-0.0626	0.1379	-0.0647	-0.1918	-0.2105
	MEDIAN	0.8304	0.8372	0.9183	0.8310	0.7120	0.6862
	SE	0.6544	0.6482	1.2487	0.6605	0.9635	0.9403
	RMSE	0.6575	0.6512	1.2563	0.6636	0.9824	0.9636
	COVER	0.8440	0.8540	0.7670	0.8420	0.8980	0.9080
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.6243	3.7815
LEVEL	0.0380	0.0490	0.0200	0.0220	0.0380	0.0210	
J-CRIT	9.4877	9.4877	9.4877	9.4877	9.7271	10.1195	

Note: cf. Table A1.

Table A3.
Results from 1,000 Monte Carlo Replications of the MC3 Experiment (T = 4, n = 100)

Estimators:		Two-step		One-step		Bootstrap	
Probabilities:		1/n	$\hat{\pi}^{\text{elm}}$	1/n	$\hat{\pi}^{\text{el}}$	$\hat{\pi}^{\text{el}}$	$\hat{\pi}^{\text{elm}}$
α	MEAN	-0.8814	-0.8879	-1.1535	-0.9051	-0.8464	-0.8254
	BIAS	0.1186	0.1121	-0.1535	0.0949	0.1536	0.1746
	MEDIAN	-0.8435	-0.8453	-0.9219	-0.8771	-0.7804	-0.7765
	SE	0.2712	0.2850	0.8422	0.3705	0.3629	0.3405
	RMSE	0.2960	0.3062	0.8561	0.3824	0.3941	0.3826
	COVER	0.7890	0.7890	0.7140	0.8290	0.8300	0.8170
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.0669	2.7466
β	MEAN	0.8986	0.8856	1.1440	0.9379	0.7391	0.7367
	BIAS	-0.1014	-0.1144	0.1440	-0.0621	-0.2609	-0.2633
	MEDIAN	0.8536	0.8187	-0.8685	0.8777	0.6867	0.6956
	SE	0.4843	0.4919	1.2675	0.5926	0.6614	0.6310
	RMSE	0.4948	0.5051	1.2756	0.5958	0.7109	0.6837
	COVER	0.8610	0.8360	0.6900	0.8480	0.9030	0.8940
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.3817	3.1501
	LEVEL	0.0250	0.1450	0.0110	0.1410	0.0200	0.0040
	J-CRIT	18.3070	18.3070	18.3070	18.3070	19.6408	22.5342

Note: cf. Table A1.

Table A4.
Results from 1,000 Monte Carlo Replications of the MC4 Experiment (T = 4, n = 200)

Estimators:		Two-step		One-step		Bootstrap	
Probabilities:		1/n	$\hat{\pi}^{\text{elm}}$	1/n	$\hat{\pi}^{\text{el}}$	$\hat{\pi}^{\text{el}}$	$\hat{\pi}^{\text{elm}}$
α	MEAN	-0.9049	-0.9057	-1.1031	-0.9229	-0.9033	-0.8834
	BIAS	0.0951	0.0943	-0.1031	0.0771	0.0967	0.1166
	MEDIAN	-0.8731	-0.8723	-0.9417	-0.8838	-0.8471	-0.8316
	SE	0.2176	0.2217	0.6279	0.2237	0.2922	0.2724
	RMSE	0.2374	0.2409	0.6363	0.2367	0.3078	0.2963
	COVER	0.8230	0.8180	0.7520	0.8340	0.8500	0.8400
	T-CRIT	1.9600	1.9600	1.9600	1.9600	2.9802	2.7299
β	MEAN	0.8794	0.8809	1.0631	0.8778	0.8259	0.8100
	BIAS	-0.1206	-0.1191	0.0631	-0.1222	-0.1741	-0.1900
	MEDIAN	0.8523	0.8492	0.9207	0.8497	0.8001	0.7758
	SE	0.3613	0.4009	0.8070	0.3753	0.4819	0.4546
	RMSE	0.3809	0.4182	0.8095	0.3947	0.5124	0.4927
	COVER	0.8640	0.8510	0.7390	0.8500	0.8980	0.9020
	T-CRIT	1.9600	1.9600	1.9600	1.9600	3.1473	3.0197
	LEVEL	0.0430	0.0870	0.0240	0.0510	0.0380	0.0110
	J-CRIT	18.3070	18.3070	18.3070	18.3070	18.9673	20.5410

Note: cf. Table A1.

References

- Abowd, J. M., and D. Card (1989): 'On the Covariance Structure of Earnings and Hours Changes', *Econometrica*, 57, 411-445.
- Altonji, J. G., and L. M. Segal (1996): 'Small-Sample Bias in GMM Estimation of Covariance Structures', *Journal of Business & Economic Statistics*, 14 (3), 353-366.
- Andrews, D. W. K., and M. Buchinsky (1997): 'A Three-Step Method for Choosing the Number of Bootstrap Repetitions', forthcoming in *Econometrica*.
- Arellano, M., and S. Bond (1991): 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, 58, 277-297.
- Back, K., and D. P. Brown (1993): 'Implied Probabilities in GMM Estimators', *Econometrica*, 61 (4), 971-975.
- Bekker, P. A. (1994): 'Alternative Approximations to the Distributions of Instrumental Variable Estimators', *Econometrica*, 62 (3), 657-681.
- Bergström, P., M. Dahlberg, and E. Johansson (1997): 'GMM Bootstrapping and Testing in Dynamic Panels', Uppsala University, Department of Economics, working paper 97/10.
- Blundell, R. W., and S. Bond (1998): 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models', *Journal of Econometrics*, 87, 115-143.
- Brown, B. W., and W. K. Newey (1995): 'Bootstrapping for GMM', MIT manuscript.
- Brown, B. W., and W. K. Newey (1998): 'Efficient Semiparametric Estimation of Expectations', *Econometrica*, 66 (2), 453-464.
- Chamberlain, G. S. (1987): 'Asymptotic Efficiency in Estimation with Conditional Moment Restrictions', *Journal of Econometrics*, 34, 305-334.
- Chamberlain, G. (1992): 'Comment: Sequential Moment Restrictions in Panel Data', *Journal of Business & Economic Statistics*, 10 (1), 20-26.
- Clark, T. E. (1996): 'Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure', *Journal of Business & Economic Statistics*, 14 (3), 367-373.
- Efron, B., and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*, New York: Chapman & Hall.
- Ferson, W., and S. Foerster (1994): 'Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models', *Journal of Financial Economics*, 36, 29-55.
- Hahn, J. (1996): 'A Note on Bootstrapping Generalized Method of Moments Estimators', *Econometric Theory*, 12, 187-197.
- Hahn, J. (1997): 'Efficient Estimation of Panel Data Models with Sequential Moment Restrictions', *Journal of Econometrics*, 79, 1-21.
- Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer.
- Hall, P., and J. L. Horowitz (1996): 'Bootstrap Critical Values for the Tests Based on Generalized-Method-Of-Moments Estimators', *Econometrica*, 64 (4), 891-916.
- Hansen, L. P. (1982): 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50 (4), 1029-1054.

- Hansen, L. P., J. Heaton, and A. Yaron (1996): 'Finite-Sample Properties of Some Alternative GMM Estimators', *Journal of Business & Economic Statistics*, 14 (3), 262-280.
- Horowitz, J. L. (1997): '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*, vol. 3, ch. 7, Cambridge: Cambridge University Press.
- Horowitz, J. L. (1998): 'Bootstrap Methods for Covariance Structures', *Journal of Human Resources*, 33 (1), 38-61.
- Imbens, G. W. (1997): 'One-Step Estimators for Over-Identified Generalized Method of Moments Models', *Review of Economic Studies*, 64, 359-383.
- Imbens, G. W., R. H. Spady, and P. Johnson (1998): 'Information-Theoretic Approaches to Inference in Moment Condition Models', *Econometrica*, 66 (2), 333-358.
- Inkmann, J. (1999): 'Misspecified Heteroskedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators', CoFE working paper no. 99/04, forthcoming in the *Journal of Econometrics*.
- Kitamura, Y., and M. Stutzer (1997): 'An Information-Theoretic Alternative to Generalized Method of Moments Estimation', *Econometrica*, 65 (4), 861-874.
- Kocherlakota, N. E. (1990): 'On Tests of Representative Consumer Asset Pricing Models', *Journal of Monetary Economics*, 26, 305-318.
- Manski, C. F. (1988): *Analog Estimation Methods in Econometrics*, New York: Chapman & Hall.
- Newey, W. K., and D. L. McFadden (1994): 'Large Sample Estimation and Hypothesis Testing', in: R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, ch. 36, Amsterdam: North-Holland.
- Owen, A. B. (1988): 'Empirical Likelihood Ratio Confidence Intervals for a Single Functional', *Biometrika*, 75, 237-249.
- Owen, A. B. (1990): 'Empirical Likelihood Ratio Confidence Regions', *The Annals of Statistics*, 18 (1), 90-120.
- Owen, A. B. (1991): 'Empirical Likelihood for Linear Models', *The Annals of Statistics*, 19 (4), 1725-1747.
- Pohlmeier, W. (1994): 'Panelökonometrische Modelle für Zähldaten: Einige neuere Schätzverfahren', *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 130, 553-574-
- Qian, H., and P. Schmidt (1999): 'Improved Instrumental Variables and Generalized Method of Moments Estimators', *Journal of Econometrics*, 91, 145-169.
- Qin, J., and J. Lawless (1994): 'Empirical Likelihood and General Estimating Equations', *The Annals of Statistics*, 22 (1), 300-325.
- Tauchen, G. (1986): 'Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data', *Journal of Business & Economic Statistics*, 4 (4), 397-416.
- Wooldridge, J. M. (1997): 'Multiplicative Panel Data Models without the Strict Exogeneity Assumption', *Econometric Theory*, 13, 667-678.
- Ziliak, J. P. (1997): 'Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators', *Journal of Business & Economic Statistics*, 15 (4), 419-431.