This paper investigates the finite-sample properties of the class of generalized empirical likelihood estimators in possibly overidentified models with weakly identified parameters. These nonparametric likelihood estimators satisfy exactly the moment conditions and automatically remove the bias that arises from a lack of centering of the moment conditions. The inference procedure suggested in the paper does not involve any explicit estimation of the variance-covariance matrix. The confidence sets for the parameters of interest are constructed by inverting the $\chi^2$ acceptance region of the criterion test.

1 Introduction

Moment condition models arise naturally from dynamic economic theory with optimizing agents. Since the seminal paper by Hansen \cite{Hansen1982}, the generalized method of moments (GMM) has become the predominant framework for estimating the structural parameters of these models. Under some general regularity conditions, the GMM estimator is consistent, asymptotically normal and efficient for the given set of moment conditions. Unfortunately, it has been found that the small-sample properties of the conventional GMM estimators (in particular, the two-step GMM) are rather poor.

In this paper, we investigate the properties of the class of generalized empirical likelihood estimators of moment condition models. Members of this class are the empirical likelihood-based GMM of Qin and Lawless \cite{Qin1994}, Imbens \cite{Imbens1997} and Imbens, Spady and Johnson \cite{Imbens1997}, and the maximum entropy-based GMM of Kitamura and Stutzer \cite{Kitamura1997} and Imbens, Spady and Johnson \cite{Imbens1997}. This class also includes the continuously-updated GMM as a special case (Imbens, Spady and Johnson \cite{Imbens1997}; Newey and Smith \cite{Newey1994}) which explains the superior small-sample performance of this estimator over the traditional two-step GMM found in Hansen, Heaton and Yaron \cite{Hansen1996}. These nonparametric likelihood estimators minimize the distance between the empirical distribution function and a distribution function that exactly satisfies the moment conditions.

One of the most attractive properties of nonparametric likelihood estimators is that they tend to remove some important sources of bias that give rise to poor finite-sample properties of the GMM estimator and GMM-based
test statistics. The first source of bias arises from the fact that the first-order conditions of the standard two-step GMM estimator (Hansen), evaluated at the true values of the parameters, are non-zero. This bias is exacerbated if the number of instruments increases (Kocherlakota). Altonji and Segal\(^1\) and Angrist, Imbens and Krueger\(^3\) proposed some \textit{ad hoc} methods for reducing the magnitude of the bias. Donald and Newey\(^9\) showed that, for the continuously-updated GMM of Hansen, Heaton and Yaron\(^{15}\), the first-order conditions are exactly satisfied at the true values of the parameters and this source of bias is automatically removed. In fact, for all members of the class of the generalized empirical estimators, the moment conditions are exactly centered at zero by construction.

Second, the estimation of the weighting matrix can be another important source of bias due to the non-zero finite sample correlation between the elements of the variance-covariance matrix of the parameters and the errors (Altonji and Segal\(^1\)). This source of bias is present for both the two-step and continuously-updated GMM estimators but disappears for the empirical likelihood (EL) estimator of Qin and Lawless\(^{29}\). Newey and Smith\(^{24}\) showed that the bias of the empirical likelihood estimator of Owen\(^{26}\) and Qin and Lawless\(^{29}\) is the same as the bias for the infeasible optimal GMM where the optimal linear combination coefficients do not have to be estimated.

Third, the small-sample properties of the GMM estimators and test statistics can be seriously affected by the choice of instruments that are only weakly correlated with the endogenous variables (Stock and Wright\(^{31}\)). In this case, the finite sample distributions of the GMM estimators and the test statistics may depart substantially from their asymptotic distributions. Stock and Wright\(^{31}\) proposed an alternative reparameterization of the moment conditions and obtained asymptotic representations with improved finite sample properties. In their framework, however, the weakly identified parameters are not consistently estimable. Fortunately, one could still conduct asymptotically valid inference by inverting criterion-based tests since their limiting \(\chi^2\)-distribution at the true values of the parameters is preserved.

In this paper, we show that the nonparametric likelihood estimators are robust in the presence of weakly identified parameters. Most importantly, the criterion-based inference procedure does not involve any explicit estimation of variance-covariance matrices. Unlike the Wald test, the confidence sets constructed by inverting the criterion test, also satisfy the requirement of infinite expected volume in the completely unidentified model (Dufour\(^{10}\)). Finally, the class of generalized empirical likelihood estimators is transformation invariant and the obtained confidence sets are transformation respecting.

The rest of the paper is structured as follows. Section 2 discusses two
approaches to estimating moment condition models that give rise to the GMM and the nonparametric likelihood estimators. The asymptotic validity of the confidence interval construction by criterion test inversion is shown in Section 3. The Monte Carlo experiment in Section 4 studies the finite-sample properties of the different estimators and their corresponding confidence intervals in a linear instrumental variable model with weakly identified parameters. In Section 5, nonparametric likelihood estimators are applied to estimating the return to education. Section 6 summarizes the conclusions.

2 General Approach to Estimating Moment Condition Models

Let $E [g(x, \theta) | F] = \int g(x, \theta) dF = 0$ be an $m \times 1$ vector of population moment conditions implied by economic theory, where $(x_1, x_2, ...)$ are independent random vectors in $\mathbb{R}^p$ with unknown continuous distribution function $F$, $\theta$ is a $k \times 1$ vector of unknown parameters from $\Theta$ and $g(.)$ is a given function $\{g(x, \theta) : \mathbb{R}^p \times \mathbb{R}^k \rightarrow \mathbb{R}^m\}$ with $m \geq k$.

Suppose that we restrict the family of possible distribution functions to the space of multinomial distributions with finite support on the observed data, denoted by $\Phi$. Also, let $F_n$ denote the empirical measure of the sample $\{x_i\}_{i=1}^n$ from $F$ that places probability mass $n^{-1}$ on each data point and $P_n$ be another probability measure that assigns multinomial weights $p_1, p_2, ..., p_n$ to each of the observations. Below, we consider two versions of the analogy principle discussed in Manski 22. The first version selects an estimator that minimizes the distance of the moment conditions from zero (GMM estimators). The second version selects an estimator that minimizes the distance between the empirical measure and a measure $P_n$ that satisfies exactly the moment conditions (nonparametric likelihood estimators).

2.1 GMM Estimators

The conventional GMM estimator minimizes the distance of the sample counterparts of these moment conditions from zero using the quadratic form

$$Q_n(\theta) = g_n(\theta)'W_n(\theta)g_n(\theta),$$

where $g_n(\theta) = E [g(x, \theta) | F_n] = \int g(x, \theta) dF_n$ and $W_n(\theta)$ is a positive definite weighting matrix. Then, $\hat{\theta} = \arg \min_{\theta \in \Theta} Q_n(\theta)$. The properties of the GMM estimator depend crucially on the choice of the weighting matrix. The optimal GMM estimator sets $W_n(\theta) = \left[ \frac{1}{n} \sum_{i=1}^n g_i(\theta)g_i(\theta)' \right]^{-1}$, where $g_i(\theta) = g(x_i, \theta)$. If a preliminary consistent (but not necessary efficient) estimator $\tilde{\theta}$ of $\theta$ is used
Finally, the continuously-updated GMM estimator proposed by Hansen, Heaton and Yaron does not require a preliminary estimate of $\theta$ and directly minimizes the criterion function

$$\hat{\theta}_{cu} = \arg \min_{\theta \in \Theta} \left[ \frac{1}{n} \sum_{i=1}^{n} g_i(\theta) \right]' \left[ \frac{1}{n} \sum_{i=1}^{n} g_i(\theta)g_i(\theta)' \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} g_i(\theta) \right].$$

(4)

The estimator is the solution to a (typically nonlinear) system of $k$ first-order conditions

$$\left( \frac{\partial g_n(\hat{\theta}_{cu})}{\partial \theta} \right)' W_n(\hat{\theta}_{cu})g_n(\hat{\theta}_{cu}) - g_n(\hat{\theta}_{cu})' W_n(\hat{\theta}_{cu}) \frac{\partial W_n(\hat{\theta}_{cu})}{\partial \theta} W_n(\hat{\theta}_{cu})g_n(\hat{\theta}_{cu}) = 0$$

or

$$\left[ \frac{1}{n} \sum_{i=1}^{n} \left( 1 + \Lambda' g_i(\hat{\theta}_{cu}) \right) \frac{\partial g_i(\hat{\theta}_{cu})}{\partial \theta} \right]' \left[ \frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{cu})g_i(\hat{\theta}_{cu})' \right]^{-1} g_n(\hat{\theta}_{cu}) = 0,$n(\hat{\theta}_{cu}) \right],$$

(5)

where $\Lambda = - \left[ \sum_{i=1}^{n} g_i(\hat{\theta}_{cu})g_i(\hat{\theta}_{cu})' \right]^{-1} g_n(\hat{\theta}_{cu})$.

Although these estimators are asymptotically equivalent, their finite sample properties may differ (see for example Hansen, Heaton and Yaron).

2.2 Nonparametric Likelihood Estimators

A second approach is to obtain a value of $\theta$ that minimizes a distance between probability measures rather than the distance of the moment conditions from zero. This data driven approach selects from the set of distributions that
satisfy exactly the moment conditions a probability measure \( P_n \) closest to the empirical measure \( F_n \) defined by the Cressie and Read\(^7\) power divergence criterion

\[
D_\rho(F_n, P_n) = \frac{2}{\rho(1+\rho)} \sum_{i=1}^{n} p_i [(np_i)^\rho - 1],
\tag{6}
\]

where \( \rho \) is a fixed scalar parameter which determines the shape of the criterion function. Cressie and Read\(^7\) proposed the family of power divergence statistics as goodness-of-fit tests. Here, we use the Cressie-Read divergence criterion for estimation purposes. The estimator is defined as the solution to

\[
\min_{P_n \in \Phi, \theta \in \Theta} D_\rho(F_n, P_n) \tag{7}
\]

subject to \( E[g(x, \theta) | P_n] = \int g(x, \theta) dP_n = 0. \tag{8} \)

This form of the analogy principle maps the empirical distribution function onto the space of feasible distribution functions and chooses the probability measure that is most likely to have generated the observed data, subject to the moment conditions (Manski\(^{22}\)). The solution to the above constrained optimization problem is a straightforward application of the Lagrange multiplier principle.

This framework embeds several interesting special cases (see Kitamura and Stutzer\(^{19}\); and Imbens, Spady and Johnson\(^{17}\)). The first two cases can also be interpreted as discrete versions of the forward and backward Kullback-Leibler discrepancy between the empirical measure and \( P_n \). If we let \( \rho \) approach 0, the estimator is the solution to the problem

\[
\min_{p, \theta} -2 \frac{1}{n} \sum_{i=1}^{n} \ln p_i \tag{9}
\]

subject to \( \sum_{i=1}^{n} p_i g(x_i, \theta) = 0 \) and \( \sum_{i=1}^{n} p_i = 1. \tag{10} \)

This is the empirical likelihood estimator of Owen\(^{26,27,28}\) and Qin and Lawless\(^{29}\) obtained as the root of the system of equations

\[
\begin{align*}
\left( \frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{EL}) / \left( 1 + \hat{\lambda} g_i(\hat{\theta}_{EL}) \right) \right) \\
\left( \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda} \left( \frac{\partial g_i(\hat{\theta}_{EL})}{\partial \theta} \right) / \left( 1 + \hat{\lambda} g_i(\hat{\theta}_{EL}) \right) \right) = 0_{m+k},
\end{align*}
\]
where $\lambda$ is a vector of Lagrange multipliers on the moment conditions.

The $(m+k) \times 1$ parameter vector $(\hat{\theta}_{EL}, \hat{\lambda})'$ can then be used to compute the vector of probability weights $p_i = \left[n \left(1 + \hat{\lambda}'g_i(\hat{\theta}_{EL})\right)\right]^{-1}$ for $i = 1, 2, \ldots, n$. This gives an efficient estimate of the distribution function $F$ given the set of moment conditions. One drawback of this method is that the form of the loss function resists heavy downweighting of specific data points which may be a problem in the presence of outliers.

If $\rho \to -1$, we obtain the maximum entropy, exponential tilting or KLIC (Kullback-Leibler information criterion) estimator of Efron\textsuperscript{11} and DiCiccio and Romano\textsuperscript{8} and discussed in Kitamura and Stutzer\textsuperscript{19}, Imbens\textsuperscript{10} and Imbens, Spady and Johnson\textsuperscript{17}). It is computed as the minimizer of the KLIC function

$$
\min_{\theta, \lambda} 2 \sum_{i=1}^{n} p_i \ln np_i
$$

subject to (10). The maximum entropy estimator solves the system of first-order conditions

$$
\begin{align*}
\left(\frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{EL}) \exp \left(\hat{\lambda}'g_i(\hat{\theta}_{EL})\right)\right) &= 0_{m+k}, \\
\frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}' \left(\frac{\partial g_i(\hat{\theta}_{EL})}{\partial \theta'}\right) \exp \left(\hat{\lambda}'g_i(\hat{\theta}_{EL})\right) &= 0_{m+k}.
\end{align*}
$$

Finally, if $\rho \to -2$, we obtain the Euclidean likelihood estimator of Owen\textsuperscript{28} given by the argument that minimizes

$$
\min_{\theta, \lambda} \frac{1}{n} \sum_{i=1}^{n} (n^2 p_i^2 - 1)
$$

subject to (10). The solution is obtained from

$$
\begin{align*}
\left(\frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{EU}) \left[1 + \hat{\lambda}'\bar{g}_i(\hat{\theta}_{EU})\right]\right) &= 0_{m+k}, \\
\frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}' \left(\frac{\partial g_i(\hat{\theta}_{EU})}{\partial \theta'}\right) \left[1 + \hat{\lambda}'\bar{g}_i(\hat{\theta}_{EU})\right] &= 0_{m+k},
\end{align*}
$$

where $\bar{g}_i(\hat{\theta}_{EU}) = g_i(\hat{\theta}_{EU}) - \frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{EU})$. From the first $m$ equations, $\hat{\lambda} = -\left[\frac{1}{n} \sum_{i=1}^{n} \bar{g}_i(\hat{\theta}_{EU})\bar{g}_i(\hat{\theta}_{EU})\right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_{EU})\right]$. Then, by substituting for
\( \hat{\lambda} \) and \( \hat{p}_i = n^{-1} \left[ 1 + \hat{\lambda} \bar{g}_i(\bar{\theta}_{EU}) \right] \) in the last \( k \) equations, we get

\[ \left[ \frac{1}{n} \sum_{i=1}^{n} \hat{p}_i \left( \frac{\partial g_i(\bar{\theta}_{EU})}{\partial \theta} \right) \right]' \left[ \frac{1}{n} \sum_{i=1}^{n} \bar{g}_i(\bar{\theta}_{EU}) \bar{g}_i(\bar{\theta}_{EU})' \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} g_i(\bar{\theta}_{EU}) \right] = 0_k. \]

It is interesting to see that this system of first-order conditions is almost identical to (5) for the continuously-updated GMM estimator and very similar to (3) for the iterated GMM estimator. Hence, the Euclidean likelihood estimator can be interpreted as a continuously-updated GMM estimator and an optimally weighted iterated GMM estimator.

Note also that the objective functions (9) and (11) implicitly impose the constraint \( p_i \geq 0 \) which validates the interpretation of \( p_i \) as probability weights. For the Euclidean estimator, we either have to inspect the positivity of the probability weights each time or introduce a nonnegativity constraint explicitly into the minimization problem. Owen \cite{Owen28} argues that in small samples, the negativity of the estimated weights may be advantageous for confidence interval construction.

### 3 Inference in Moment Condition Models

Consider the estimators discussed in Section 2. Let \( \theta_0 \) denote the true value of the parameter vector and suppose that the following regularity conditions are satisfied.

**Assumption A1.** Assume that \( W_n \xrightarrow{p} W \), where \( W \) is a nonstochastic symmetric positive definite matrix; \( g(x_i, \theta) \) is continuous in \( \theta \); \( E[\sup_{\theta \in \Theta} |g(x_i, \theta)|] < \infty \); \( \sup_{\theta} E[g(x_i, \theta)g(x_i, \theta)'] < \infty \) for all \( \theta \) and \( \Theta \) is a compact subset of \( \mathbb{R}^k \).

**Assumption A2.** There is a unique \( \theta_0 \) such that \( E[g(x_i, \theta_0)] = 0 \) and \( E[g(x_i, \theta)] \neq 0 \) for all \( \theta \neq \theta_0 \in \Theta \).

**Assumption A3.** Assume that \( M = E\left( \frac{\partial g(x_i, \theta_0)}{\partial \theta'} \right) \) is of full rank \( k \); \( g(x_i, \theta) \) is continuous in \( \theta \) and \( E\left[ \sup_{\theta \in N(\theta_0)} \left| \frac{\partial g(x_i, \theta)}{\partial \theta'} \right| \right] < \infty \) for some neighborhood of \( \theta_0, N(\theta_0) \).

**Theorem 1.** Under Assumptions A1-A2,

\[ \hat{\theta}_n \overset{p}{\to} \theta_0 \text{ as } n \to \infty \]
\[ \sqrt{n}(\hat{\theta}_\rho - \theta_0) \xrightarrow{d} N(0, \Omega), \]

where \( \Omega = (M'V^{-1}M)^{-1} \) and \( V = E(g(x, \theta_0)g(x, \theta_0)') \).

**Proof.** See Imbens\textsuperscript{16}, Qin and Lawless\textsuperscript{29}, and Newey and Smith\textsuperscript{24}.

**Theorem 2.** Let \( \theta = (\alpha, \beta)' \), where \( \alpha \in \Theta_1 \) and \( \beta \in \Theta_2 \) are \( p \times 1 \) and \( (k-p) \times 1 \) vectors, respectively. Then, under Assumptions A1-A3,

(i) test for overidentifying restrictions

\[ nD_\rho(\hat{\theta}_\rho) \xrightarrow{d} \chi^2_{(m-k)} \text{ as } n \to \infty \]

(ii) test of \( H_0 : \theta = \theta_0 \)

\[ n \left[ D_\rho(\theta_0) - D_\rho(\hat{\theta}_\rho) \right] \xrightarrow{d} \chi^2_{(k)} \text{ as } n \to \infty \]

(iii) test for a subset \( \alpha \) with \( H_0 : \alpha = \alpha_0 \)

\[ n \left[ D_\rho(\alpha_0, \hat{\beta}_\rho) - D_\rho(\hat{\alpha}_\rho, \hat{\beta}_\rho) \right] \xrightarrow{d} \chi^2_{(p)} \text{ as } n \to \infty \]

where \( D_\rho(\theta_0) \) for \( \rho = -2, -1 \) and 0 is the criterion function defined in (12), (11) and (9), \( \hat{\theta}_\rho = (\hat{\alpha}_\rho, \hat{\beta}_\rho) \) is the unrestricted nonparametric likelihood estimator that minimizes \( D_\rho(\theta) \) and \( \hat{\beta}_\rho \) is the minimizer of \( D_\rho(\alpha, \beta) \) subject to \( \alpha = \alpha_0 \).

**Proof.** See Imbens\textsuperscript{16}, Qin and Lawless\textsuperscript{29}, and Newey and Smith\textsuperscript{24}.

The results in Theorems 1 and 2 show that we can conduct asymptotically valid inference such as testing for overidentifying restrictions and constructing confidence intervals by inverting the \( \chi^2 \) acceptance region of the criterion test. The \( 100\eta \% \) confidence set for the parameter of interest \( \theta \) is then given by the set of values of \( \theta \) satisfying

\[ C_\eta(x) = \{ \theta \in \Theta : D_\rho(\theta) \leq q_\eta \}, \]

where \( q_\eta \) is the \( 100\eta \text{th} \) quantile of the distribution of \( D_\rho(\theta) \). Equivalently, \( C_\eta(x) = \{ \theta \in \Theta : x \in A(\theta) \} \), where \( A(\theta) \) is the acceptance region of the test \( D_\rho(\theta) \). The endpoints of the confidence set are the infimum and the supremum over \( C_\eta(x) \), respectively. In particular, the two-sided, equal-tailed confidence interval with nominal coverage \( \eta \) is given by \( C_\eta(x) = [\theta_L, \theta_U] \), where
the confidence limits are defined to satisfy \( \theta_L = \inf \{ \theta \in \Theta : \Pr(D_{\rho}(\theta) \leq q_\eta | H_0) \geq \eta \} \) and \( \theta_U = \sup \{ \theta \in \Theta : \Pr(D_{\rho}(\theta) \leq q_\eta | H_0) \geq \eta \} \).

For the weak instrument case, Stock and Wright 31 parameterized the moment condition as a function of the sample size and developed an alternative limiting theory which yields a better approximation to the finite-sample distributions of the estimator and corresponding test statistics. In particular, Stock and Wright 31 replace Assumption 2 with the assumption that

\[ E[g(x_i, \theta)] = n^{-1/2}m(\theta) \text{ uniformly in } \theta \in \Theta, \]

where \( m(\theta) \) is continuous in \( \theta \) and bounded on \( \Theta \) with \( m(\theta_0) = 0 \). Under this assumption, the GMM and nonparametric likelihood estimators are no longer consistent \((\hat{\theta}_\rho - \theta_0 = O_p(1))\) although the \( \chi^2 \) asymptotic approximation for the distributions of the test for overidentifying restrictions and a test of a subvector of weakly identifiable parameters is still valid.

4 Monte Carlo Study

The poor small sample performance of the two-step GMM estimator in linear homoskedastic instrumental variables models with weakly identified parameters has been well documented in Nelson and Startz 23, Maddala and Jeong 21, Bound, Jaeger and Baker 5 and Staiger and Stock 30 among others. In this section, we assess the robustness of the nonparametric likelihood estimators in the presence of weak instruments.

The structure of the Monte Carlo experiment is similar to the one considered by Angrist, Imbens and Krueger 3. It is designed to study the finite sample bias of the different estimators and the size properties of hypothesis tests and the test for overidentifying restrictions with a large number of irrelevant instruments. The data are generated from the model

\[ y_i = \theta_0 + \theta_1 x_i + e_i, \]
\[ x_i = \gamma_0 + \sum_{j=1}^{m-1} \gamma_j z_{ij} + u_i, \]

where \( z \sim N(0, I), \left( \begin{array}{c} e_i \\ u_i \end{array} \right) = \text{chol}(\Sigma) \xi_i, \xi_i \sim iid(0, I), \Sigma = \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}, \]

\text{chol}(\Sigma) \text{ denotes Cholesky decomposition of } \Sigma, \theta_0 = 0, \theta_1 = 1, \gamma_0 = 0, \gamma_1 = 0.15 \text{ and } \gamma_l = 0 \text{ for } l = 2, ..., m - 1.

The optimal two-step GMM estimator in this setting is asymptotically equivalent to the two-stage least squares (2SLS) estimator with \( W_n = (Z'Z)^{-1} \)

\[ \hat{\theta}_{2\text{step}} = (X'P_z X)^{-1} (X'P_z y), \]

where \( X = (1, x), Z = (1, z) \) and \( P_z = Z(Z'Z)^{-1}Z. \)
We also consider the limited information maximum likelihood (LIML) estimator which is given by
\[
\hat{\theta}_{LIML} = (X'(I - kM_z)X)^{-1} (X'(I - kM_z)y),
\]
where \(M_z = I - P_z\), \(k\) is the smallest characteristic root of \((Y'Y)(Y'M_zY)^{-1}\) and \(Y = (y \ X)\).

The confidence intervals for the OLS, 2SLS and LIML estimators are constructed by inverting the Wald test using the \(\chi^2\) critical values. The confidence intervals for the empirical likelihood (EL) and the Euclidean likelihood (Euclid) estimators are obtained by inverting the \(\chi^2\) acceptance regions of the corresponding criterion-based tests discussed above. The results for the exponential tilting (KLIC) estimator are very similar to the results for the EL estimator and are not reported. The number of Monte Carlo replications is 10,000.

Table 1. Monte Carlo results for model (13) with \(T=500\) and Gaussian errors.

<table>
<thead>
<tr>
<th>estimator</th>
<th>quantiles around ((\hat{\theta}_1 - \theta_1))</th>
<th>test for (\theta_1)</th>
<th>coverage rate of CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>(m - k = 1, \gamma_1 = .15, \gamma_2 = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.6991</td>
<td>0.7154</td>
<td>0.7336</td>
</tr>
<tr>
<td>2SLS</td>
<td>-0.1826</td>
<td>-0.0693</td>
<td>0.0348</td>
</tr>
<tr>
<td>LIML</td>
<td>-0.2297</td>
<td>-0.1028</td>
<td>0.0165</td>
</tr>
<tr>
<td>EL</td>
<td>-0.2362</td>
<td>-0.1120</td>
<td>0.0005</td>
</tr>
<tr>
<td>Euclid</td>
<td>-0.2354</td>
<td>-0.1124</td>
<td>0.0011</td>
</tr>
<tr>
<td>(m - k = 5, \gamma_1 = .15, \gamma_2 = ... = \gamma_5 = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.6987</td>
<td>0.7157</td>
<td>0.7339</td>
</tr>
<tr>
<td>2SLS</td>
<td>-0.1098</td>
<td>-0.0148</td>
<td>0.0830</td>
</tr>
<tr>
<td>LIML</td>
<td>-0.2260</td>
<td>-0.0975</td>
<td>0.0193</td>
</tr>
<tr>
<td>EL</td>
<td>-0.2427</td>
<td>-0.1175</td>
<td>-0.0011</td>
</tr>
<tr>
<td>Euclid</td>
<td>-0.2402</td>
<td>-0.1179</td>
<td>-0.0022</td>
</tr>
<tr>
<td>(m - k = 10, \gamma_1 = .15, \gamma_2 = ... = \gamma_{11} = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.6985</td>
<td>0.7149</td>
<td>0.7338</td>
</tr>
<tr>
<td>2SLS</td>
<td>-0.0145</td>
<td>0.0663</td>
<td>0.1484</td>
</tr>
<tr>
<td>LIML</td>
<td>-0.2291</td>
<td>-0.1020</td>
<td>0.0186</td>
</tr>
<tr>
<td>EL</td>
<td>-0.2394</td>
<td>-0.1154</td>
<td>-0.0002</td>
</tr>
<tr>
<td>Euclid</td>
<td>-0.2408</td>
<td>-0.1161</td>
<td>-0.0010</td>
</tr>
</tbody>
</table>
First, we assess the effect of increasing the number of redundant moment restrictions on the magnitude of the bias of the estimators and the size properties of the corresponding criterion tests. Tables 1 reports the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles of the distribution of \((\hat{\theta}_1 - \theta_1)\) as well as the empirical size of the test for overidentifying restrictions (OIR) with nominal level 0.1 and the coverage properties of the 90\% confidence intervals for \(\theta_1\) in a model with Gaussian errors.

Newey and Smith \(^{24}\) showed that in model (13) with symmetric errors 
\[
\text{bias}(\hat{\theta}_{LIML}) = \text{bias}(\hat{\theta}_{EL}) = \text{bias}(\hat{\theta}_{EU}) = -\delta/n, \quad \text{where} \quad \delta = \Omega \sigma_{ex} / \sigma_{e}^2.
\]
Thus, the LIML, EL and Euclidean estimators are higher-order asymptotically equivalent and their bias does not depend on the number of instruments. By contrast, 
\[
\text{bias}(\hat{\theta}_{2sle}) = (m - 2)\delta/n \quad \text{which increases linearly with the number of instruments.}
\]

To investigate the finite-sample sensitivity of the bias of the estimators with respect to the number of irrelevant instruments, we consider three cases of overidentification: \(m - 1 = 2\) (1 overidentifying restriction), \(m - 1 = 6\) (5 overidentifying restrictions) and \(m - 1 = 11\) (10 overidentifying restrictions). Tables 1 contains the results for a sample size of 500 which is commonly encountered in economic applications.

As expected, the OLS estimator is severely upward biased. The 2SLS is slightly biased when \(m - k = 1\), but its bias starts to approach the bias of the OLS estimator as \(m\) increases. Also, the size properties of the test based on the 2SLS deteriorate significantly as the number of instruments gets large. The magnitude of the bias of the LIML estimator is small and insensitive to the degree of overidentification of the model which is consistent with the theoretical results.

The bias of the nonparametric likelihood methods is negligible and it is practically unchanged as the number of the overidentifying restrictions increases. The confidence intervals based on the nonparametric likelihood estimators slightly undercover with the coverage rate of the Euclidean likelihood being closest to the nominal level. Similar results were obtained for sample size \(T = 150\) but these results are not reported due to space limitations. It is interesting to note that the dominance of the Euclidean over the EL estimator in terms of coverage rates is more pronounced for the smaller sample size. This requires further theoretical investigation of the higher-order properties of these tests using Edgeworth expansions.

The higher-order asymptotic equivalence of the LIML, EL and Euclidean estimators derived by Newey and Smith \(^{24}\) is valid only for models with symmetric errors. Newey and Smith \(^{25}\) show that all members of the class of nonparametric likelihood estimators except EL have an additional bias term

11
coming from the estimation of the variance-covariance matrix $\Omega$. To investigate the sensitivity of the results to fat tailed and asymmetric distributions, we also report results from weakly identified models with $t$-distributed errors with 4 degrees of freedom and $\chi^2$-distributed errors with 1 degree of freedom. In addition, we vary the correlation of the endogenous explanatory variable with the instruments.

The simulation results in Table 2 show that the nonparametric likelihood estimators provide reliable inference in the presence of weak instruments regardless of the distribution of the errors. Similar to the results in Table 1, the Euclidean estimator dominates in terms of coverage rate with empirical levels within 3 percentage points from the nominal level. Although the bias of the empirical and Euclidean likelihood estimators could be significant for asymmetric errors (for instance, the bottom left corner of Table 2), it is still considerably smaller than the 2SLS and LIML estimators. In summary, the nonparametric likelihood estimators appear to possess good finite-sample properties in overidentified models with weak instruments.

| estimator $\xi_i \sim N(0, I)$ | $\gamma_1 = .05,\gamma_2 = .02$ & $\gamma_1 = .05,\gamma_2 = .05$ & $\gamma_1 = .05,\gamma_2 = .1$ |
|-------------------------------|---------------------------------|---------------------------------|---------------------------------|
| OLS                           | median 0.7909 test -0.0000 cov.rate 0.7844 test -0.0000 | median 0.7621 test -0.0000 cov.rate 0.7844 test -0.0000 | median 0.7844 test -0.0000 cov.rate 0.7844 test -0.0000 |
| 2SLS                          | median 0.1300 test 0.1352 cov.rate 0.8182 test 0.8405 | median 0.1266 test 0.1352 cov.rate 0.8182 test 0.8405 | median 0.1126 test 0.1352 cov.rate 0.8182 test 0.8405 |
| LIML                          | median 0.0458 test 0.0995 cov.rate 0.8552 test 0.8645 | median 0.0285 test 0.0995 cov.rate 0.8552 test 0.8645 | median 0.0285 test 0.0995 cov.rate 0.8552 test 0.8645 |
| EL                            | median 0.0149 test 0.0859 cov.rate 0.8946 test 0.8946 | median 0.0040 test 0.0957 cov.rate 0.8946 test 0.8946 | median 0.0040 test 0.0957 cov.rate 0.8946 test 0.8946 |
| Euclid                        | median 0.0140 test 0.0852 cov.rate 0.8946 test 0.8946 | median 0.0037 test 0.0950 cov.rate 0.8946 test 0.8946 | median 0.0037 test 0.0950 cov.rate 0.8946 test 0.8946 |

| estimator $\xi_i \sim t(4)$ | $\gamma_1 = .05,\gamma_2 = .02$ & $\gamma_1 = .05,\gamma_2 = .05$ & $\gamma_1 = .05,\gamma_2 = .1$ |
|-------------------------------|---------------------------------|---------------------------------|---------------------------------|
| OLS                           | median 0.7900 test -0.0000 cov.rate 0.7841 test -0.0000 | median 0.7620 test -0.0000 cov.rate 0.7841 test -0.0000 | median 0.7841 test -0.0000 cov.rate 0.7841 test -0.0000 |
| 2SLS                          | median 0.1332 test 0.1390 cov.rate 0.8190 test 0.8479 | median 0.0333 test 0.1124 cov.rate 0.8190 test 0.8479 | median 0.0333 test 0.1124 cov.rate 0.8190 test 0.8479 |
| LIML                          | median 0.0443 test 0.1006 cov.rate 0.8502 test 0.8715 | median 0.0337 test 0.1396 cov.rate 0.8502 test 0.8715 | median 0.0337 test 0.1396 cov.rate 0.8502 test 0.8715 |
| EL                            | median 0.0105 test 0.0930 cov.rate 0.8842 test 0.8836 | median 0.0031 test 0.1064 cov.rate 0.8842 test 0.8836 | median 0.0031 test 0.1064 cov.rate 0.8842 test 0.8836 |
| Euclid                        | median 0.0083 test 0.0890 cov.rate 0.8943 test 0.8943 | median 0.0026 test 0.1016 cov.rate 0.8943 test 0.8943 | median 0.0026 test 0.1016 cov.rate 0.8943 test 0.8943 |

| estimator $\xi_i \sim \chi^2(1)$ | $\gamma_1 = .05,\gamma_2 = .02$ & $\gamma_1 = .05,\gamma_2 = .05$ & $\gamma_1 = .05,\gamma_2 = .1$ |
|-------------------------------|---------------------------------|---------------------------------|---------------------------------|
| OLS                           | median 0.7957 test -0.0000 cov.rate 0.7922 test -0.0000 | median 0.7795 test -0.0000 cov.rate 0.7922 test -0.0000 | median 0.7922 test -0.0000 cov.rate 0.7922 test -0.0000 |
| 2SLS                          | median 0.2661 test 0.1527 cov.rate 0.7453 test 0.8043 | median 0.0618 test 0.1168 cov.rate 0.7453 test 0.8043 | median 0.0618 test 0.1168 cov.rate 0.7453 test 0.8043 |
| LIML                          | median 0.2136 test 0.1720 cov.rate 0.7314 test 0.7799 | median 0.0333 test 0.1463 cov.rate 0.7314 test 0.7799 | median 0.0333 test 0.1463 cov.rate 0.7314 test 0.7799 |
| EL                            | median 0.1126 test 0.0893 cov.rate 0.8545 test 0.8644 | median 0.0032 test 0.1019 cov.rate 0.8545 test 0.8644 | median 0.0032 test 0.1019 cov.rate 0.8545 test 0.8644 |
| Euclid                        | median 0.1027 test 0.0825 cov.rate 0.8702 test 0.8776 | median 0.0027 test 0.0940 cov.rate 0.8702 test 0.8776 | median 0.0027 test 0.0940 cov.rate 0.8702 test 0.8776 |
The finite-sample properties of the constructed confidence intervals for the nonparametric likelihood methods can be further improved by bootstrap methods. For the efficient bootstrap suggested by Brown and Newey\textsuperscript{6} and Hall and Presnell\textsuperscript{13}, the data can be resampled using the implied probability weights $\hat{p}_i$ from the estimation problem rather than the empirical measure ($p_i = \frac{1}{n}$ for all $i$) as in the conventional bootstrap. Also, the asymptotic validity of the conventional bootstrap requires explicit recentering of the moment conditions (Hall and Horowitz\textsuperscript{12}) whereas for the efficient bootstrap the moment conditions, evaluated at the true parameter vector, are centered at zero by construction.

5 Empirical Illustration: Return to Education

Estimating the return to education is of central interest to labour economists. It shows the predicted percentage increase in wage for an additional year of education. Since education is believed to be endogenous, Angrist and Krueger\textsuperscript{2} suggested the quarter of birth as an instrument for education. However, Bound, Jaeger and Baker\textsuperscript{5} challenged the results obtained by Angrist and Krueger\textsuperscript{2} arguing that the two-step GMM could be severely biased in the presence of a large number of weak instruments.

Here we use the Angrist-Krueger data set which consists of a random sample from 1980 census of 329,500 men who were born between 1930 and 1939. See Angrist and Krueger\textsuperscript{2} for a detailed description of the data and model specification. Following Angrist and Krueger\textsuperscript{2}, 30 instruments are constructed by interacting quarter and year of birth. Then we draw random subsamples of 500 observations from the original sample without replacement. The results in Table 3 are obtained from 5,000 repetitions and report the median of the parameter estimates and their corresponding standard errors.

<table>
<thead>
<tr>
<th>parameter estimate</th>
<th>OLS</th>
<th>2SLS</th>
<th>EL</th>
<th>Euclid</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard error</td>
<td>0.0087</td>
<td>0.0437</td>
<td>0.0404</td>
<td>0.0373</td>
</tr>
</tbody>
</table>

The estimated return to schooling for all methods is in the range of 6.5\% and 7.3\%. This is a bit surprising since the weak instruments employed in the estimation are expected to produce a large upward bias in the OLS and 2SLS estimates. This does not seem to be the case and all the estimates do not appear significantly different from one another. It is also interesting to note the smaller standard errors for the nonparametric likelihood estimators compared to the two-step GMM estimator.
6 Concluding Remarks

This paper shows the robustness of nonparametric likelihood estimators of moment condition models to the presence of weak instruments and nonnormal errors. The computational procedure does not involve any explicit bias correction or estimation of variance-covariance matrices. The confidence intervals are obtained directly from the criterion function by inverting its asymptotic acceptance region.

One interesting finding that emerges from the study is the existence of noticeable differences in the coverage properties within the class of generalized empirical likelihood estimators. Since the criterion-based test statistics are asymptotically equivalent, higher-order expansions are necessary to appraise the statistical significance of these results.

Acknowledgments

I would like to thank Gordon Fisher and two anonymous referees for helpful comments and suggestions. Financial support from the FRDP of Concordia University is gratefully acknowledged.

References


