QUASI EMPIRICAL LIKELIHOOD ESTIMATION OF MOMENT CONDITION MODELS

SHANE M. SHERLUND

Abstract. In this paper, I develop a quasi empirical likelihood estimator that has good finite sample properties when there are many moment conditions. I show that the quasi empirical likelihood estimator, which uses semiparametric efficient estimation, is an approximation to the empirical likelihood estimator, which has been shown to have good statistical properties. The quasi empirical likelihood estimator is a consistent estimator and has a normal asymptotic distribution. As with the full-blown empirical likelihood estimator, the quasi empirical likelihood estimator reduces finite-sample bias, but is much simpler to compute than the empirical likelihood estimator. Monte Carlo experiments and a quick validation exercise confirm my theoretical results.

1. Introduction

Moment condition models arise frequently in applied economics, including the instrumental variable estimation of supply or demand functions, Euler equations implied by dynamic optimization, and even dynamic panel data models. The two-stage least squares (2SLS) and generalized method of moments (GMM) estimators, not surprisingly, have received much attention in the literature because, under fairly general regularity conditions, these estimators are consistent, efficient, and asymptotically normal (Newey and McFadden, 1994). Recent work in this area, however, has called

Date: January 2004.

Prelimianry. Comments are very much appreciated.

Keywords: GMM, empirical likelihood, bias reduction, instrumental variables.

Contact information: Federal Reserve Board, Mail Stop 93, Washington, DC 20551. Phone: 202-452-3589. Fax: 202-728-5887. E-mail: shane.m.sherlund@frb.gov. Web: http://www.sit.wisc.edu/ smsherlund.

I thank Bruce Hansen, Yuichi Kitamura, Gautam Tripathi, Chris Barrett, Michael Carter, Bob Staiger, Scott Taylor, Ken West, Limor Golan, and seminar audiences at the Federal Reserve Board, Wisconsin, Brigham Young, Arkansas, and Wyoming for helpful comments and discussions. Financial assistance from a Christensen Award at the University of Wisconsin is gratefully acknowledged. This research does not necessarily reflect the views of the Board of Governors of the Federal Reserve System, its members, or its staff.

into question the validity of these statistical properties in the finite-sample setting (e.g., Bekker, 1994; Altonji and Segal, 1996; Staiger and Stock, 1997; Newey and Smith, 2004). In particular, finite-sample bias can arise due to a large number of moment conditions or because of weak identification (weak instruments, in the linear instrumental variables setting). This finite-sample bias can result in misleading confidence intervals and potentially meaningless hypothesis testing. Therefore, researchers must either implement alternatives to 2SLS and GMM or choose the number of moment conditions so as to minimize some measure of badness (such as mean squared error) to reduce finite sample bias.

Alternatives to 2SLS and GMM include the limited information maximum likelihood estimator (LIML), the continuous updating estimator (CUE; Hansen, Heaton, and Yaron, 1996), the empirical likelihood estimator (EL; Owen, 1988, 1990, 1991, 2001; Qin and Lawless (1994), and the exponential tilting estimator (ET; Kitamura and Stutzer, 1997; Imbens, 1997; Imbens, Spady, and Johnson, 1998). Newey and Smith (2004) show that CUE, EL, and ET, begin part of the class of generalized empirical likelihood estimators, are as unbiased as infeasible GMM, in which the optimal weight matrix is known a priori. Further, the EL estimator eliminates all bias asymptotically, regardless of the number of moment conditions. An additional benefit of EL is that its empirical likelihood ratio test is at least as powerful as any other overidentifying restrictions test under some size constraint (Kitamura, 2001). Unfortunately, these estimators do not come without cost. EL becomes computationally burdensome (and possibly even intractable) when the number of moment conditions is even moderately large. LIML and CUE typically suffer from fat-tailed finite-sample distributions.

In contrast to implementing an alternative estimation technique, Donald and Newey (2001) suggest choosing the number of instruments (or, synonymously, moment conditions) to minimize approximate mean squared error in order to reduce finite-sample bias. While this may indeed be an improvement over the naive use of all available instruments, this approach may not prove particularly useful in empirical applications for several reasons. First, even when the number of instruments is chosen optimally, the number of instruments may still be large relative to the sample size. Thus, finite-sample bias may still be a problem (see, for instance, Bekker, 1994). Second, once one has chosen the optimal number of instruments, the choice among instruments is arbitrary.

Researchers using the same data and following the same methodology can come to completely different conclusions – all based on the selection of particular (sets of) instruments.

I therefore propose the quasi empirical likelihood estimator (QEL), which reduces finite-sample bias in estimating moment condition models but is very simple to compute (much akin to two-step GMM), even when the number of moment conditions is large. QEL approximates EL via semiparametric efficient estimation (Brown and Newey, 1998). I show that QEL is nearly unbiased under higher-order asymptotic theory. Because of this, QEL can drastically reduce finite-sample bias, but at a fraction of the computational cost of EL. Further, because this result is robust to the number of moment conditions, QEL eliminates the need to choose among moment conditions arbitrarily. In addition, QEL has an intuitive GMM interpretation and is robust to general forms of heteroskedasticity.

QEL, however, is not a perfect estimator. First, as with all estimators, QEL assumes that the moment conditions hold. That is, the underlying moment conditions are assumed to hold in the population. Second, QEL trades bias for variance. So even though bias is lower, variance is higher. In general, however, the contribution of reduced bias in the mean squared error is larger than the contribution of increased variance, leading to mean squared error gains. Finally, QEL relies on a consistent initial estimator of the underlying population parameters. An initial estimator suffering from finite-sample bias, for instance, may lead to increased biased or variance over infeasible QEL. This suggests that QEL iterations could prove useful in reducing finite-sample bias and/or variance.

The remainder of this paper is organized as follows. Section 2 develops a conditional moment model and reviews the estimators of interest. Section 3 describes Brown and Newey's (1998) approach to the efficient estimation of expectation functions under semiparametric assumptions. Section 4 derives the new QEL estimator and Section 5 contains the higher-order asymptotic results (proofs are contained in the Appendix). Section 6 provides Monte Carlo evidence as a check on the robustness of my theoretical results and then validates the estimators against a well-known data set. Section 7 concludes.

2. Conditional Moment Models and Estimators

The statistical model I consider is one with a large, but finite, number of moment conditions. To describe the model, let $\mathcal{Z}_i$ $(i = 1, \ldots, n)$ be *i.i.d.* observations on a data vector $\mathcal{Z}$, β be a

$p \times 1$ parameter vector, and $g(\mathcal{Z}_i, \beta)$ be an $m \times 1$ vector of moment conditions, where $m \geq p$. The population has a true, unknown parameter β_0 satisfying the moment condition

$$\mathbb{E}[g(\mathcal{Z}_i, \beta_0)] = 0, \tag{2.1}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator with respect to the distribution of $\mathcal{Z}_i$. Throughout the remainder of this paper, let $g_i(\beta) = g(\mathcal{Z}_i, \beta)$, $\bar{g}(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\beta)$, and $\hat{\Omega}(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)'$. Finally, let $\bar{\beta}$ be some consistent preliminary estimator of β_0 and let $\mathcal{B}$ denote the parameter space.

An important estimator of β_0 is the optimal GMM estimator of Hansen (1982):

$$\hat{\beta}_{GMM} = \arg\min_{\beta \in \mathcal{B}} \bar{g}(\beta)' \hat{\Omega}(\bar{\beta})^{-1} \bar{g}(\beta). \tag{2.2}$$

In this setting, $\hat{\Omega}(\bar{\beta})$ is the optimal weight matrix evaluated at a consistent preliminary estimate of β_0. When the moment condition is linear in β, as is the case with the linear simultaneous equations model (for instance, $y_i = x_i'\beta + u_i$, $x_i = \Pi' z_i + v_i$, and u_i correlated with v_i), the optimal two-step GMM estimator takes the form

$$\hat{\beta}_{2GMM} = [X'Z\hat{\Omega}(\bar{\beta})^{-1}Z'X]^{-1} X'Z\hat{\Omega}(\bar{\beta})^{-1}Z'Y, \tag{2.3}$$

where X, Y, and Z are just the stacked versions of x_i, y_i, and z_i, respectively. The 2SLS estimator (Theil, 1953) assumes homoskedasticity of ε_i, so it uses the weight matrix $n^{-1} \sum_{i=1}^{n} z_i z_i'$ instead of $\hat{\Omega}$, yielding

$$\hat{\beta}_{2SLS} = [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'Y. \tag{2.4}$$

Another important estimator of β_0 is the CUE of Hansen, Heaton, and Yaron (1996). The CUE is analogous to GMM except that the objective function is also minimized over β in $\hat{\Omega}(\beta)^{-1}$:

$$\hat{\beta}_{CUE} = \arg\min_{\beta \in \mathcal{B}} \bar{g}(\beta)' \hat{\Omega}(\beta)^{-} \bar{g}(\beta), \tag{2.5}$$

where A^{-} denotes any generalized inverse of the matrix A satisfying $AA^{-}A = A$. The LIML estimator (Anderson and Rubin, 1949) can be computed as

$$\hat{\beta}_{LIML} = [X'(\mathbb{I} - \gamma M)X]^{-1} X'(\mathbb{I} - \gamma M)Y, \tag{2.6}$$

where γ is the smallest characteristic root of $W_1 W^{-1}$, $W = (Y, X)' M (Y, X)$, $W_1 = (Y, X)' M_1 (Y, X)$, $M = \mathbb{I} - Z(Z'Z)^{-1}Z'$, $M_1 = \mathbb{I} - \iota(\iota'\iota)^{-1}\iota'$, and ι is an n-vector of ones (in this case, where no exogenous variables, save the constant, are included as explanatory variables). Some studies have shown some equivalence between CUE and LIML.

EL (Owen, 1988, 1990, 1991, 2001; Qin and Lawless, 1994), in contrast to the estimators so far, utilizes an alternative form of the analogy principle (Manski, 1988), minimizing a distance between probability measures rather than the distance of the population moment conditions from their sample counterparts. That is, EL assigns multinomial weights $\{p_i\}_{i=1}^n$ to each of the observations, $\{\mathcal{Z}_i\}_{i=1}^n$, so that $\sum_{i=1}^n p_i g_i(\beta) = 0$. This allows EL to choose probabilities so that the sample moment conditions hold exactly. The optimal p_is maximize the empirical log-likelihood

$$\frac{1}{n}\sum_{i=1}^{n} ln(p_i) \tag{2.7}$$

subject to the sample moment conditions

$$\sum_{i=1}^{n} p_i g_i(\beta) = 0, \tag{2.8}$$

and constraints on the EL probabilities

$$\begin{aligned} \textstyle\sum_{i=1}^{n} p_i &= 1, \\ and\; p_i &\geq 0 \end{aligned} \tag{2.9}$$

for all $i = 1, \ldots, n$. The optimal EL probabilities, $\{p_i^*\}_{i=1}^n$, can be shown to take the form

$$p_i^* = \frac{1}{n}[1 + \lambda' g_i(\beta)]^{-1}, \tag{2.10}$$

where λ is the m-vector of Lagrange multipliers for the sample moment condition constraints. Substituting the optimal EL probabilities back into the empirical log-likelihood function yields

$$lnL_{EL} = -\frac{1}{n}\sum_{i=1}^{n} ln[1 + \lambda' g_i(\beta)]. \tag{2.11}$$

The EL estimator of β_0 is then

$$\hat{\beta}_{EL} = \arg\min_{\beta \in \mathcal{B}} \max_{\lambda \in \Lambda} \frac{1}{n}\sum_{i=1}^{n} ln[1 + \lambda' g_i(\beta)]. \tag{2.12}$$

The first-order necessary condition for an optimum of the empirical log-likelihood function are

$$\frac{1}{n}\sum_{i=1}^{n}[1+\hat{\lambda}_{EL}'g_i(\hat{\beta}_{EL})]^{-1}g_i(\hat{\beta}_{EL})=0 \tag{2.13}$$

with respect to λ and

$$\frac{1}{n}\sum_{i=1}^{n}\{[1+\hat{\lambda}_{EL}'g_i(\hat{\beta}_{EL})]^{-1}\frac{\partial g_i(\hat{\beta}_{EL})}{\partial\beta}\}'\hat{\lambda}_{EL}=0 \tag{2.14}$$

with respect to β, where $\hat{\lambda}_{EL}$ and $\hat{\beta}_{EL}$ denote the implicit solutions to these first-order conditions. Unfortunately, a closed-form solution exists for neither $\hat{\lambda}_{EL}$ or $\hat{\beta}_{EL}$, so one must either solve the first-order conditions numerically or optimize the empirical log-likelihood function directly.

Newey and Smith (2004) show that EL is asymptotically unbiased even when the number of moment conditions is large. This is because EL implicitly estimates the optimal instruments. That is, EL implicitly uses optimal estimates of the Jacobian and optimal weight matrices, $G(\beta)=\mathbb{E}[\partial g_i(\beta)/\partial\beta]$ and $\Omega(\beta)=\mathbb{E}[g_i(\beta)g_i(\beta)']$, respectively. By comparison, optimal GMM uses the simple sample analogs of $G(\beta)$ and $\Omega(\beta)$ – $\hat{G}$ and $\hat{\Omega}$, respectively. The QEL estimator of this paper implements semiparametric efficient estimates of $G(\beta)$ and $\Omega(\beta)$, which turn out to be weighted averages, to mimic the properties of the EL estimator.

3. Semiparametric Efficient Estimation

Brown and Newey (1998) develop techniques to estimate expectation functions efficiently under semiparametric assumptions. In particular, they consider three semiparametric assumptions: independence ($\varepsilon_i \perp z_i$), zero conditional mean ($\mathbb{E}[\varepsilon_i|z_i]=0$), and zero unconditional mean ($\mathbb{E}[g_i(\beta_0)]=0$). The estimators we have considered thus far utilize the latter type of assumption. This section therefore discusses the efficient estimation of $G(\beta)=\mathbb{E}[\partial g_i(\beta)/\partial\beta]$ and $\Omega(\beta)=\mathbb{E}[g_i(\beta)g_i(\beta)']$ under the semiparametric assumption $\mathbb{E}[g_i(\beta_0)]=0$.

But first, consider the estimation of an arbitrary expectation function, $\mathbb{E}[m_i(\beta)]$ when we assume $\mathbb{E}[g_i(\beta_0)]=0$. Brown and Newey show that the semiparametric efficient estimator of $\mathbb{E}[m_i(\beta)]$ takes the form

$$\tilde{m}(\bar{\beta})=\frac{1}{n}\sum_{i=1}^{n}[m_i(\bar{\beta})-c(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}g_i(\bar{\beta})], \tag{3.1}$$

where

$$c(\bar{\beta}) = \frac{1}{n}\sum_{j=1}^{n} m_j(\bar{\beta}) g_j(\bar{\beta}) \tag{3.2}$$

and $\hat{\Omega}$ and $\bar{\beta}$ are defined as before. The semiparametric efficient estimator of the expectation function may be written in a simpler, more intuitive form. Let

$$w_i = 1 - \bar{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1} g_i(\bar{\beta}). \tag{3.3}$$

Then the semiparametric estimator of $\mathbb{E}[m_i(\beta)]$ is

$$\tilde{m}(\bar{\beta}) = \frac{1}{n}\sum_{i=1}^{n} w_i m_i(\bar{\beta}). \tag{3.4}$$

The weights w_i/n are the optimal probability weights of the discrete points of the empirical distribution. This is similar to the result reported in Back and Brown (1993), in which the sample moment conditions are used to construct the implied probabilities associated with the empirical distribution function. Indeed, as I now show, these weights are a first-order Taylor expansion of the optimal EL probabilities.

Proposition 1. *The weights w_i/n proposed by Brown and Newey (1998) are a first-order Taylor expansion of the optimal empirical likelihood probabilities, p_i^*.*

The weights $w_i \equiv 1 - q_i$, where $q_i = \bar{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1} g_i(\bar{\beta})$, have an intuitive interpretation. Suppose, for a moment, that $\bar{g}(\bar{\beta}) < 0$ and $m = 1$ (only one moment condition). Observations with $g_i(\bar{\beta}) < 0$, thus contributing to $\bar{g}(\bar{\beta})$ being negative, have $q_i > 0$. Thus $w_i < 1$ for these observations. Observations with $g_i(\bar{\beta}) > 0$, however, have $q_i < 0$ and $w_i > 1$. Therefore, the method proposed by Brown and Newey gives less weight to observations that contribute to the sample moment conditions not holding, while giving more weight to observations that can cause the sample moment conditions to hold more closely. An analogous result holds for the case in which $\bar{g}(\bar{\beta}) > 0$. This mechanism will cause the sample moment conditions to hold exactly for a given $\bar{\beta}$.

Proposition 2. *If $w_i = 1 - \bar{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1} g_i(\bar{\beta})$ then $n^{-1}\sum_{i=1}^{n} w_i g_i(\bar{\beta}) = 0$.*

Note that $\tilde{m}(\bar{\beta}) = \bar{m}(\bar{\beta}) - \breve{m}(\bar{\beta})$, where $\bar{m}(\bar{\beta}) = n^{-1}\sum_{i=1}^{n} m_i(\bar{\beta})$ and $\breve{m}(\bar{\beta}) = n^{-1}\sum_{i=1}^{n} q_i m_i(\bar{\beta})$. The Brown and Newey weights therefore make a linear adjustment to $\bar{m}(\bar{\beta})$.

The efficient estimators of $G(\beta) = \mathbb{E}[\partial g_i(\beta)/\partial\beta]$ and $\Omega(\beta) = \mathbb{E}[g_i(\beta)g_i(\beta)']$ under the semiparametric assumption $\mathbb{E}[g_i(\beta_0)] = 0$ are thus

$$\tilde{G}(\bar{\beta}) = \frac{1}{n}\sum_{i=1}^{n} w_i \frac{\partial g_i(\bar{\beta})}{\partial\beta} \tag{3.5}$$

and

$$\tilde{\Omega}(\bar{\beta}) = \frac{1}{n}\sum_{i=1}^{n} w_i g_i(\bar{\beta}) g_i(\bar{\beta})', \tag{3.6}$$

where again, $\bar{\beta}$ is a consistent initial estimator of β_0. Under certain conditions (most notably $\mathbb{E}[g_i^3] \neq 0$), Brown and Newey show that these weighted averages are semiparametric efficient relative to their sample average counterparts, since the latter do not use the information contained in the moment conditions. Thus, I expect higher-order gains from using the weighted averages over using the simple averages. I now show how the semiparametric estimation of expectation functions provides a link between GMM and EL estimation.

4. Quasi Empirical Likelihood Estimation

Recall that a closed-form solution exists for neither $\hat{\lambda}_{EL}$ nor $\hat{\beta}_{EL}$, even if $g_i(\beta)$ is linear in β. The first-order condition with respect to λ may be rewritten as

$$\begin{aligned} \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \frac{1}{1+\hat{\lambda}_{EL}' g_i(\hat{\beta}_{EL})} g_i(\hat{\beta}_{EL})' &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \frac{1+\hat{\lambda}_{EL}' g_i(\hat{\beta}_{EL}) - \hat{\lambda}_{EL}' g_i(\hat{\beta}_{EL})}{1+\hat{\lambda}_{EL}' g_i(\hat{\beta}_{EL})} g_i(\hat{\beta}_{EL})' \\ &= \bar{g}(\hat{\beta}_{EL})' - \hat{\lambda}_{EL}' \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \frac{1}{1+\hat{\lambda}_{EL}' g_i(\hat{\beta}_{EL})} g_i(\hat{\beta}_{EL}) g_i(\hat{\beta}_{EL})' \\ &= \bar{g}(\hat{\beta}_{EL})' - \hat{\lambda}_{EL}' \tfrac{1}{n}\textstyle\sum_{i=1}^{n} p_i^* g_i(\hat{\beta}_{EL}) g_i(\hat{\beta}_{EL})' = 0. \end{aligned} \tag{4.1}$$

Similarly, the first-order condition with respect to β is

$$\frac{1}{n}\sum_{i=1}^{n} [p_i^* \frac{\partial g_i(\hat{\beta}_{EL})}{\partial\beta}]' \hat{\lambda}_{EL} = 0. \tag{4.2}$$

Given that the Brown and Newey weights provide a good approximation for the optimal EL probabilities, I propose using an estimate of these weights in the EL first-order conditions instead. This substitution results in the quasi empirical likelihood first-order conditions

$$\begin{aligned} 0 &= \bar{g}(\bar{\beta})' - \hat{\lambda}_{QEL}' \tfrac{1}{n}\textstyle\sum_{i=1}^{n} w_i g_i(\bar{\beta}) g_i(\bar{\beta})' \\ &= \bar{g}(\bar{\beta})' - \hat{\lambda}_{QEL}' \tilde{\Omega}(\bar{\beta}) \end{aligned} \tag{4.3}$$

with respect to λ and

$$\begin{aligned} 0 &= \tfrac{1}{n}\sum_{i=1}^{n}\left[w_i \tfrac{\partial g_i(\bar{\beta})}{\partial \beta}\right]'\hat{\lambda}_{QEL} \\ &= \tilde{G}(\bar{\beta})'\hat{\lambda}_{QEL} \end{aligned} \tag{4.4}$$

with respect to β. Solving the QEL first-order condition with respect to λ for $\hat{\lambda}_{QEL}$ yields

$$\hat{\lambda}_{QEL} = \tilde{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta}). \tag{4.5}$$

Finally, substitute this into the QEL first-order condition with respect to β to obtain the concentrated QEL first-order condition:

$$\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta}) = 0. \tag{4.6}$$

Note that the optimal GMM estimator uses the first-order condition

$$\hat{G}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta}) = 0, \tag{4.7}$$

the components of which are based on simple averages, rather than the semiparametric efficient weighted averages of the QEL estimator.

In the linear simultaneous equations model, $g_i(\beta) = z_i(y_i - x_i'\beta)$ so that $G_i(\beta) = -z_i x_i'$. We can then solve the concentrated QEL first-order condition for β:

$$\hat{\beta}_{QEL} = \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i y_i, \tag{4.8}$$

where, once again, $\bar{\beta}$ is a consistent initial estimator of β_0. The QEL estimator is the estimator whose asymptotic properties I consider throughout the remainder of this paper.

5. Asymptotic Results

The following assumptions are necessary to show consistency of the QEL estimator.

Assumption 1. *$\beta_0 \in \mathcal{B}$ is the unique solution to $\mathbb{E}[g(\mathcal{Z}_i, \beta)] = 0$, $g(\mathcal{Z}_i, \beta)$ is continuous at each $\beta \in \mathcal{B}$, Ω is nonsingular, $\mathcal{B}$ is compact, and $\bar{\beta}$ is a consistent estimator of β_0.*

These are typical assumptions for showing consistency of the two step GMM estimator.

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

The following assumptions are necessary to show asymptotic normality of the QEL estimator.

Assumption 2. *$\beta_0 \in int(\mathcal{B})$, $rank(G) = m \geq p$, and $g(\mathcal{Z}_i, \beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0.*

As with the consistency assumptions, these assumptions are typical for showing asymptotic normality of the two step GMM estimator.

Theorem 2. *If Assumptions 1 and 2 are satisfied then $\sqrt{n}(\hat{\beta}_{QEL} - \beta_0) \xrightarrow{d} N(0, \Sigma)$, where $\Sigma = (G'\Omega^{-1}G)^{-1}$.*

Theorem 3. *If Assumptions 1 and 2 are satisfied then $Bias(\hat{\beta}_{QEL}) = B_I + B^*_{\bar{\beta}}$.*

Newey and Smith show that GMM has asymptotic bias of the form $B_I + B_G + B_\Omega + B_{\bar{\beta}}$, where B_I is the inescapable bias resulting from using the true optimal instruments, B_G is the bias arising from the estimation of G_i, B_Ω is the bias due to the estimation of Ω, and $B_{\bar{\beta}}$ is the bias resulting from the choice of the initial estimator. Newey and Smith also show that EL has asymptotic bias B_I, i.e., EL eliminates three sources of bias. I show that QEL eliminates two sources of bias (B_G and B_Ω) while mitigating the extent of the third ($B^*_{\bar{\beta}} \leq B_{\bar{\beta}}$), thereby nearly replicating the performance of EL but at a much lower cost.

6. Monte Carlo Studies and Validation

I use a simple, linear simultaneous equations model to examine the finite-sample properties of various estimators in a Monte Carlo experiment. The dependent variable, $y_i \in \mathbb{R}$, is linearly related to a single endogenous explanatory variable, $x_i \in \mathbb{R}$, subject to statistical error, $\varepsilon_i \in \mathbb{R}$ so that

$$y_i = x_i'\beta + u_i. \tag{6.1}$$

The endogenous explanatory variable is related to a set of exogenous instrumental variables, $z_i \in \mathbb{R}^m$, subject to statistical error, $v_i \in \mathbb{R}$:

$$x_i = \Pi' z_i + v_i. \tag{6.2}$$

u_i and v_i have a variance-covariance matrix with σ_u^2 and σ_v^2 along the diagonal and σ_{uv} on the off-diagonal.

For the Monte Carlo experiments in this paper, I consider various sample sizes ($n = 250, 1000$), number of moment conditions or instrumental variables ($m = 10, 50$), and degrees of endogeneity

($\sigma_{uv} = 0.3, 0.9$). I choose the first-stage R^2 to be 0.30, thereby pinning down Π (since I assume that $\Pi = c \times \iota$) and avoiding the weak instruments case. I set σ_u^2, σ_v^2, and σ_z^2 equal to one and β_0 equal to zero. I then draw 1,000 replicates of the error terms and the instrumental variables from the standard normal distribution. The key restriction with this setup is that $\mathbb{E}[z_i \varepsilon_i] = 0$, where $\varepsilon_i = u_i - v_i'\beta_0$.

As shown in Tables 1-4, finite-sample bias can indeed be a problem for the 2SLS and GMM estimators. This bias tends to increase with the number of instruments, the degree of endogeneity, small sample sizes, and (in results not reported here) the weakness of the instrument set. Even though the 2SLS and GMM estimators have the smallest variances, finite-sample bias can offset their finite-sample distributions enough that coverage probabilities can be adversely affected. The LIML, CUE, EL, and QEL estimators, however, do not suffer as much from finite-sample bias, but have slightly larger variances than do 2SLS and GMM. Because the bias reduction is so much larger than the variance increase, LIML, CUE, EL, and QEL tend to perform better than 2SLS or GMM in terms of mean-squared error or median absolute error, especially when the number of moment conditions is large, the sample size is small, or when the degree of endogeneity is high. Also, LIML, CUE, EL, and QEL generally exhibit coverage probabilities closer to nominal than do 2SLS and GMM. The QEL estimator seems to do quite well in approximating EL, but at a much lower time cost. Relative to 2SLS and GMM, QEL reduces finite-sample bias while increasing variance and competes well against the LIML and EL estimators.

In all, these simulations support the theoretical results of this and other papers. 2SLS and GMM can exhibit substantial finite-sample bias. LIML, CUE, EL, and QEL reduce this finite-sample bias greatly. However, the reduction in bias generally comes at the cost of increased variance. But in terms of root mean squared error of median absolute error, LIML, EL, and QEL tend to do best. This is particularly the case when the number of moment conditions is large, the sample size is small, or the degree of endogeneity is large.

6.1. **Heteroskedasticity.** As an additional experiment, I take the linear simultaneous equations model and introduce multiplicative heteroskedasticity of the form $\varepsilon_i = u_i z_{1,i}^2 - v_i'\beta_0$. Due to their ability to accommodate heteroskedasticity, GMM, CUE, EL, and QEL have a distinct theoretical advantage over 2SLS and LIML in the presence of heteroskedasticity, since 2SLS and LIML assume homoskedastic errors. This effect shows up clearly in Table 5, in which I assume $m = 50$ and

$\sigma_{uv} = 0.90$ – all other parameters are specified as before. 2SLS and GMM again exhibit finite-sample bias, affecting their coverage probabilities. But now 2SLS and LIML do not have the best variance properties among these estimators. Taking into account both bias and variance, CUE, EL, and QEL tend to do best in terms of root mean squared error and median absolute bias. This result is not surprising since these estimators reduce finite-sample bias and are first-order efficient even under heteroskedasticity.

6.2. **Labor Supply for Married, Working Women.** The empirical application I consider comes from Mroz (1987) as presented in Wooldridge (2002). They consider a labor supply function for working, married women in the United States during 1975 and specify a linear labor supply function,

$$hours = \gamma_{12} ln(wage) + \delta_{10} + \delta_{11} edu + \delta_{12} age + \delta_{13} clt6 + \delta_{14} cge6 + \delta_{15} nwinc + u, \tag{6.3}$$

and a linear wage offer function,

$$ln(wage) = \gamma_{21} hours + \delta_{20} + \delta_{21} edu + \delta_{22} exp + \delta_{23} exp^2 + v, \tag{6.4}$$

where $clt6$ is the number of children under age 6, $cge6$ is the number of children older than age 6, and $nwinc$ is non-wife labor income. The set of instrumental variables include the explanatory variables from the wage offer equation and from the labor supply equation, except for $ln(wage)$. These variables are largely exogenous because of their fixed nature, at least in the short run. The moment conditions are therefore

$$\mathbb{E}[z_i \varepsilon_i] = 0, \tag{6.5}$$

where z_i is the m-vector instrumental variables for observation i, and ε_i is the reduced-form residual from the labor supply equation. Table 6 contains parameter estimates using various estimators. While all the estimators produce similar quantitative results, CUE, EL, and QEL depict a labor supply function that is more responsive to wages, education, age, young children, and non-wife labor income – but less responsive to older children – relative to 2SLS, GMM, and LIML.[1]

[1]Note that we have only one overidentifying restriction, which is one potential reason why the parameter estimates look so similar across estimators.

7. Conclusions

In this paper, I propose the quasi empirical likelihood (QEL) estimator that reduces finite-sample bias in estimating moment condition models but remains computationally simple (much akin to two-step GMM) even when the number of instruments is large. The QEL estimator approximates the method of empirical likelihood via semiparametric efficient estimation. I show that the QEL estimator is nearly unbiased under a higher-order asymptotic approach. QEL, therefore, retains the unbiasedness of EL, but at a fraction of the computational cost. Further, because QEL is robust to the inclusion of many instrumental variables, QEL eliminates the need to choose instruments arbitrarily. QEL also has an intuitive GMM interpretation and is robust to general forms of heteroskedasticity. Simple Monte Carlo experiments confirm these theoretical results.

QEL, however, is not the perfect estimation technique. First, as with all instrumental variable estimators, QEL assumes that the moment conditions hold. That is, the underlying structural errors must be mean independent of the instrumental variables. Second, the QEL estimator essentially trades bias for variance. So even though it reduces finite-sample bias, it is not, in general, efficient. Finally, QEL relies on a consistent initial estimator of the underlying population parameters. A poor initial estimate can lead to increased asymptotic bias or variance, suggesting that iteration of the QEL estimator could prove beneficial.

The QEL estimator proves particularly useful when I validate the estimators against a well-known data set. Mroz (1987) and Wooldridge (2002) examine married, working women labor supply in the United States during 1975. They specify a linear labor supply function and a wage offer function. To the extent that we believe the theoretical finite-sample results, the 2SLS and GMM estimators do seem to exhibit some finite-sample bias in this application, as evidenced by departure from the EL parameter estimates. QEL, on the other hand, seems to eliminate most of this bias.

Several topics for future research are apparent. First, a test of overidentifying restrictions in the QEL context would be useful. This would enable the testing of the validity of the moment conditions. One possible solution is to compare the empirical log-likelihood at the QEL parameter estimates to the unrestricted (no moment conditions) empirical log-likelihood. This test would behave much like the traditional likelihood ratio test. Second, an estimator that achieves Hahn's many-instruments efficiency bound would prove particularly useful. This would require the formulation of a 'Bekker-optimal' weight matrix, in the GMM and QEL context, that minimizes the asymptotic variance of the estimator under Bekker's (1994) asymptotic approach. Lastly, dynamic panel data models offer a very interesting context for QEL estimation. As the panel lengthens (time dimension increases), the number of moment conditions increases at a geometric rate (see, for instance, Alvarez and Arellano, 2003). Since the number of time periods (and especially the number of moment conditions) is typically large when compared to the number of countries, for instance, QEL would be a natural choice for empirical dynamic panel data applications.

Appendix A. Proofs of Main Results

Proof. **Proposition 1**: The weights w_i/n proposed by Brown and Newey (1998) are a first-order Taylor expansion of the optimal empirical likelihood probabilities, p_i^*.

Brown and Newey's weights take the form $w_i = 1 - q_i$, where $q_i = \bar{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}g_i(\bar{\beta})$ and $\bar{\beta} \xrightarrow{p} \beta_0$. The optimal empirical likelihood probabilities take the form $p_i^* = n^{-1}(1+q_i)^{-1}$. The Taylor series approximation of p_i^* around β_0 is $n^{-1}(1-q_i)+O(q_i^2/n) = w_i/n+O(q_i^2) \approx w_i/n$. Thus, the first-order Taylor series approximation of the optimal empirical likelihood probabilities is simply the weighting proposed by Brown and Newey (1998). A similar result holds for the optimal exponential tilting weights of Kitamura and Stutzer (1997). □

Proof. **Proposition 2**: If $w_i = 1 - \bar{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}g_i(\bar{\beta})$ then $n^{-1}\sum_{i=1}^n w_i g_i(\bar{\beta}) = 0$.

$$\begin{aligned} \tfrac{1}{n}\textstyle\sum_{i=1}^n w_i g_i(\bar{\beta}) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n g_i(\bar{\beta})\left[1 - g_i(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta})\right] \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^n g_i(\bar{\beta}) - \tfrac{1}{n}\sum_{i=1}^n g_i(\bar{\beta})g_i(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta}) \\ &= \bar{g}(\bar{\beta}) - \hat{\Omega}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\bar{g}(\bar{\beta}) \\ &= \bar{g}(\bar{\beta}) - \bar{g}(\bar{\beta}) = 0. \end{aligned} \tag{A.1}$$

□

Lemma 1. *If Assumption 1 is satisfied then* $\tilde{G}(\bar{\beta}) \xrightarrow{p} G(\beta_0) = \mathbb{E}[\partial g_i(\beta)/\partial\beta|_{\beta=\beta_0}]$ *and* $\tilde{\Omega}(\bar{\beta}) \xrightarrow{p} \Omega(\beta_0) = \mathbb{E}[g_i(\beta_0)g_i(\beta_0)']$.

Proof. Note that for some generic function $m(\bar{\beta})$, $\tilde{m}(\bar{\beta}) = \bar{m}(\bar{\beta}) - \breve{m}(\bar{\beta})$, where $\bar{m}(\bar{\beta}) = n^{-1}\sum_{i=1}^n m_i(\bar{\beta})$ and $\breve{m}(\bar{\beta}) = n^{-1}\sum_{i=1}^n q_i m_i(\bar{\beta})$. Assumption 1 gives $\bar{\beta} \xrightarrow{p} \beta_0$. So by the continuous mapping theorem, $\bar{m}(\bar{\beta}) - \bar{m}(\beta_0) = o_p(1)$ and $\breve{m}(\bar{\beta}) - \breve{m}(\beta_0) = o_p(1)$. By the law of large numbers, $\bar{m}(\beta_0) \xrightarrow{p} m(\beta_0) = \mathbb{E}[m_i(\beta_0)]$. It remains to be shown that $\breve{m}(\beta_0) \xrightarrow{p} 0$. To show this, write $\breve{m}(\beta_0)$ as

$$\begin{aligned} \tfrac{1}{n}\textstyle\sum_{i=1}^n q_i m_i(\beta_0) &= \tfrac{1}{n^2}\textstyle\sum_{i=1}^n\sum_{j=1}^n g_j(\beta_0)'\Omega(\beta_0)^{-1}g_i(\beta_0)m_i(\beta_0) \\ &= \tfrac{1}{n^2}\textstyle\sum_{i=1}^n g_i(\beta_0)'\Omega(\beta_0)^{-1}g_i(\beta_0)m_i(\beta_0) \\ &\quad + \tfrac{2}{n^2}\textstyle\sum_{i=1}^n\sum_{j<i} g_j(\beta_0)'\Omega(\beta_0)^{-1}g_i(\beta_0)m_i(\beta_0). \end{aligned} \tag{A.2}$$

The first term vanishes asymptotically as long as $\mathbb{E}[g_i'\Omega^{-1}g_i m_i]$ is bounded. Then the first term is $n^{-2}nO_p(1) = O_p(n^{-1}) = o_p(1)$ so that $n^{-2}\sum_{i=1}^n g_i(\beta_0)'\Omega(\beta_0)^{-1}g_i(\beta_0)m_i(\beta_0) \xrightarrow{p} 0$. The second term also vanishes asymptotically, but all that is needed here is independence across observations. Thus, the second term is $n^{-2}n^2 o_p(1) = o_p(1)$ which implies (via V-statistic theorems) that $n^{-2}\sum_{i=1}^{n}\sum_{j<i} g_j(\beta_0)'\Omega(\beta_0)^{-1}g_i(\beta_0)m_i(\beta_0) \xrightarrow{p} 0$. Therefore, since the two terms both converge in probability to zero, $\breve{m}(\beta_0) \xrightarrow{p} 0$. Finally, since $\bar{m}(\beta_0) \xrightarrow{p} m(\beta_0) = \mathbb{E}[m_i(\beta_0)]$ and $\breve{m}(\beta_0) \xrightarrow{p} 0$, $\tilde{m}(\bar{\beta}) = \bar{m}(\bar{\beta}) - \breve{m}(\bar{\beta}) \xrightarrow{p} m(\beta_0) = \mathbb{E}[m_i(\beta_0)]$. Substituting G and Ω in for m yields the desired result. □

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

Write $\hat{\beta}_{QEL}$ as

$$\begin{aligned}
\hat{\beta}_{QEL} &= \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\hat{G}(\bar{\beta})\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\tfrac{1}{n}\textstyle\sum_{i=1}^{n} z_i y_i \\
&= \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\hat{G}(\bar{\beta})\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\tfrac{1}{n}\textstyle\sum_{i=1}^{n} z_i(x_i'\beta_0 + \varepsilon_i) \\
&= \beta_0 + \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\hat{G}(\bar{\beta})\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\tfrac{1}{n}\textstyle\sum_{i=1}^{n} z_i\varepsilon_i \\
&= \beta_0 + \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\hat{G}(\bar{\beta})\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\bar{g}(\beta_0).
\end{aligned} \tag{A.3}$$

As shown in Lemma 1, $\tilde{G}(\bar{\beta}) \xrightarrow{p} G(\beta_0)$ and $\tilde{\Omega}(\bar{\beta}) \xrightarrow{p} \Omega(\beta_0)$. Also, $\bar{g}(\beta_0) \xrightarrow{p} \mathbb{E}[g_i(\beta_0)]$ by the law of large numbers. Thus, $\hat{\beta}_{QEL} \xrightarrow{p} [G'\Omega^{-1}G]^{-1}G'\Omega^{-1}\mathbb{E}[g_i(\beta_0)] = 0$. □

Proof. **Theorem 2**: If Assumptions 1 and 2 are satisfied then $\sqrt{n}(\hat{\beta}_{QEL} - \beta_0) \xrightarrow{d} N(0, \Sigma)$.

Write $\sqrt{n}(\hat{\beta}_{QEL} - \beta_0)$ as

$$\sqrt{n}(\hat{\beta}_{QEL} - \beta_0) = \left[\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\hat{G}(\bar{\beta})\right]^{-1}\tilde{G}(\bar{\beta})'\tilde{\Omega}(\bar{\beta})^{-1}\sqrt{n}\bar{g}(\beta_0). \tag{A.4}$$

As shown in Lemma 1, $\tilde{G}(\bar{\beta}) \xrightarrow{p} G(\beta_0)$ and $\tilde{\Omega}(\bar{\beta}) \xrightarrow{p} \Omega(\beta_0)$. Also, $\sqrt{n}\bar{g}(\beta_0) \xrightarrow{d} N(0, \Omega)$ by the central limit theorem. Thus,

$$\sqrt{n}(\hat{\beta}_{QEL} - \beta_0) \xrightarrow{d} \left(G'\Omega^{-1}G\right)^{-1}G'\Omega^{-1}N(0, \Omega) \tag{A.5}$$

and

$$\sqrt{n}(\hat{\beta}_{QEL} - \beta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1}) \equiv N(0, \Sigma). \tag{A.6}$$

Note that this is the same asymptotic covariance matrix as GMM. □

Proof. **Theorem 3**: If Assumptions 1 and 2 are satisfied then $Bias(\hat{\beta}_{QEL}) = B_I + B^*_{\bar{\beta}}$.

Forthcoming. □

Appendix B. Tables

Table 1: Monte Carlo Results ($m = 10$, $\sigma_{uv} = 0.50$)

$\beta_0 = 0$	Estimator	Mean Bias	Median Bias	Std.Dev.	RMSE	MAE	CovProb	Time
$n = 250$	2SLS	.0289	.0306	.0958	.1001	.0666	.945	0:01
	GMM	.0284	.0312	.0992	.1032	.0698	.947	0:02
	CUE	-.0130	-.0063	.1137	.1144	.0762	.959	0:20
	LIML	-.0116	-.0088	.1060	.1066	.0712	.955	0:03
	EL	-.0129	-.0069	.1121	.1128	.0733	.960	0:26
	QEL	-.0121	-.0074	.1106	.1113	.0718	.960	0:03
	GMM*	.0268	.0287	.0937	.0975	.0656	.951	0:01
	QEL*	-.0113	-.0086	.1109	.1114	.0740	.957	0:02
$n = 1000$	2SLS	.0073	.0097	.0464	.0470	.0330	.946	0:03
	GMM	.0073	.0099	.0465	.0471	.0318	.941	0:05
	CUE	-.0032	.0006	.0480	.0481	.0294	.950	1:02
	LIML	-.0031	-.0009	.0477	.0478	.0305	.953	0:09
	EL	-.0032	.0002	.0479	.0480	.0297	.949	1:17
	QEL	-.0032	-.0001	.0479	.0480	.0298	.949	0:07
	GMM*	.0072	.0097	.0457	.0463	.0311	.941	0:02
	QEL*	-.0032	-.0001	.0480	.0481	.0298	.950	0:04

*=Infeasible estimator of β_0.

Table 2: Monte Carlo Results ($m = 10$, $\sigma_{uv} = 0.90$)

$\beta_0 = 0$	Estimator	Mean Bias	Median Bias	Std.Dev.	RMSE	MAE	CovProb	Time
$n = 250$	2SLS	.0576	.0655	.0913	.1079	.0829	.919	0:01
	GMM	.0576	.0635	.0943	.1105	.0824	.915	0:02
	CUE	-.0166	-.0051	.1134	.1146	.0735	.956	0:23
	LIML	-.0153	-.0081	.1061	.1072	.0683	.949	0:03
	EL	-.0162	-.0069	.1118	.1129	.0725	.955	0:29
	QEL	-.0199	-.0102	.1116	.1133	.0708	.954	0:03
	GMM*	.0542	.0601	.0891	.1043	.0772	.912	0:01
	QEL*	-.0158	-.0073	.1156	.1166	.0764	.954	0:02
$n = 1000$	2SLS	.0148	.0181	.0458	.0481	.0336	.939	0:03
	GMM	.0148	.0171	.0459	.0483	.0334	.938	0:06
	CUE	-.0039	.0005	.0480	.0482	.0296	.949	1:04
	LIML	-.0038	-.0002	.0476	.0478	.0300	.948	0:07
	EL	-.0039	.0000	.0478	.0480	.0298	.949	1:17
	QEL	-.0042	.0000	.0479	.0481	.0298	.950	0:09
	GMM*	.0146	.0167	.0451	.0474	.0328	.937	0:02
	QEL*	-.0040	.0002	.0485	.0487	.0305	.948	0:04

*=Infeasible estimator of β_0.

Table 3: Monte Carlo Results ($m = 50$, $\sigma_{uv} = 0.50$)

$\beta_0 = 0$	Estimator	Mean Bias	Median Bias	Std.Dev.	RMSE	MAE	CovProb	Time
$n = 250$	2SLS	.1558	.1572	.0752	.1730	.1572	.449	0:09
	GMM	.1561	.1573	.0842	.1774	.1573	.534	0:19
	CUE	-.0087	.0130	.2037	.2039	.1222	.943	5:07
	LIML	-.0089	.0009	.1273	.1276	.0754	.948	0:08
	EL	.0014	.0144	.1680	.1680	.0981	.0946	18:38
	QEL	.0475	.0501	.1359	.1440	.0998	.942	0:29
	GMM*	.1117	.1118	.0672	.1304	.1118	.633	0:09
	QEL*	.0368	.0444	.1335	.1384	.0999	.951	0:19
$n = 1000$	2SLS	.0477	.0501	.0434	.0644	.0515	.0820	0:30
	GMM	.0477	.0501	.0459	.0662	.0522	.821	1:01
	CUE	-.0053	-.0011	.0540	.0543	.0365	.941	12:59
	LIML	-.0053	-.0028	.0495	.0498	.0321	.944	0:27
	EL	-.0053	-.0016	.0531	.0534	.0356	.943	32:17
	QEL	-.0053	-.0019	.0530	.0533	.0355	.943	1:31
	GMM*	.0435	.0459	.0422	.0606	.0478	.0820	0:29
	QEL*	-.0048	-.0014	.0534	.0536	.0361	.946	0:59

*=Infeasible estimator of β_0.

Table 4: Monte Carlo Results ($m = 50$, $\sigma_{uv} = 0.90$)

$\beta_0 = 0$	Estimator	Mean Bias	Median Bias	Std.Dev.	RMSE	MAE	CovProb	Time
$n = 250$	2SLS	.2821	.2852	.0581	.2880	.2852	.004	0:09
	GMM	.2829	.2864	.0666	.2906	.2864	.019	0:18
	CUE	-.0228	.0029	.1789	.1803	.1058	.941	5:51
	LIML	-.0118	-.0008	.1115	.1121	.0683	.952	0:09
	EL	-.0104	.0011	.1559	.1563	.0894	.0953	22:15
	QEL	.0568	.0625	.1195	.1323	.0921	.919	0:27
	GMM*	.2122	.2122	.0580	.2200	.2122	.050	0:09
	QEL*	.0702	.0782	.1335	.1508	.1137	.924	0:18
$n = 1000$	2SLS	.0889	.0919	.0404	.0977	.0919	.393	0:30
	GMM	.0890	.0911	.0429	.0988	.0911	.440	1:01
	CUE	-.0059	-.0011	.0524	.0527	.0356	.945	15:19
	LIML	-.0059	-.0032	.0482	.0486	.0319	.948	0:27
	EL	-.0079	-.0045	.0521	.0527	.0355	.944	33:09
	QEL	-.0118	-.0090	.0526	.0539	.0350	.942	1:30
	GMM*	.0814	.0837	.0397	.0906	.0837	.446	0:29
	QEL*	-.0049	-.0017	.0551	.0553	.0377	.947	0:58

*=Infeasible estimator of β_0.

Table 5: Monte Carlo Results (Heteroskedasticity, $m = 50$, $\sigma_{uv} = 0.90$)

$\beta_0 = 0$	Estimator	Mean Bias	Median Bias	Std.Dev.	RMSE	MAE	CovProb	Time
$n = 250$	2SLS	.2285	.2267	.0674	.2382	.2267	.066	0:10
	GMM	.1766	.1751	.0599	.1865	.1751	.162	0:20
	CUE	-0.0069	-.0065	.1296	.1298	.0770	.948	7:04
	LIML	-0.0181	-0.0088	.1233	.1246	.0779	.941	0:09
	EL	-.0029	-.0031	.1912	.1912	.1071	.956	41:04
	QEL	-.0023	-.0025	.1130	.1130	.771	.947	0:30
	GMM*	.1249	.1251	.0494	.1343	.1251	.281	0:10
	QEL*	-.0071	-.0087	.1275	.1277	.1023	.948	0:19
$n = 1000$	2SLS	.0717	.0753	.0434	.0838	.0753	.606	0:30
	GMM	.0639	.0662	.0403	.0756	.0662	.634	1:01
	CUE	-.0042	-.0020	.0489	.0490	.0328	.954	14:45
	LIML	-.0065	-.0044	.0509	.0513	.0331	.939	0:28
	EL	-.0019	-.0017	.0539	.0539	.0332	.953	36:19
	QEL	-.0017	-.0014	.0546	.0546	.0338	.948	1:29
	GMM*	.0587	.0606	.0376	.0698	.0606	.643	0:29
	QEL*	-.0018	-.0015	.0612	.0612	.0406	.951	0:57

*=Infeasible estimator of β_0.

Table 6: Labor Supply Function Estimation Results

Estimator	$Const$	$ln(wage)$	edu	age	$clt6$	$cge6$	$nwinc$
OLS	2114.7	-17.4	-14.4	-7.7	-342.5	-115.0	-4.2
	(340.1)	(54.2)	(18.0)	(5.5)	(100.0)	(30.8)	(3.7)
2SLS	2432.2	1544.8	-177.4	-10.8	-210.8	-47.6	-9.2
	(594.2)	(480.7)	(58.1)	(9.6)	(176.9)	(56.9)	(6.5)
GMM	2421.9	1638.3	-184.8	-10.8	-229.8	-44.3	-9.7
	(611.2)	(592.9)	(66.5)	(10.6)	(203.2)	(56.4)	(5.2)
CUE	2482.3	1838.6	-205.0	-11.9	-228.3	-37.4	-10.3
	(690.1)	(670.2)	(75.3)	(11.9)	(227.5)	(63.7)	(5.9)
LIML	2449.3	1629.1	-186.2	-10.9	-203.7	-43.9	-9.5
	(615.6)	(498.1)	(60.2)	(9.9)	(183.3)	(59.0)	(6.7)
EL	2479.0	1828.0	-204.1	-11.7	-221.3	-37.8	-10.3
	(694.4)	(694.7)	(78.0)	(11.9)	(224.2)	(63.5)	(6.1)
QEL	2474.3	1839.1	-205.3	-11.6	-221.5	-37.5	-10.4
	(600.8)	(537.7)	(61.8)	(10.2)	(202.4)	(55.8)	(5.2)

References

[1] Altonji, J. and L.M. Segal (1996), 'Small sample bias in GMM estimation of covariance structures,' *Journal of Business and Economic Statistics*, **14**, 353-366.

[2] Alvarez, J. and M. Arellano (2003), 'The time series and cross-section asymptotics of dynamic panel data estimators,' *Econometrica*, **71**, 1121-1159.

[3] Anderson, T.W. and H. Rubin (1949), 'Estimator of the parameters of a single equation in a complete system of stochastic equations,' *Annals of Mathematical Statistics*, **20**, 46-63.

[4] Back, K. and D.P. Brown (1993), 'Implied probabilities in GMM estimators,' *Econometrica*, **61**, 971-975.

[5] Bekker, P.A. (1994), 'Alternative approximations to the distributions of instrumental variable estimators,' *Econometrica*, **62**, 657-681.

[6] Brown, B.W. and W.K. Newey (1998), 'Efficient semiparametric estimation of expectations,' *Econometrica*, **66**, 453-464.

[7] Donald, S.G. and W.K. Newey (2001), 'Choosing the number of instruments,' *Econometrica*, **69**, 1161-1191.

[8] Hahn, J. (2002), 'Optimal inference with many instruments,' *Econometric Theory*, **18**, 140-168.

[9] Hansen, L.P. (1982), 'Large sample properties of generalized method of moments estimators,' *Econometrica*, **50**, 1029-1054.

[10] Hansen, L.P., J. Heaton, and A. Yaron (1996), 'Finite-sample properties of some alternative GMM estimators,' *Journal of Business and Economic Statistics*, **14**, 262-280.

[11] Imbens, G.W. (1997), 'One-step estimators for over-identified generalized method of moments models,' *Review of Economic Studies*, **64**, 359-383.

[12] Imbens, G.W., R.H. Spady, and P. Johnson (1998), 'Information theoretic approaches to inference in moment condition models,' *Econometrica*, **66**, 333-357.

[13] Kitamura, Y. (2001), 'Asymptotic optimality of empirical likelihood for testing moment restrictions,' *Econometrica*, **69**, 1661-1672.

[14] Kitamura, Y. and M. Stutzer (1997), 'An information theoretic alternative to generalized method of moments estimation,' *Econometrica*, **65**, 861-874.

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

Mroz, T.A. (1987), 'The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions,' *Econometrica*, **55**, 765-799.

[16] Newey, W.K. and D. McFadden (1994), 'Large sample estimation and hypothesis testing,' in Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, New York: North Holland.

[17] Newey, W.K. and R.J. Smith (2004), 'Higher order properties of GMM and generalized empirical likelihood estimators,' *Econometrica*, **72**, 219-255.

[18] Owen, A.B. (2001), *Empirical Likelihood*, New York: Chapman and Hall.

[19] Owen, A. (1991), 'Empirical likelihood for linear models,' *Annals of Statistics*, **19**, 1725-1747.

[20] Owen, A. (1990), 'Empirical likelihood ratio confidence regions,' *Annals of Statistics*, **18**, 90-120.

[21] Owen, A. (1988), 'Empirical likelihood ratio confidence intervals for a single functional,' *Biometrika*, **75**, 237-249.

[22] Qin, J. and J. Lawless (1994), 'Empirical likelihood and general estimating equations,' *Annals of Statistics*, **22**, 300-325.

[23] Staiger, D. and J.H. Stock (1997), 'Instrumental variable regression with weak instruments,' *Econometrica*, **65**, 557-586.

[24] Theil, H. (1953), 'Repeated least-squares applied to complete equation systems,' mimeo, The Hague: Central Planning Bureau.

[25] Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge (MA): MIT Press.