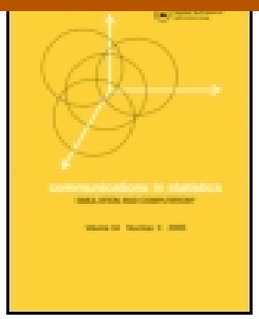

Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

On the Finite-sample Properties of Conditional Empirical Likelihood Estimators

Federico Crudu & Zsolt Sándor

To cite this article: Federico Crudu & Zsolt Sándor (2015): On the Finite-sample Properties of Conditional Empirical Likelihood Estimators, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2014.999090

To link to this article: http://dx.doi.org/10.1080/03610918.2014.999090

Accepted author version posted online: 17 Sep 2015.

ON THE FINITE-SAMPLE PROPERTIES OF CONDITIONAL EMPIRICAL LIKELIHOOD ESTIMATORS*

Federico Crudu†
Pontificia Universidad Católica de Valparaíso
University of Sassari and CRENoS

Zsolt Sándor‡
Sapientia University, Miercurea Ciuc
Affiliate Fellow at CERGE-EI, Prague

Abstract

We provide Monte Carlo evidence on the finite sample behavior of the conditional empirical likelihood (CEL) estimator of Kitamura, Tripathi, and Ahn (2004) and the conditional Euclidean empirical likelihood (CEEL) estimator of Antoine, Bonnal, and Renault (2007) in the context of a heteroskedastic linear model with an endogenous regressor. We compare these estimators with three heteroskedasticity-consistent instrument-based estimators and the Donald, Imbens and Newey (2003) estimator in terms of various performance measures. Our results suggest that the CEL and CEEL with fixed bandwidths may suffer from the no-moment problem, similarly to the unconditional generalized empirical likelihood estimators studied by Guggenberger (2008). We also study the CEL and CEEL estimators with automatic bandwidths selected through cross-validation. We do not find evidence that these suffer from the no-moment problem. When the instruments are weak, we find CEL and CEEL to have finite sample properties –in terms of mean squared error and coverage probability of confidence

*Financial support from Marie Curie Excellence Grant MEXT-CT-2006-042471 is gratefully acknowledged.

†federico.crudu@gmail.com, Pontificia Universidad Católica de Valparaíso, Instituto de Estadística, Avenida Errazuriz 2734, Valparaíso, Chile. Financial support from Regione Autonoma della Sardegna Master and Back Grant PRR-MAB-A2011-24192 is gratefully acknowledged.

‡sandorzsolt@sapientia.siculorum.ro, Sapientia University, Faculty of Economic and Human Sciences, Piata Libertatii 1, 530104 Miercurea Ciuc, Romania.

intervals– poorer than the heteroskedasticity-consistent Fuller (HFUL) estimator. In the strong instruments case the CEL and CEEL estimators with automatic bandwidths tend to outperform HFUL in terms of mean squared error, while the reverse holds in terms of the coverage probability, although the differences in numerical performance are rather small.

ACCEPTED MANUSCRIPT

1 Introduction

Motivated by the practical importance of models defined by conditional moment restrictions, a number of recent important contributions have proposed empirical likelihood-based techniques for estimation and inference of this class of models. Kitamura, Tripathi, and Ahn (2004, KTA henceforth) develop a conditional empirical likelihood estimator of these models. Antoine, Bonnal, and Renault (2007, ABR henceforth) introduce an estimator based on a related idea that instead of the empirical likelihood uses the Euclidean likelihood. A common way of dealing with conditional moment restrictions is to reduce them to unconditional ones by means of instruments. This procedure typically leads to a large number of potentially useful instruments, which implies the problem of selecting a reduced number of most useful instruments. An alternative way to proceed is to construct asymptotically optimal instruments (e.g., Newey 1993). The KTA and ABR estimators are appealing from an asymptotic theoretical point of view as they are able to achieve semiparametric first-order asymptotic efficiency without computing the optimal instruments.

The literature has paid considerable attention to conditional empirical likelihood estimators. Smith (2007) generalizes the conditional empirical likelihood (CEL) of KTA and the conditional euclidean empirical likelihood (CEEL) of ABR to the class of local Cressie-Read discrepancies, where the term *local* refers to the explicit use of kernel weights. He shows that the estimators

ACCEPTED MANUSCRIPT

of the local Cressie-Read class are first order asymptotically equivalent to the CEL and CEEL (CE(E)L for short) estimators. A few other recent contributions stress the potential of the conditional generalized empirical likelihood (GEL) framework from an asymptotic theory point of view. Gospodinov and Otsu (2012) show that in an AR(1) model with iid errors the local GMM estimator, which is essentially the same as CEEL, has a higher order asymptotic bias smaller than the OLS estimator. Tripathi and Kitamura (2003) show that a test statistic for conditional moment restrictions based on the CEL objective function is asymptotically optimal in terms of a certain average power criterion.

A conclusion of these papers is that empirical likelihood-based estimators are rather appealing for conditional moment restriction models from an asymptotic theory point of view. However, although some of these papers present finite sample studies of these estimators, none of them provides information on their finite sample performance in the important class of models with endogenous regressors. Another problem that is important in practice is that, although the CE(E)L are instrument-free methods, they depend on additional unknown parameters, that is, bandwidths. The asymptotic theory of these estimators specifies the rate at which the bandwidths should change with the sample size in order to obtain asymptotic efficiency, but this does not provide a clear indication on how to choose the bandwidths in practice. For some models (e.g., the linear heteroskedastic

ACCEPTED MANUSCRIPT

model in KTA, or the AR(1) model with ARCH errors in Gospodinov and Otsu, 2012) different bandwidth values lead to similar estimates. For models with endogenous regressors, however, it is not known to what extent the finite sample performance of these estimators is affected, if one uses different bandwidths, or if one uses some bandwidth selection procedure.

There are at least two reasons to expect CE(E)L to perform poorly in models with endogenous regressors, especially when the instruments are weak. First, these estimators are the result of a saddle point optimization problem, which may have extensive flat parts near the optimum. Second, for a linear model with an endogenous regressor, Guggenberger (2008) finds that the unconditional GEL estimators suffer from the no-moment problem. These estimators are also obtained as the outcome of a saddle point optimization problem, compared to which the dimensionality of the optimization problem increases considerably in the conditional moment case.

Due to these considerations we find it important to investigate how the CE(E)L estimators perform in finite samples. In order to do so, we conduct a Monte Carlo experiment, in which we estimate a one-parameter linear model with an endogenous regressor and heteroskedasticity using several estimators: CEL (KTA), CEEL (ABR), GMM, HLIM, HFUL (the latter two from Hausman et al., 2012) and the estimator proposed by Donald, Imbens and Newey (2003). For the CE(E)L estimators we use a grid search on a very fine grid in a rather large interval around the true value in

ACCEPTED MANUSCRIPT

order to circumvent possible convergence problems of standard algorithms like Newton-Raphson or nonderivative simplex search. Since the CE(E)L estimators depend on unknown bandwidths, we compute these estimators for a small grid of fixed bandwidth values, and then out of these we select the best bandwidth according to a cross-validation criterion proposed by Newey (1993). We then evaluate the performance of the estimators according to a range of criteria. Newey's cross-validation criterion uses the second-order remainder term from estimation of the optimal instruments. We provide intuition for the fact that this cross-validation acts in a way similar to regularization in that it tends to eliminate estimates that are far from zero.

Due to their similarity to unconditional GEL estimators, the CE(E)L estimators may also suffer from the no-moment problem. Therefore, interpretation of quadratic loss measures such as standard deviation and mean square error computed from Monte Carlo samples should be dealt with care. In order to avoid potential problems of interpretation, in addition to the standard measures of performance, we also look at performance measures like the median absolute error, the nine-decile range, and the tail probability, which do not depend on moments. A related concept is the probability of concentration used by Morimune (1983). Fiebig (1985) provides examples on how some estimators with no moments may be preferred to others that have moments. He suggests as a general evaluation criterion in this case the concentration of the estimator around the true parameter.

ACCEPTED MANUSCRIPT

In this respect, his probability of concentration criterion (Fiebig, 1985, equation (2)) is virtually the same as the tail probability statistic used in this paper and also in Guggenberger (2008).

Our results suggest that CEL and CEEL perform rather similarly. Both estimators computed with fixed bandwidths may suffer from the no-moment problem. We draw this conclusion from the fact that both estimators perform similarly to the HLIM estimator, which is known to have the no-moment problem (Hausman et al., 2012). We do not find evidence that the CE(E)L estimators with bandwidths computed by cross-validation have the no-moment problem. In addition, these estimators outperform their fixed bandwidth counterparts, especially in the weak instruments case. In this case, these estimators are outperformed by the HFUL estimator (Hausman et al., 2012), but in the strong instruments case they have competitive finite sample properties with respect to the other estimators.

The remainder of the paper is organized as follows: Section 2 describes the Monte Carlo setup and the estimators, while in Section 3 we discuss the implementation and the results. Section 4 collects some final remarks and, finally, the Appendix contains the tables and some technical details on estimation and cross validation.

ACCEPTED MANUSCRIPT

2 Monte Carlo experiment

In this section we describe the data generating process (DGP) in our Monte Carlo experiment and present the estimators that we study. For our DGP we consider a linear model with heteroskedastic errors that is similar to the one considered by Hausman et al. (2012). Specifically,

$$y_i = \beta_0 x_i + \varepsilon_i, \quad i = 1, ..., n,$$

where x_i is expected to be endogenous and the exogenous variable z_i is observed. The parameter β_0 is identified by the conditional moment restriction

$$E\left(g\left(y_i, x_i, \beta\right) | z_i\right) = 0, \tag{1}$$

where $g\left(y_i, x_i, \beta\right) = y_i - \beta x_i$. Regarding the primitives of our DGP we assume that

$$x_i = \pi z_i + u_i$$

where $z_i \sim N(0, 1)$, $u_i \sim N(0, 1)$, and

$$\varepsilon_i = \rho u_i + \sqrt{\frac{1 - \rho^2}{\phi^2 + .86^4}} \left(\phi v_{1i} + .86 v_{2i}\right)$$

ACCEPTED MANUSCRIPT

with $v_{1i} \sim N\left(0, z_i^2\right)$, $v_{2i} \sim N\left(0, .86^2\right)$. The parameter ϕ is computed from the theoretical R^2 for the regression of ε_i^2 on z_i^2, that is,

$$R^2 = \frac{Var\left(E\left(\varepsilon_i^2 \left| z_i^2\right.\right)\right)}{Var\left(E\left(\varepsilon_i^2 \left| z_i^2\right.\right)\right) + E\left(Var\left(\varepsilon_i^2 \left| z_i^2\right.\right)\right)}$$

for given values of R^2. This latter quantity measures the degree of heteroskedasticity, while ρ determines the degree of endogeneity because $corr\left(x_i, \varepsilon_i\right) = \rho / \sqrt{(1+\pi^2)}$. We consider two sample sizes $n = 100$ and $n = 500$. For each sample size we have the following parameter combinations

$$(\rho, R^2, \phi) = (0.75, 0.1, 1.863521),$$

$$(\rho, R^2, \phi) = (0, 0.1, 0.5765206),$$

$$(\rho, R^2, \phi) = (0.75, 0, 0).$$

The first parameter combination implies a rather large degree of endogeneity accompanied by heteroskedasticity; the second parameter combination includes heteroskedasticity but has no endogeneity; the third parameter combination considers the presence of endogeneity and homoskedastic errors.[1] We vary the strength of instruments z_i by taking $\pi = 0.4$ and $\pi = 0.04$; the latter value provides instruments with strength comparable to that in Guggenberger (2008), where in the case

[1]It would be desirable to study the case of high degree of heteroskedasticity as well. However, this does not seem to be possible within the current DGP because the restriction that ε_i has unconditional variance equal to 1 restricts ρ and R^2 so that R^2 cannot take values much higher than 0.2.

ACCEPTED MANUSCRIPT

of one instrument the lowest correlation between the endogenous regressors and instruments is 0.032.[2] Whenever we use estimators that require instruments, we consider the following two sets of 10 and 30 instruments:

$$\bar{z}_i = \left(1, z_i, z_i^2, z_i^3, z_i^4, z_i D_{1i}, ..., z_i D_{5i}\right)', \tag{2}$$

$$\bar{z}_i = \left(1, z_i, z_i^2, z_i^3, z_i^4, z_i D_{1i}, ..., z_i D_{25i}\right)',$$

where the variable D_{ki} is a dummy variable that takes value 1 with probability 0.5. Similar dummies are used by Hausman et al. (2012).

In the next sections we describe the estimators that we consider.

2.1 Conditional empirical likelihood estimators

In this section we describe the CEL and CEEL estimators. These estimators are the result of a constrained optimization of certain nonparametric objective functions, where one of the constraints is the sample analog of the conditional moment restriction. The nonparametric objective functions are a nonparametric version of the log-likelihood function for CEL, and a local quadratic Cressie-Read discrepancy criterion for CEEL, respectively (see KTA and ABR for further details, as well as Smith (2007) for a unified treatment based on Cressie-Read discrepancy). In practice

[2]The results for $\pi = 0.4$ are shown only in the case of $n = 100$.

both estimators can be obtained from unconstrained optimizations of the so-called dual objective functions, which are derived from the first order conditions of the constrained optimization. These dual problems have the feature that they are saddle point optimization problems.

In particular, the CEL estimator of β_0 is

$$\widehat{\beta}_{CEL} = \arg\min_{\beta} \max_{\gamma_i, i=1,\ldots,n} \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} \log\left(1 + \gamma_i g\left(y_j, x_j, \beta\right)\right), \tag{3}$$

where w_{ij}, $i, j = 1, \ldots, n$ are defined as

$$w_{ij} = \frac{K\left(\frac{z_i - z_j}{b_n}\right)}{\sum_{j=1}^{n} K\left(\frac{z_i - z_j}{b_n}\right)}, \tag{4}$$

that is, the weights of the Nadaraya-Watson nonparametric regression estimator, K is a density function on $\mathbb{R}$, symmetric around 0, playing the role of a kernel function, and γ_i, $i = 1, \ldots, n$ are the Lagrange multipliers in the constrained maximization of the original objective function. Determining the CEL estimator from the dual (3) involves the first step maximization with respect to these Lagrange multipliers. A computationally efficient method for determining the Lagrange multipliers is discussed in the Appendix in Section B.1.

The CEEL estimator is

$$\widehat{\beta}_{CEEL} = \arg\min_{\beta} \sum_{i=1}^{n} \left(\frac{\widehat{g}(\beta)^2}{\sum_{j=1}^{n} w_{ij} g\left(y_j, x_j, \beta\right)\left(g\left(y_j, x_j, \beta\right) - \widehat{g}(\beta)\right)} \right), \tag{5}$$

where $\widehat{g}_i(\beta) = \sum_{j=1}^{n} w_{ij} g\left(y_j, x_j, \beta\right)$ with weights given in (4). Differently from the CEL estimator, the CEEL estimator does not require optimization with respect to the Lagrange multipliers. This

is because the quadratic Cressie-Read discrepancy criterion implies first-order conditions of the constrained optimization that allow for explicit expressions of the Lagrange multipliers. Therefore, although not directly visible in the CEEL-objective function (5), CEEL estimation is also a saddle point problem. We also note that the CEEL estimator is numerically identical to a conditional generalization of the continuously updated GMM estimator of Hansen et al. (1996) (see ABR for further details).

The limiting distribution of the CE(E)L estimators is the same. That is, for $k = CEL, CEEL$

$$\sqrt{n}\left(\widehat{\beta}_k - \beta_0\right) \to_d N(0, V),$$

where

$$V = \left(E\left[D(z)^2\,\Omega(z)^{-1}\right]\right)^{-1}$$

is the semiparametric lower bound with $D(z) = E\left[\frac{\partial g(y,x,\beta)}{\partial \beta}|z\right]$, $\Omega(z) = E\left[g(y,x,\beta)^2|z\right]$. The asymptotic variance of $\widehat{\beta}_k$, $k = CEL, CEEL$, can be estimated as

$$\widehat{V}_k = \left(\sum_{i=1}^{n} \widehat{D}(z_i)^2 \left(\widehat{\Omega}\left(\widehat{\beta}_k, z_i\right)\right)^{-1}\right)^{-1},$$

where $\widehat{\Omega}(z_i)$, $\widehat{\Omega}\left(\widehat{\beta}_k, z_i\right)$ are nonparametric regression estimators of $D(z_i)$, $\Omega\left(\widehat{\beta}_k, z_i\right)$. Specifically,

the Nadaraya-Watson nonparametric kernel regression estimators for our DGP are

$$\widehat{D}(z_i) = -\sum_{j=1}^{n} w_{ij} x_j,$$

and

$$\widehat{\Omega}\left(\widehat{\beta}_{CE(E)L}, z_i\right) = \sum_{j=1}^{n} w_{ij} g\left(y_j, x_j, \widehat{\beta}_{CE(E)L}\right)^2.$$

We use cross-validation for both the CEL and CEEL estimators. The cross-validation criterion used is presented in Appendix B.2. There we also show that in the case of our model for large values of the estimate the cross-validation criterion is also large. This appears to suggest that the cross-validation criterion tends to select values that are closer to zero. This is a feature similar to regularized estimators (e.g., Hausman et al. 2011).

2.2 The Donald, Imbens and Newey (2003) estimator

DIN consider efficient estimation of conditional moment restriction models via empirical likelihood estimation. Conditional moments are transformed into unconditional ones by means of splines or other approximating functions. The problem we consider is the same problem defined in equation (1). The DIN estimator requires the specification of a $K \times 1$ vector of spline approximating

functions

$$q^K(z) = (1, z, z^2, \ldots, z^s, \xi(z - t_1)^s), \ldots \xi(z - t_{K-s-1})^s)'$$

where $\xi(z) = \mathbf{1}(z > 0)z$, $\mathbf{1}(\cdot)$ is the indicator function taking value 1 if the condition $\cdot$ is true and 0 otherwise and $t_1, \ldots, t_{K-s-1}$ is a set of knots. In our simulations $s = 3$ and $K = 2, 3, 5, 10, 15$. The knots are to be found equispaced in the range of the variable z. In practice we choose the knots in a set of 21 equispaced quantiles of z. The conditional moment restriction (1) implies

$$E(q^K(z_i)(y_i - x_i\beta)) = 0.$$

From this set of moments we can define the EL problem

$$P(\beta, \pi_i, \gamma, \mu) = \sum_{i=1}^{n} \log \pi_i + \gamma' \sum_{i=1}^{n} \pi_i q^K(z_i)(y_i - x_i\beta) + \mu\Big(\sum_{i=1}^{n} \pi_i - 1\Big).$$

The estimation of the parameters of interests can be implemented by solving the saddle point problem

$$(\widehat{\beta}_{DIN}, \widehat{\gamma}_{DIN})' = \arg\min_{\beta} \max_{\gamma} \sum_{i=1}^{n} \log(1 - \gamma' q^K(z_i)(y_i - x_i\beta))$$

and $\widehat{\pi} = (n(1+\widehat{\gamma}' q^K(z_i)(y_i - x_i\widehat{\beta})))^{-1}$. The asymptotic distribution of $\widehat{\beta}_{DIN}$ is the same as the CE(E)L estimators. An estimator of the asymptotic variance is given by

$$\widehat{V}_{DIN} = (\widehat{D}'\widehat{\Omega}^{-1}\widehat{D})^{-1}$$

where $\widehat{D} = \sum_{i=1}^{n} \widehat{\pi}_i q^K(z_i) x_i$ and $\widehat{\Omega} = \sum_{i=1}^{n} \widehat{\pi}_i (y_i - x_i\beta)^2 q^K(z_i) q^K(z_i)'$.

2.3 Instrumental variable estimation

Suppose that we have an $L \times 1$ vector of instrumental variables $\overline{z}_i$ as described in (2). Then the conditional moment (1) implies the unconditional moment restrictions

$$E\left(\overline{z}_i\left(y_i - x_i\beta\right)\right) = 0,$$

which leads to estimation by means of GMM. GMM estimation generally requires a two step procedure. The first step estimator is given by the minimum of

$$Q_{GMM}(\beta) = (y - x\beta)' \, ZWZ' \, (y - x\beta)$$

ACCEPTED MANUSCRIPT

for y and x being $n \times 1$ vectors of observations and Z is a $n \times L$ matrix, such that its *ith* row is $\overline{z}_i'$. The resulting first step estimator is defined as

$$\widehat{\beta}_1 = (x'ZWZ'x)^{-1} x'ZWZ'y$$

for a certain positive definite matrix W. In our simulations W is chosen to be the identity matrix. In order to achieve efficiency and robustness with respect to heteroskedasticity, in the second step we use an Eicker-White matrix (White, 1980):

$$\widehat{\beta}_{GMM} = \left(x'Z\widehat{\Omega}^{-1}Z'x\right)^{-1} x'Z\widehat{\Omega}^{-1}Z'y$$

where $\widehat{\Omega} = \sum_{i=1}^{n} \left(y_i - x_i\widehat{\beta}_1\right)^2 \overline{z}_i\overline{z}_i'$. The GMM estimator is normally distributed

$$\sqrt{n}\left(\widehat{\beta}_{GMM} - \beta\right) \rightarrow_d N\left(0, V_{GMM}\right)$$

and $V_{GMM} = \left(E\left(x'Z\right)\Omega^{-1}E\left(Z'x\right)\right)^{-1}$, for $\Omega = E\left(Z'\left(y - x\beta_0\right)\left(y - x\beta_0\right)'Z\right)$, which we estimate by $\widehat{V}_{GMM} = \left(x'Z\widehat{\Omega}^{-1}Z'x\right)^{-1}$.

In a recent paper Hausman et al. (2012) describe a simple one-step estimator that is robust to the presence of heteroskedasticity and many instruments. Such an estimator is similar to LIML and it is based on jackknife techniques. Let us first define the projection matrix $P_Z = Z\left(Z'Z\right)^{-1}Z'$

ACCEPTED MANUSCRIPT

and the diagonal matrix D_{P_Z}, whose diagonal elements are the diagonal entries of P_Z. Then, the so called HLIM estimator is computed as the minimum of

$$Q_{HLIM}(\beta) = \frac{\binom{1}{-\beta}' A \binom{1}{-\beta}}{\binom{1}{-\beta}' B \binom{1}{-\beta}}$$

with

$$A = (y, x)' \left(P_Z - D_{P_Z}\right)(y, x), \quad B = (y, x)'(y, x)$$

and is equal to

$$\widehat{\beta}_{HLIM} = \left(x'\left(P_Z - D_{P_Z}\right)x - \lambda_{HLIM} x'x\right)^{-1}\left(x'\left(P_Z - D_{P_Z}\right)y - \lambda_{HLIM} x'y\right) \tag{6}$$

where λ_{HLIM} is the minimum eigenvalue of the matrix $B^{-1}A$. This estimator shares some features with LIML, most notably it may not have moments (Hausman et al., 2012, p. 217) in the weak instruments case. These authors propose a correction in the spirit of Fuller (1977), where the eigenvalue λ_{HLIM} is replaced by

$$\lambda_{HFUL} = \frac{\lambda_{HLIM} - \frac{1-\lambda_{HLIM}}{n} C}{1 - \frac{1-\lambda_{HLIM}}{n} C}.$$

The parameter C is chosen by the econometrician and following the suggestion of Hausman et al.

(2012) we set $C = 1$. The so called HFUL estimator is then defined as

$$\widehat{\beta}_{HFUL} = \left(x'\left(P_Z - D_{P_Z}\right)x - \lambda_{HFUL}x'x\right)^{-1}\left(x'\left(P_Z - D_{P_Z}\right)y - \lambda_{HFUL}x'y\right). \tag{7}$$

For $k = HLIM, HFUL$ we have the following convergence in distribution:

$$\frac{\widehat{\beta}_k - \beta_0}{\sqrt{\widehat{V}_k}} \to_d N(0,1),$$

where

$$\widehat{V}_k = \widehat{M}^{-1}\widehat{S}\widehat{M}^{-1}, \quad \widehat{M} = x'\left(P_Z - D_{P_Z}\right)x - \lambda_k x'x$$

and

$$\widehat{S} = \sum_{i=1}^{n}\left(\dot{x}_i^2 - 2p_{ii}\widehat{x}_i\dot{x}_i\right)\widehat{\varepsilon}_i^2 + \sum_{t=1}^{L}\sum_{s=1}^{L}\left(\sum_{i=1}^{n}\widetilde{Z}_{it}\widetilde{Z}_{is}\widehat{\varepsilon}_i\right)\left(\sum_{j=1}^{n}Z_{jt}Z_{js}\widehat{\varepsilon}_j\right),$$

for, $\widetilde{Z} = Z\left(Z'Z\right)^{-1}$, $\widehat{\varepsilon} = y - x\widehat{\beta}_k$, $\widehat{x} = x - \frac{\widehat{\varepsilon}x'\widehat{\varepsilon}}{\widehat{\varepsilon}'\widehat{\varepsilon}}$, $\dot{x} = P_Z\widehat{x}$; furthermore, p_{ii} is the ith diagonal element of P_Z. The limit of $\widehat{V}_k$ is provided in Hausman et al. (2012).

3 Implementation and results

We implement the CE(E)L estimators by using the Epanechnikov kernel:

$$K(u) = \frac{3}{4}\left(1 - u^2\right)\cdot \mathbf{1}\left(|u| \le 1\right),$$

where $\mathbf{1}(\cdot)$ is the indicator function. For the two sample sizes $n = 100$ and 500 we use the bandwidths b_n in the set

$$b_n \in \{0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9\}.$$

As mentioned in the previous section, the CE(E)L estimators are the solution of a saddle point problem. Therefore, in certain situations that typically occur when the instruments are weak, the corresponding objective function may be very flat in the neighborhood of the optimum, causing the failure of standard optimization routines. In order to avoid this, we solve the optimization problem by means of a grid search. The grid we consider is between -25 and 25 and has step length 0.01.[3] In order to provide a fair comparison of performance, we also restrict the other estimates to the interval $[-25, 25]$. We note that Guggenberger (2008) uses the same grid search approach in his study of unconditional GEL.

In order to provide some insight on the difficulty of solving a saddle-point optimization problem, we make a few remarks on the behavior of the CE(E)L estimators for different bandwidths. First, in cases when the objective functions in (3) and (5) do not have flat parts around the optimum for any given bandwidth, the objective functions are similar, and, as a consequence, the estimates corresponding to different bandwidths will also be similar (the cases studied by KTA in their Monte

[3]This approach is not attractive from a computational point of view, in particular when the dimension of the parameter of interest is larger than one. However, it is ideal for our simulation environment where we have to tackle situations where the instruments provided by our DGP are particularly weak.

ACCEPTED MANUSCRIPT

Carlo experiments appear to be of this type). Second, whenever, for some bandwidths the objective function is flat near the optimum, the estimates corresponding to different bandwidths may be very different. We illustrate this phenomenon by plotting the objective function in these two cases.

In Figure 1 we present the CEEL objective function for $n = 100$ in a case with low endogeneity ($\rho = 0.3$) and strong instruments ($\pi = 0.4$) for four different bandwidth values. We can see that the objective function is well-behaved in the sense that we can clearly distinguish a global minimum in the case of each bandwidth. The global minima in the four cases occur at values close to 0, which is the true parameter value.

On the contrary, in the presence of weak instruments and high endogeneity, the objective function may be characterized by multiple local optima and extensive flat parts in the vicinity of the minimum. This situation is well depicted in Figure 2, where we plot the CEEL objective function again for $n = 100$ and four different bandwidth values in a case with high endogeneity ($\rho = 0.75$) and weak instruments ($\pi = 0.04$). In this figure we can see that for bandwidths $b_n = 0.5$ and 0.7 there are two minima for which the value of the objective function is quite similar. For bandwidth values $b_n = 1.7$ and 1.9 the objective function degenerates so that the minimum falls in a region where the objective function is very flat. This figure illustrates the pathological features of the optimization problem in the case of weak instruments, and provides an argument for using grid

ACCEPTED MANUSCRIPT

search instead of standard optimization routines such as Newton-Raphson or simplex search.

In order to compare the performance of the estimators, we conduct a Monte Carlo experiment based on 1000 simulation repetitions in each case. Tables 1-9 contain the results; each table corresponds to a different DGP. The leftmost columns list all the estimators and the bandwidth values for the CE(E)L. In the other columns we report the results for various performance measures, such as mean and median bias (referred to as Mean and Median in the tables), median absolute error (MAE), standard deviation (StD) and root mean square error (RMSE). In addition to these standard measures we consider the nine-decile range (9-DR), the tail probability (TailPr) and the coverage probability of a 95% confidence interval (CovPr).[4] The former provides us with information on how spread out is the distribution of the estimator between the 5th and 95th percentile. The tail probability is computed as the relative frequency of the estimates for which $\left|\widehat{\beta}\right| > 22.5$ (we follow Guggenberger (2008) in choosing this number), and it conveys information on the fatness of the tails of the distribution of the estimators. The coverage probability of the symmetric 95% confidence interval is estimated by the relative frequency of the event $\left(\left|\widehat{\beta} - \beta_0\right| \leq 1.96 \cdot \widehat{\sigma}\right)$ for a certain estimator $\widehat{\beta}$ of the true value β_0, where $\widehat{\sigma}$ is an estimator of the standard error of $\widehat{\beta}$, which may differ across the various estimators we consider.

[4]Since Guggenberger (2008) uses similar simulation setup and performance measures for studying the finite sample properties of unconditional GEL estimators, we can directly compare the tail probabilities for our estimators to his estimators.

ACCEPTED MANUSCRIPT

The focus of the Monte Carlo experiment is on the performance of the CE(E)L estimators in comparison with the instrument-based methods presented above. The latter may perform differently if few or many instruments are included, specifically, in theory many instruments lead to asymptotic efficiency gains, but in practice they may lead to biased estimates. Therefore, for the three instrumental variable-based estimators we use two instrument sets of $L = 10$ and 30 instruments. Another objective in analyzing the results is to compare CEL to CEEL. CEEL has a computational advantage compared to CEL due to the fact that the Lagrange multipliers can be expressed explicitly and need not be estimated via numerical optimization as for CEL (see equations (3) and (5)).

Before discussing the details with respect to the performance measures, we provide some general remarks. Most of the estimators suffer considerably from the presence of endogeneity. Having strong instruments may mitigate the effects of endogeneity. When instruments are weak and endogeneity is present none of the estimators performs well in terms of bias and coverage probability. Moreover, having or not heteroskedasticity seems to play a relatively minor role. Apart from the case of GMM in the absence of endogeneity, in none of the tables can we find an estimator that dominates all the others in the sense that it performs better with respect to all measures.

The HLIM estimator is often similar to CE(E)L estimators with some fixed bandwidth, espe-

cially in the weak instruments case (Tables 4-9). In the case of endogeneity the GMM estimator tends to perform well in terms of precision (MAE, StD), but performs poorly in terms of bias (Mean, Median) and coverage probability. On the other hand, when $\rho = 0$ GMM is the best estimator according to nearly all performance measures and its coverage probabilities (CovPr) are quite close to the nominal probabilities. HFUL has a rather sound performance compared to the other estimators in all the cases. The DIN estimator seems to display quite clear patterns with respect to our performance measures as we vary K. More specifically, we notice that as K grows all the measures of dispersion and the TailPr get larger. On the other hand, the coverage probabilities tend to zero. Regarding the effect of increasing the sample size from 100 to 500 we note that, contrary to our expectations, the dispersion does not decrease while the bias (both Mean and Median) gets slightly smaller. The tail probabilities, where strictly positive, tend to go down.

The two conditional empirical likelihood estimators, CEL and CEEL, have a rather similar performance.[5] Their performance is much better with automatic bandwidths than with fixed bandwidths in most of the cases. This is remarkable, because it contrasts the findings for a linear heteroskedastic model with an exogenous regressor, where CEL is only slightly better with automatic bandwidths than with fixed bandwidths (see KTA). This contrast is rather sharp in the weak

[5]This can be seen in the $n = 100$ cases. Due to this similarity we do not present results for CEL in the $n = 500$ cases.

ACCEPTED MANUSCRIPT

instruments case. We believe this is due to the fact that the cross-validation criterion tends to select lower estimates (see Appendix B.2), and in the weak instruments case the proportion of excessively large estimates, as shown by the TailPr, is larger.

In what follows we make some distinctive comments on these and the strong instruments case, and then we discuss the properties of the estimators for each performance measure.

Weak instruments case (Tables 4-9). The CE(E)L estimators with fixed bandwidths have large tail probabilities, similarly to the HLIM estimator, which is known to suffer from the no-moment problem (Hausman et al., 2012). Therefore, the CE(E)L estimators with fixed bandwidths may also have the no-moment problem in the weak instruments case. Besides the tail probabilities, these estimators perform rather poorly also with respect to the 9-DR.

The CE(E)L estimators with automatic bandwidths perform much better than their counterparts with fixed bandwidths. Their most remarkable feature is that they all have tail probabilities equal to 0 (the TailPr of CEEL for $n = 500$ and $\rho = 0$ is actually 0.001), which suggests that these estimators do not suffer from the no-moment problem. In addition, their performance with respect to the two measures of dispersion MAE and 9-DR improves dramatically, although the latter values still remain high relative to those of GMM and HFUL. The same observation holds for the StD and RMSE. Apart from the no endogeneity case, it is difficult to rank the CE(E)L and GMM even if we

ACCEPTED MANUSCRIPT

restrict the comparison to the criteria RMSE and CovPr, because in most cases GMM has lower RMSE but poorer CovPr.

Strong instruments case (Tables 1-3). The CE(E)L estimators with fixed bandwidths have small TailPr, possibly except for some low bandwidth values. Consequently, for most of the bandwidths the second moments of these estimators are finite. For the lowest bandwidth values ($b_n = 0.5$) CEL tends to perform poorly compared to HLIM in terms of the 9-DR and TailPr, so in these cases these estimators may suffer from the no-moment problem. For $n = 100$ (Tables 1-3) CE(E)L are rather competitive regarding the MAE, but poor regarding the RMSE, for several fixed bandwidth values.

The CE(E)L estimators with automatic bandwidths perform better than their counterparts with fixed bandwidths. These estimators are rather competitive compared to the other estimators as well. A clear ranking is difficult to establish even if we restrict the comparison to RMSE and CovPr, but we can claim that CEEL has rather good CovPr and low RMSE in all four cases. Compared to HFUL, CEEL has similar CovPr and lower RMSE in almost all the cases.

Mean bias. The mean bias tends to be small in absolute value when we remove endogeneity, while it seems to be indifferent to the presence of heteroskedasticity. In the weak instruments case the CE(E)L have different bias values for different bandwidths. The CE(E)L with automatic

bandwidths have a performance comparable to the other estimators. Apart from the case of $n = 500$ without endogeneity, the bias increases slightly for the GMM and HFUL estimators as the number of instruments L increases from 10 to 30, while for the HLIM the change is ambiguous. In the strong instruments case, the CE(E)L are only biased for some very low fixed bandwidth values, while the CE(E)L with automatic bandwidths are virtually unbiased. The bias of GMM and HFUL tends to decrease with the strength of instruments, and decrease with the number of observations n. The bias of GMM increases substantially as the number of instruments L increases; the bias of HLIM is small in most cases. The size of the bias of the DIN is comparable to that of the CE(E)L in the fixed bandwidth case.

Median bias. In the weak instruments case the CE(E)L estimators have similar median bias values for different bandwidths, for both fixed and automatic bandwidths. These bias values are rather similar to the median biases of the other estimators. The median bias increases slightly for the GMM, HLIM and HFUL estimators as L increases from 10 to 30. In the strong instruments case and in the case of absence of endogeneity the CE(E)L estimators tend to be median-unbiased for any choice of bandwidth. The median bias is small if we shut off endogeneity. HFUL and especially HLIM have small median bias values in most of the cases, while in the case of endogeneity GMM has considerable median bias. This bias increases with L, with the degree of endogeneity,

ACCEPTED MANUSCRIPT

and decreases with n. The median bias of the DIN estimator does not follow a specific pattern. However, in the case of endogeneity and weak instruments the size of the median bias is comparable to that of the CE(E)L estimators. Moreover, it is negative in the strong instruments case and large and positive in the case of endogeneity and weak instruments.

MAE. The MAE is a measure of dispersion that is robust to the no-moment problem. It decreases with the strength of instruments and with n, while the effect of the degree of endogeneity is ambiguous.[6] For HLIM and HFUL, MAE increases with L, while for GMM the effect of L is ambiguous. Except for very low bandwidths like $b_n = 0.5,\ 0.7$, the CE(E)L estimators with different fixed bandwidths have rather similar MAE values. In the weak instruments case these values are also similar to the MAE of HLIM and larger than the MAE of GMM and HFUL. In this case, the CE(E)L estimators with automatic bandwidths have very competitive MAE, they are comparable to those of the HFUL when $L = 10$ and they are only outperformed by GMM in the low endogeneity cases. In the weak instruments case with endogeneity (Table 4) CEL with bandwidths $b_n \in \{1.1, 1.3, 1.5, 1.7, 1.9\}$ dominates the other estimators. In most of the strong instruments cases the CE(E)L with automatic bandwidths have the lowest MAE, and they dominate HLIM and HFUL in all these cases. For $K = 2, 3$ the MAE of the DIN is comparable to that of CE(E)L with fixed

[6]The ambiguity may come from the feature of the DGP that a change in the degree of endogeneity is accompanied by a change in the degree of heteroskedasticity.

ACCEPTED MANUSCRIPT

bandwidths, while for $K = 5, 10, 15$ it tends to be larger. For those latter values of K the MAE of the DIN is often the larger value in the tables.

9-DR. The 9-DR is a measure of dispersion that can be estimated consistently for estimators that suffer from the no-moment problem. In general the performance of all the estimators with respect to the 9-DR improves with the strength of instruments, but their relative performance is specific to this feature. In all the weak instruments case GMM has the lowest 9-DR followed by HFUL, which is followed by the CE(E)L with automatic bandwidths. The CE(E)L with fixed bandwidths have rather large 9-DR values, which tend to decrease with the bandwidth. The largest values of the 9-DR are found for the DIN estimator with large K. HLIM has 9-DR values similar to those of the CE(E)L corresponding to the highest bandwidths. Compared to these, the 9-DR values of the CE(E)L with automatic bandwidths are lower by a factor ranging roughly between 2 and 3. In the strong instruments case GMM still has the lowest 9-DR in all the cases, but here this is followed by the CE(E)L with automatic bandwidths, which tends to outperform HFUL in most of the cases. The CE(E)L with some larger fixed bandwidths outperform HFUL in most of the cases, while for some lower fixed bandwidths they have 9-DR values similar to HLIM.

In general for all the estimators the 9-DR increases with the degree of heteroskedasticity. For GMM the 9-DR decreases with L, but the reverse holds for HLIM and HFUL. In the weak in-

ACCEPTED MANUSCRIPT

struments case the 9-DR of GMM, HLIM, HFUL tend to increase with n, while for the CE(E)L with automatic bandwidths it tends to decrease; for the CE(E)L with fixed bandwidths it changes ambiguously. The 9-DR of the DIN estimator tends to be large and to grow in K. In the strong instruments case, for $K = 2$, the DIN estimator is comparable to the CE(E)L estimator with automatic bandwidths. However, in the weak instruments case, its smallest value is similar to the 9-DR of the HLIM.

StD. We can repeat here the qualitative remarks made in the first paragraph of the discussion on the 9-DR. Therefore, we only mention the differences and make some further quantitative remarks. The StD still increases with the degree of heteroskedasticity in most cases, except for CE(E)L in the strong instruments case. In this case the StD of CE(E)L changes in an ambiguous way, which is most probably due to the presence of some non-zero tail probabilities. For GMM the StD still decreases with L, but the reverse only holds for HFUL, while for HLIM it does so only in the strong instruments case. In the weak instruments case the StD of HLIM changes very little and ambiguously with L. Further, in this case the StD of GMM and HFUL tend to increase with n, while for the CE(E)L with automatic bandwidths and HLIM it tends to decrease; for the CE(E)L with fixed bandwidths it changes little and ambiguously.

In the weak instruments case (Tables 4-9) the StD values of CE(E)L are improved by a factor

ACCEPTED MANUSCRIPT

ranging roughly between 2.5 and 3.5 with automatic bandwidths. In the weak instruments case the DIN with $K = 2$ is similar to the HLIM. It is interesting to note that in this case, the numerical StD values of the CE(E)L for fixed (large) bandwidths and HLIM are rather similar to the StD of the unconditional GEL and LIML estimators in Guggenberger's (2008) weak instruments case (Tables 1(a) and 1(b)).

RMSE. The RMSE values, although in some cases numerically different, qualitatively behave like the StD values. Therefore, the discussion on the performance of the estimators regarding the StD is also valid here.

CovPr. In an overall sense, the estimator with the best CovPr tends to be HLIM, which outperforms HFUL most of the times. However, we can make a few remarks. In particular, for K small the DIN estimator displays competitive CovPr's in all the cases. The HFUL estimator outperforms the CE(E)L with automatic bandwidths. In almost all cases GMM performs rather poorly in the presence of endogeneity where its CovPr is below 0.5 in several cases. When $\rho = 0$ the CovPr of the GMM is quite close to the nominal coverage. The poorest CovPr of the CE(E)L with automatic bandwidths is 0.586 (Table 6), where the CovPr of HLIM is 0.770. In the absence of endogeneity, the CovPr of the CE(E)L is very close to the nominal coverage. In the case of $n = 500$ and $\rho = 0$ the CovPr of CEEL is exactly 0.950. The CovPr of the CE(E)L with fixed bandwidths increases

ACCEPTED MANUSCRIPT

with the bandwidth values.

The CovPr improves with the strength of instruments and it gets poorer with higher endogeneity. In the weak instruments case the effect of n is not clear. The CovPr for HLIM and HFUL increases in L, while for GMM it decreases in L; the latter is remarkably poor for $L = 30$. A similar phenomenon is observed for the DIN estimator.

TailPr. The TailPr of the CE(E)L with automatic bandwidths, GMM and HFUL are 0 in nearly all the cases. The CE(E)L with fixed bandwidths and HLIM have strictly positive TailPr in several cases. In the weak instruments case these are typically rather large for the former estimator, ranging from 0.012 to 0.054, while they are slightly lower, ranging from 0.017 to 0.036 for the latter estimator.[7] In the strong instruments case, these estimators have their TailPr equal to 0 or below 0.01 in most of the cases. Some exceptions to these can be found for CEL for bandwidths $b_n = 0.5,\ 0.7$, where the TailPr values range from 0.15 to 0.20, and for HLIM with $n = 100$, where the TailPr could be as high as 0.036. The TailPr of the DIN is an increasing function of K. For $K = 2$ we find values that are comparable to the TailPr of the HLIM estimator. The TailPr's of CE(E)L, HLIM and DIN tend to go down as the sample size changes from $n = 100$ to 500.

We use the TailPr together with the fact that HLIM suffers from the no-moment problem (Haus-

[7]For comparison, we mention that the corresponding tail probability of the standard Cauchy distribution, whose first absolute moment does not exist, is about 0.028.

ACCEPTED MANUSCRIPT

man et al., 2012, p.217) as a practical indicator of the existence of moments. Our conclusions earlier in this section regarding the no-moment problem for the CE(E)L with fixed bandwidths are based on this indicator. For further comparison purposes we note that the unconditional GEL and LIML in the weak instruments case discussed by Guggenberger (2008, Tables 1(a)-(b), 3(a)-(b)) have tail probabilities ranging from 0.1 to 0.3. These values are rather close to those found in our weak instruments case for HLIM and slightly lower than those found for the CE(E)L with fixed bandwidths.

4 Conclusions

In this paper we find evidence that the CE(E)L estimators with certain fixed bandwidths have standard deviations and tail probabilities similar to the HLIM estimator, which is known to have the no-moment problem. This suggests that the CE(E)L with fixed bandwidths may also suffer from the no-moment problem. We also study these estimators with automatic bandwidths obtained through the cross-validation method proposed by Newey (1993). Our results suggest that the CE(E)L estimators with automatic bandwidths do not have the no-moment problem. This is remarkable for two reasons. First, the closely related unconditional GEL estimators also suffer from the no-moment problem (Guggenberger, 2008). Second, in linear heteroskedastic models without

endogenous regressors the CE(E)L with fixed and automatic bandwidths have similar finite sample properties (KTA and Gospodinov and Otsu, 2012). We provide arguments that the cross-validation criterion in the model we consider tends to eliminate estimates with large values, which implies that estimators with automatic bandwidths have low tail probabilities.

In linear models with endogenous regressors and weak instruments we find CE(E)L to have finite sample properties poorer than the HFUL estimator. This holds regardless of whether the bandwidth is fixed or automatic, although the latter considerably improves the performance of CE(E)L under the various performance measures. The relative performances change significantly in the strong instruments case. Automatic bandwidths for CE(E)L still improve over fixed bandwidths in most cases, but the improvement is not as large as in the weak instruments case. Further, the CE(E)L with automatic bandwidths tend to outperform HFUL in terms of RMSE, while the reverse holds in terms of the coverage probability, although the differences in performance are numerically rather small.

Based on these considerations, we recommend the use of HFUL. This advice also takes into account the computational burden that CEEL, and in particular CEL, entail, which increases further when the automatic bandwidth is calculated. Still, in cases when the RMSE is the relevant loss function, and the instruments are known to be strong, one may prefer CE(E)L. In this situ-

ation, since CEL and CEEL deliver similar results, we recommend the computationally simpler CEEL. Since even in the strong instruments case it may happen for some fixed bandwidths that the CEEL estimator has a large tail probability, we recommend estimation by using at least a few fixed bandwidths followed by the selection of the best bandwidth.

The conclusions regarding the relative performance of the CE(E)L estimators may be different in nonlinear models. In such models, since the HFUL estimator has been developed for linear models, the performance of CE(E)L should be compared to other estimators, which are suited to nonlinear models. Such estimators have recently been developed by Domínguez and Lobato (2004) and Lavergne and Patilea (2012) based on unconditional moment restrictions that are equivalent to the conditional moment restriction that identifies the model. Future research will focus on the finite samples properties of CE(E)L compared to these estimators, as well as to the efficient GMM estimator (Newey, 1993) for a nonlinear model.

ACCEPTED MANUSCRIPT

A Appendix: Tables

A.1 $n = 100$

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.4, \rho = 0.75, R^2 = 0.1$							
CEL								
automatic	0.018	0.080	0.210	1.300	0.472	0.472	0.877	0.000
$b_n = 0.5$	-0.605	-0.020	0.350	4.040	4.026	4.071	0.732	0.020
$b_n = 0.7$	-0.485	-0.010	0.320	3.200	3.484	3.518	0.798	0.015
$b_n = 0.9$	-0.218	-0.010	0.240	1.680	1.978	1.990	0.852	0.003
$b_n = 1.1$	-0.172	-0.010	0.240	1.460	1.267	1.279	0.853	0.002
$b_n = 1.3$	-0.143	-0.010	0.240	1.440	1.491	1.498	0.916	0.003
$b_n = 1.5$	-0.140	-0.010	0.230	1.440	1.497	1.503	0.905	0.003
$b_n = 1.7$	-0.088	-0.010	0.240	1.470	1.857	1.859	0.948	0.005
$b_n = 1.9$	-0.037	0.000	0.250	1.530	1.502	1.503	0.959	0.002
CEEL								
automatic	0.025	0.060	0.220	1.170	0.450	0.451	0.891	0.000
$b_n = 0.5$	-0.333	-0.040	0.290	2.553	2.347	2.370	0.716	0.004
$b_n = 0.7$	-0.209	-0.040	0.280	2.112	2.039	2.050	0.797	0.004
$b_n = 0.9$	-0.109	-0.010	0.235	1.560	1.018	1.023	0.882	0.000
$b_n = 1.1$	-0.073	-0.010	0.230	1.391	0.793	0.797	0.915	0.000
$b_n = 1.3$	-0.085	-0.010	0.230	1.381	0.585	0.591	0.926	0.000
$b_n = 1.5$	-0.055	-0.010	0.240	1.321	0.965	0.967	0.931	0.001
$b_n = 1.7$	-0.080	-0.020	0.240	1.361	0.682	0.687	0.936	0.000
$b_n = 1.9$	-0.035	-0.010	0.240	1.431	1.239	1.239	0.941	0.001
GMM								
$L = 10$	0.295	0.311	0.144	0.779	0.244	0.383	0.633	0.000
$L = 30$	0.532	0.533	0.091	0.470	0.144	0.551	0.073	0.000
HLIM								
$L = 10$	-0.108	0.026	0.266	1.910	1.483	1.487	0.912	0.001
$L = 30$	-0.109	0.068	0.375	4.222	3.436	3.437	0.921	0.011
HFUL								
$L = 10$	0.067	0.093	0.213	1.136	0.372	0.378	0.892	0.000
$L = 30$	0.202	0.181	0.252	1.449	0.441	0.485	0.907	0.000
DIN								
$K = 2$	-0.133	-0.004	0.250	1.350	1.206	1.213	0.910	0.002
$K = 3$	-0.230	-0.029	0.261	1.547	1.828	1.842	0.870	0.004
$K = 5$	-0.621	-0.134	0.475	14.870	5.553	5.587	0.478	0.024
$K = 10$	-0.716	-0.295	1.159	23.460	7.269	7.305	0.187	0.040
$K = 15$	-0.081	-0.259	1.273	24.262	7.284	7.284	0.141	0.040

Table 1: strong instruments, endogeneity, heteroskedasticity

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.4, \rho = 0, R^2 = 0.1$							
CEL								
automatic	-0.013	0.005	0.225	1.270	0.492	0.492	0.975	0.000
$b_n = 0.5$	-0.340	0.000	0.380	3.930	3.643	3.659	0.594	0.017
$b_n = 0.7$	-0.167	0.010	0.330	3.080	2.989	2.993	0.647	0.009
$b_n = 0.9$	0.035	0.010	0.260	1.670	1.675	1.676	0.693	0.002
$b_n = 1.1$	-0.025	0.000	0.240	1.540	1.304	1.304	0.722	0.002
$b_n = 1.3$	-0.016	0.010	0.250	1.540	1.333	1.333	0.751	0.002
$b_n = 1.5$	-0.050	0.000	0.260	1.510	1.634	1.635	0.774	0.003
$b_n = 1.7$	-0.013	0.000	0.270	1.590	1.613	1.613	0.796	0.002
$b_n = 1.9$	-0.033	0.000	0.280	1.740	1.685	1.686	0.810	0.003
CEEL								
automatic	-0.002	-0.020	0.200	1.300	0.422	0.422	0.982	0.000
$b_n = 0.5$	-0.009	-0.020	0.310	2.612	2.145	2.145	0.682	0.004
$b_n = 0.7$	-0.042	-0.020	0.280	2.072	1.499	1.500	0.728	0.002
$b_n = 0.9$	-0.020	-0.020	0.250	1.613	0.859	0.859	0.775	0.000
$b_n = 1.1$	-0.034	-0.010	0.240	1.491	0.978	0.979	0.798	0.001
$b_n = 1.3$	0.039	-0.010	0.230	1.482	1.098	1.099	0.834	0.001
$b_n = 1.5$	0.011	-0.020	0.230	1.581	1.314	1.314	0.850	0.002
$b_n = 1.7$	-0.005	-0.010	0.250	1.622	1.010	1.010	0.866	0.000
$b_n = 1.9$	-0.021	-0.020	0.260	1.791	0.783	0.783	0.884	0.000
GMM								
$L = 10$	-0.010	-0.007	0.170	0.904	0.279	0.279	0.936	0.000
$L = 30$	0.002	0.000	0.119	0.609	0.188	0.188	0.925	0.000
HLIM								
$L = 10$	0.007	-0.004	0.289	2.279	1.836	1.836	0.972	0.002
$L = 30$	-0.002	-0.004	0.479	5.343	3.592	3.592	0.991	0.011
HFUL								
$L = 10$	-0.006	-0.002	0.251	1.554	0.453	0.453	0.966	0.000
$L = 30$	0.011	-0.002	0.380	2.103	0.621	0.621	0.988	0.000
DIN								
$K = 2$	0.007	-0.015	0.242	1.340	0.961	0.961	0.947	0.000
$K = 3$	0.093	-0.009	0.249	1.739	2.166	2.168	0.870	0.006
$K = 5$	-0.073	-0.044	0.561	15.143	5.748	5.748	0.411	0.030
$K = 10$	0.033	-0.030	1.589	24.647	7.347	7.347	0.156	0.044
$K = 15$	0.515	-0.001	2.154	28.694	8.301	8.317	0.107	0.061

Table 2: strong instruments, no endogeneity, heteroskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.4, \rho = 0.75, R^2 = 0$							
CEL								
automatic	0.082	0.140	0.190	1.130	0.382	0.390	0.869	0.000
$b_n = 0.5$	-0.660	0.020	0.270	3.100	3.728	3.786	0.656	0.019
$b_n = 0.7$	-0.449	0.030	0.240	1.880	3.259	3.290	0.700	0.016
$b_n = 0.9$	-0.121	0.010	0.210	1.360	1.053	1.060	0.735	0.001
$b_n = 1.1$	-0.133	0.005	0.215	1.250	1.212	1.219	0.769	0.002
$b_n = 1.3$	-0.087	0.020	0.210	1.250	1.109	1.112	0.807	0.001
$b_n = 1.5$	-0.036	0.020	0.220	1.330	1.278	1.279	0.829	0.002
$b_n = 1.7$	-0.085	0.020	0.220	1.480	1.416	1.418	0.851	0.002
$b_n = 1.9$	-0.137	0.020	0.240	1.640	1.697	1.703	0.875	0.001
CEEL								
automatic	0.047	0.100	0.190	1.070	0.411	0.414	0.899	0.000
$b_n = 0.5$	-0.138	0.010	0.240	1.840	1.753	1.759	0.719	0.003
$b_n = 0.7$	-0.185	0.020	0.220	1.490	1.505	1.516	0.757	0.003
$b_n = 0.9$	-0.131	0.000	0.210	1.290	0.995	1.004	0.801	0.001
$b_n = 1.1$	-0.093	0.000	0.200	1.330	1.014	1.018	0.854	0.001
$b_n = 1.3$	-0.113	0.010	0.200	1.440	0.764	0.772	0.880	0.000
$b_n = 1.5$	-0.114	0.010	0.200	1.470	0.897	0.904	0.894	0.000
$b_n = 1.7$	-0.056	0.010	0.210	1.440	1.736	1.737	0.910	0.004
$b_n = 1.9$	-0.018	0.010	0.220	1.550	1.204	1.204	0.927	0.001
GMM								
$L = 10$	0.298	0.311	0.122	0.654	0.201	0.359	0.498	0.000
$L = 30$	0.520	0.519	0.083	0.427	0.131	0.536	0.023	0.000
HLIM								
$L = 10$	-0.115	0.016	0.218	1.713	1.532	1.536	0.936	0.002
$L = 30$	0.002	0.090	0.377	4.270	3.089	3.089	0.902	0.009
HFUL								
$L = 10$	0.058	0.097	0.173	0.963	0.305	0.311	0.911	0.000
$L = 30$	0.247	0.205	0.269	1.658	0.487	0.546	0.867	0.000
DIN								
$K = 2$	-0.105	-0.001	0.194	1.152	0.905	0.911	0.916	0.001
$K = 3$	-0.118	-0.007	0.211	1.248	0.999	1.005	0.879	0.001
$K = 5$	-0.821	-0.108	0.352	11.449	4.617	4.690	0.515	0.014
$K = 10$	-0.599	-0.228	0.767	21.824	6.595	6.623	0.218	0.032
$K = 15$	-0.608	-0.224	0.870	23.411	6.878	6.905	0.174	0.031

Table 3: strong instruments, endogeneity, homoskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.04, \rho = 0.75, R^2 = 0.1$							
CEL								
automatic	0.698	0.730	0.400	3.640	1.514	1.667	0.615	0.000
$b_n = 0.5$	0.236	0.740	1.090	16.270	6.183	6.188	0.367	0.039
$b_n = 0.7$	0.648	0.730	1.050	14.880	5.736	5.772	0.459	0.025
$b_n = 0.9$	0.656	0.730	0.865	13.140	5.666	5.704	0.561	0.030
$b_n = 1.1$	0.636	0.740	0.720	10.870	5.632	5.668	0.687	0.036
$b_n = 1.3$	0.674	0.740	0.650	11.410	5.387	5.429	0.723	0.028
$b_n = 1.5$	0.666	0.740	0.615	10.140	5.262	5.304	0.797	0.025
$b_n = 1.7$	0.703	0.740	0.590	8.980	4.878	4.929	0.850	0.022
$b_n = 1.9$	0.564	0.740	0.570	9.130	4.721	4.755	0.855	0.019
CEEL								
automatic	0.656	0.700	0.440	3.650	1.318	1.472	0.656	0.000
$b_n = 0.5$	0.528	0.710	1.040	14.490	5.683	5.708	0.360	0.029
$b_n = 0.7$	0.626	0.740	1.050	13.808	5.965	5.997	0.497	0.037
$b_n = 0.9$	0.736	0.700	0.860	12.498	5.762	5.809	0.589	0.033
$b_n = 1.1$	0.514	0.695	0.745	10.802	5.490	5.514	0.663	0.031
$b_n = 1.3$	0.460	0.680	0.670	10.330	5.489	5.508	0.737	0.034
$b_n = 1.5$	0.408	0.680	0.645	9.837	4.861	4.863	0.770	0.023
$b_n = 1.7$	0.557	0.695	0.920	8.847	4.393	4.427	0.798	0.016
$b_n = 1.9$	0.709	0.700	0.610	8.105	4.713	4.765	0.819	0.022
GMM								
$L = 10$	0.728	0.735	0.180	1.026	0.326	0.798	0.294	0.000
$L = 30$	0.752	0.751	0.092	0.481	0.144	0.766	0.010	0.000
HLIM								
$L = 10$	0.506	0.730	0.568	9.782	4.702	4.729	0.718	0.020
$L = 30$	0.564	0.751	0.619	8.534	4.908	4.940	0.822	0.025
HFUL								
$L = 10$	0.726	0.739	0.346	1.720	0.512	0.888	0.649	0.000
$L = 30$	0.753	0.746	0.387	1.783	0.542	0.927	0.794	0.000
DIN								
$K = 2$	0.590	0.684	0.845	9.798	4.682	4.719	0.808	0.017
$K = 3$	0.260	0.685	1.144	16.567	6.144	6.144	0.678	0.034
$K = 5$	-0.273	0.657	1.885	27.869	7.962	7.966	0.331	0.062
$K = 10$	-0.374	0.616	2.658	38.041	9.289	9.296	0.085	0.090
$K = 15$	-0.610	0.488	2.829	36.804	9.166	9.187	0.057	0.087

Table 4: weak instruments, endogeneity, heteroskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.04, \rho = 0, R^2 = 0.1$							
CEL								
automatic	0.047	0.030	0.560	5.010	1.725	1.726	0.943	0.000
$b_n = 0.5$	-0.425	0.020	1.330	20.540	7.165	7.177	0.587	0.054
$b_n = 0.7$	0.079	0.050	1.350	19.040	6.908	6.908	0.720	0.046
$b_n = 0.9$	0.123	0.050	1.155	17.080	6.583	6.585	0.730	0.047
$b_n = 1.1$	0.078	0.040	1.015	16.200	6.396	6.396	0.730	0.045
$b_n = 1.3$	0.166	0.025	0.985	15.810	6.289	6.291	0.735	0.041
$b_n = 1.5$	0.092	0.010	0.940	12.760	5.668	5.669	0.730	0.031
$b_n = 1.7$	-0.007	0.005	0.915	12.190	5.545	5.545	0.738	0.034
$b_n = 1.9$	-0.039	0.010	0.880	10.980	5.268	5.268	0.748	0.026
CEEL								
automatic	0.085	0.020	0.600	4.850	1.840	1.842	0.959	0.000
$b_n = 0.5$	-0.340	-0.025	1.220	19.450	6.449	6.458	0.602	0.034
$b_n = 0.7$	0.010	0.055	1.200	16.080	6.141	6.141	0.731	0.034
$b_n = 0.9$	0.197	0.010	1.090	14.170	6.237	6.240	0.774	0.042
$b_n = 1.1$	0.066	0.005	1.025	15.320	6.196	6.196	0.773	0.037
$b_n = 1.3$	-0.046	0.010	0.970	16.160	6.079	6.079	0.772	0.031
$b_n = 1.5$	0.028	0.010	0.980	14.820	6.078	6.078	0.771	0.036
$b_n = 1.7$	0.043	0.020	0.940	13.200	5.750	5.751	0.770	0.032
$b_n = 1.9$	-0.030	0.010	0.920	13.120	5.429	5.429	0.776	0.029
GMM								
$L = 10$	-0.003	-0.003	0.268	1.411	0.443	0.443	0.952	0.000
$L = 30$	0.007	-0.004	0.138	0.738	0.221	0.221	0.924	0.000
HLIM								
$L = 10$	-0.231	-0.009	0.904	13.285	5.815	5.819	0.990	0.036
$L = 30$	0.031	0.033	0.992	11.619	5.629	5.629	0.993	0.030
HFUL								
$L = 10$	0.007	-0.005	0.522	2.434	0.726	0.726	0.983	0.000
$L = 30$	0.017	0.019	0.603	2.755	0.839	0.840	0.993	0.000
DIN								
$K = 2$	-0.015	0.006	1.069	14.278	5.785	5.785	0.994	0.030
$K = 3$	0.300	0.052	1.345	17.393	6.462	6.469	0.935	0.036
$K = 5$	-0.181	-0.097	2.063	33.875	8.627	8.681	0.413	0.078
$K = 10$	-0.289	0.059	2.933	36.961	9.326	9.330	0.089	0.090
$K = 15$	-0.167	0.082	2.825	37.129	9.426	9.428	0.061	0.091

Table 5: weak instruments, no endogeneity, heteroskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 100$	$\pi = 0.04, \rho = 0.75, R^2 = 0$							
CEL								
automatic	0.741	0.750	0.340	2.890	0.994	1.240	0.586	0.000
$b_n = 0.5$	0.561	0.740	0.670	9.550	4.631	4.665	0.368	0.020
$b_n = 0.7$	0.552	0.730	0.650	9.880	4.818	4.850	0.444	0.022
$b_n = 0.9$	0.675	0.710	0.670	9.610	4.554	4.604	0.556	0.019
$b_n = 1.1$	0.586	0.700	0.680	8.490	4.370	4.410	0.614	0.017
$b_n = 1.3$	0.848	0.700	0.700	7.940	4.320	4.402	0.659	0.018
$b_n = 1.5$	0.853	0.720	0.700	7.850	4.108	4.196	0.688	0.014
$b_n = 1.7$	0.783	0.720	0.710	8.210	4.138	4.211	0.716	0.014
$b_n = 1.9$	0.931	0.735	0.685	7.950	3.934	4.043	0.739	0.013
CEEL								
automatic	0.748	0.745	0.355	3.020	1.372	1.562	0.628	0.000
$b_n = 0.5$	0.595	0.670	0.660	11.120	4.886	4.922	0.402	0.019
$b_n = 0.7$	0.572	0.700	0.660	9.550	4.555	4.591	0.452	0.018
$b_n = 0.9$	0.877	0.720	0.700	9.110	4.685	4.766	0.548	0.019
$b_n = 1.1$	0.661	0.730	0.700	8.650	4.960	5.004	0.627	0.027
$b_n = 1.3$	0.672	0.720	0.675	7.830	4.519	4.569	0.682	0.022
$b_n = 1.5$	0.470	0.710	0.665	8.790	4.478	4.502	0.721	0.019
$b_n = 1.7$	0.629	0.720	0.650	8.060	4.523	4.566	0.732	0.018
$b_n = 1.9$	0.558	0.705	0.640	8.730	4.282	4.319	0.744	0.013
GMM								
$L = 10$	0.746	0.751	0.157	0.772	0.246	0.786	0.136	0.000
$L = 30$	0.754	0.757	0.095	0.452	0.142	0.767	0.002	0.000
HLIM								
$L = 10$	0.823	0.771	0.672	9.741	4.694	4.766	0.770	0.022
$L = 30$	0.643	0.778	0.696	8.577	4.925	4.966	0.763	0.023
HFUL								
$L = 10$	0.757	0.765	0.351	1.549	0.472	0.892	0.671	0.000
$L = 30$	0.755	0.769	0.407	1.814	0.554	0.937	0.703	0.000
DIN								
$K = 2$	0.308	0.679	0.674	8.520	4.779	4.789	0.776	0.022
$K = 3$	0.277	0.715	0.680	12.451	5.354	5.361	0.618	0.028
$K = 5$	0.364	0.736	0.973	17.669	6.337	6.348	0.318	0.040
$K = 10$	0.325	0.743	1.621	23.085	7.457	7.464	0.123	0.054
$K = 15$	0.319	0.744	1.718	23.088	7.604	7.611	0.085	0.085

Table 6: weak instruments, endogeneity, homoskedasticity

ACCEPTED MANUSCRIPT

A.2 $n = 500$

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 500$	$\pi = 0.04, \rho = 0.75, R^2 = 0.1$							
CEEL								
automatic	0.538	0.640	0.420	3.760	1.510	1.603	0.669	0.000
$b_n = 0.5$	0.495	0.540	0.790	11.100	4.692	4.718	0.472	0.016
$b_n = 0.7$	0.081	0.520	0.720	10.460	5.238	5.238	0.554	0.028
$b_n = 0.9$	0.517	0.550	0.690	9.780	5.255	5.280	0.612	0.029
$b_n = 1.1$	0.423	0.545	0.695	10.240	4.707	4.726	0.653	0.017
$b_n = 1.3$	0.306	0.540	0.665	8.650	4.645	4.655	0.683	0.018
$b_n = 1.5$	0.369	0.540	0.650	8.320	4.691	4.706	0.702	0.018
$b_n = 1.7$	0.350	0.540	0.630	8.340	4.568	4.581	0.720	0.019
$b_n = 1.9$	0.521	0.560	0.640	8.250	4.523	4.553	0.725	0.020
GMM								
$L = 10$	0.704	0.701	0.196	1.119	0.335	0.780	0.392	0.000
$L = 30$	0.734	0.735	0.113	0.568	0.174	0.755	0.029	0.000
HLIM								
$L = 10$	0.651	0.674	0.595	9.263	4.737	4.782	0.716	0.022
$L = 30$	0.464	0.685	0.681	8.590	4.810	4.832	0.728	0.020
HFUL								
$L = 10$	0.671	0.691	0.396	2.275	0.666	0.945	0.674	0.000
$L = 30$	0.682	0.694	0.500	2.611	0.778	1.035	0.704	0.000
DIN								
$K = 2$	-0.044	0.431	0.731	9.475	5.108	5.108	0.842	0.026
$K = 3$	-0.176	0.426	0.842	15.900	5.850	5.852	0.743	0.033
$K = 5$	0.321	0.453	1.152	16.021	6.254	6.262	0.568	0.038
$K = 10$	0.026	0.492	1.293	18.864	6.640	6.641	0.318	0.040
$K = 15$	-0.160	0.544	1.334	20.472	7.044	7.045	0.209	0.052

Table 7: weak instruments, endogeneity, heteroskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 500$	$\pi = 0.04, \rho = 0, R^2 = 0.1$							
CEEL								
automatic	0.005	0.010	0.555	4.780	2.006	2.006	0.950	0.001
$b_n = 0.5$	0.135	0.040	0.960	13.580	5.661	5.662	0.738	0.026
$b_n = 0.7$	-0.014	0.020	0.880	10.990	5.264	5.264	0.806	0.025
$b_n = 0.9$	-0.202	0.015	0.875	10.240	5.141	5.145	0.813	0.027
$b_n = 1.1$	-0.004	0.045	0.875	9.900	5.052	5.052	0.821	0.024
$b_n = 1.3$	-0.115	0.020	0.860	9.040	4.693	4.694	0.821	0.021
$b_n = 1.5$	-0.020	0.030	0.870	9.320	4.923	4.923	0.830	0.022
$b_n = 1.7$	0.074	0.020	0.895	9.930	4.950	4.950	0.831	0.020
$b_n = 1.9$	0.061	0.015	0.905	10.510	5.229	5.229	0.839	0.022
GMM								
$L = 10$	0.006	0.014	0.260	1.389	0.433	0.433	0.977	0.000
$L = 30$	0.002	0.010	0.153	0.762	0.231	0.285	0.958	0.000
HLIM								
$L = 10$	0.140	0.018	0.776	9.969	5.373	5.375	0.991	0.029
$L = 30$	-0.048	-0.022	0.913	11.310	5.109	5.109	0.992	0.022
HFUL								
$L = 10$	0.010	0.013	0.525	2.919	0.852	0.852	0.989	0.000
$L = 30$	0.002	-0.019	0.691	3.524	1.028	1.454	0.990	0.000
DIN								
$K = 2$	0.133	0.024	0.837	9.502	5.252	5.253	0.994	0.029
$K = 3$	0.286	0.043	0.980	12.065	5.959	5.966	0.972	0.039
$K = 5$	0.202	0.013	1.254	16.767	5.993	5.996	0.758	0.026
$K = 10$	-0.017	0.033	1.617	21.157	7.344	7.344	0.356	0.055
$K = 15$	0.029	0.026	1.579	21.124	7.261	7.261	0.196	0.050

Table 8: weak instruments, no endogeneity, heteroskedasticity

ACCEPTED MANUSCRIPT

	Mean	Median	MAE	9-DR	StD	RMSE	CovPr	TailPr
$n = 500$	$\pi = 0.04, \rho = 0.75, R^2 = 0$							
CEEL								
automatic	0.595	0.640	0.380	3.690	1.479	1.594	0.652	0.000
$b_n = 0.5$	0.617	0.520	0.650	9.190	4.981	5.019	0.496	0.025
$b_n = 0.7$	0.453	0.490	0.625	8.140	4.343	4.366	0.565	0.016
$b_n = 0.9$	0.367	0.480	0.610	8.150	4.244	4.260	0.617	0.016
$b_n = 1.1$	0.524	0.480	0.600	7.660	3.940	3.975	0.662	0.012
$b_n = 1.3$	0.362	0.490	0.590	7.980	4.030	4.046	0.712	0.015
$b_n = 1.5$	0.433	0.505	0.615	7.740	4.379	4.400	0.747	0.018
$b_n = 1.7$	0.589	0.510	0.625	8.650	4.679	4.716	0.758	0.021
$b_n = 1.9$	0.695	0.530	0.645	7.940	4.740	4.790	0.778	0.023
GMM								
$L = 10$	0.698	0.701	0.160	0.756	0.233	0.735	0.155	0.000
$L = 30$	0.735	0.737	0.088	0.434	0.132	0.746	0.004	0.000
HLIM								
$L = 10$	0.638	0.584	0.645	7.909	4.308	4.355	0.813	0.017
$L = 30$	0.656	0.672	0.624	7.847	4.340	4.397	0.764	0.020
HFUL								
$L = 10$	0.649	0.619	0.387	1.710	0.530	0.838	0.734	0.000
$L = 30$	0.706	0.689	0.439	2.118	0.639	0.952	0.731	0.000
DIN								
$K = 2$	0.297	0.415	0.585	9.148	4.111	4.121	0.813	0.010
$K = 3$	0.269	0.442	0.586	9.125	4.569	4.577	0.710	0.020
$K = 5$	0.386	0.493	0.677	10.158	4.991	5.006	0.530	0.024
$K = 10$	0.414	0.606	0.829	10.500	4.799	4.817	0.323	0.019
$K = 15$	0.348	0.595	0.865	9.928	4.843	4.856	0.249	0.020

Table 9: weak instruments, endogeneity, homoskedasticity

B Appendix: Notes on computation

B.1 Lagrange multipliers for CEL

The Lagrange multiplier $\lambda(z_i, \beta)$ is the solution, for any $i = 1, ..., n$, of the maximization problem

$$\lambda(z_i, \beta) = \arg\max_{\gamma} \sum_{j=1}^{n} w_{ij} \log\left(1 + \gamma g\left(y_j, x_j, \beta\right)\right).$$

For simplicity of notation drop the subscript i from w_{ij} and let $g_j = g\left(y_j, x_j, \beta\right)$. Then, the Lagrange multiplier corresponding to this generic case is found by maximizing

$$f(\gamma) = \sum_{j=1}^{n} w_j \log\left(1 + g_j \gamma\right).$$

This is a function strictly concave in γ unless $g_j = 0$ for all j.

In order to search for γ values for which $1+g_j\gamma > 0$ for all j, we compute $c = \max\left\{\frac{-1}{g_j} | g_j > 0, w_j > 0\right\} < 0$ and $d = \min\left\{\frac{-1}{g_j} | g_j < 0, w_j > 0\right\} > 0$;[8] then for $c < \gamma < d$ it holds that $1 + g_j\gamma > 0$ for all j. We use the Newton-Raphson algorithm to find the Lagrange multiplier. In order to ensure that the algorithm does not take γ values outside the interval (c, d), we maximize in fact the function

$$F(t) = \sum_{j=1}^{n} w_j \log\left(1 + g_j \frac{c + de^t}{1 + e^t}\right), \quad t \in \mathbb{R};$$

suppose we obtain $t^* = \arg\max_t F(t)$. Then, the Lagrange multiplier is determined as

$$\lambda = \frac{c + de^{t^*}}{1 + e^{t^*}} \in (c, d).$$

This method for computing the Lagrange multipliers $\lambda(z_i, \beta)$ has worked very well for our DGP's.

[8]Note that, since we use the Epanechnikov kernel, not all weights w_j are necessarily strictly positive.

ACCEPTED MANUSCRIPT

B.2 Cross-validation

The cross-validation criterion proposed by Newey (1993, p.433) adapted to our model is

$$\widehat{CV}(b_n) = \sum_{i=1}^{n} \widehat{R}^2(z_i)\,\Omega(z_i),$$

where

$$\widehat{R}(z_i) = \left\{\widehat{D}(z_i) - D(z_i) + B(z_i)\left[\widehat{\Omega}(z_i) - \Omega(z_i)\right]\right\},$$

$B(z) = D(z)\,\Omega(z)^{-1}$ and $\widehat{D}(z)$, $\widehat{\Omega}(z)$ are nonparametric kernel regression estimators of $D(z)$, $\Omega(z)$, respectively, where

$$D(z) = E\left[\frac{\partial g(y, x, \beta_0)}{\partial \beta}|z\right], \quad \Omega(z) = E\left[g^2(y, x, \beta_0)\,|z\right].^{9} \tag{7}$$

The expression $\widehat{R}(z_i)$ cannot be computed; Newey proposes to estimate it by

$$\overline{R}(z_i) = \left[\frac{\partial g\left(y_i, x_i, \widehat{\beta}\right)}{\partial \beta} - \widehat{D}_{-i} + \widehat{B}_{-i}\left(\widehat{g}_i^2 - \widehat{\Omega}_{-i}\right)\right],$$

[9] For our DGP these expressions are

$$\begin{aligned} D(z) &= E[-x|z] = -\pi z, \\ \Omega(z) &= E\left[\left(\rho u + \sqrt{\frac{1-\rho^2}{\phi^2 + .86^4}}\,(\phi v_1 + .86 v_2)\right)^2 |z\right] = \rho^2 + \left(1-\rho^2\right)\frac{\phi^2 z^2 + .86^4}{\phi^2 + .86^4}. \end{aligned}$$

ACCEPTED MANUSCRIPT

where $\widehat{D}_{-i}$ and $\widehat{\Omega}_{-i}$ are leave-one-out estimators of $D(z_i)$ and $\Omega(z_i)$. Specifically,

$$\begin{aligned}
\widehat{D}_{-i} &= -\sum_{j=1,\ j\neq i}^{n} x_j w_{-ij}, \\
\widehat{\Omega}_{-i} &= \sum_{j=1,\ j\neq i}^{n} \widehat{g}_j^2 w_{-ij}, \\
\widehat{B}_{-i} &= \widehat{D}_{-i}\widehat{\Omega}_{-i}^{-1},
\end{aligned}$$

where

$$\begin{aligned}
w_{-ij} &= \frac{K\left(\frac{z_i - z_j}{b_n}\right)}{\sum_{j=1,\ j\neq i}^{n} K\left(\frac{z_i - z_j}{b_n}\right)}, \\
\widehat{g}_i &= g\left(y_i, x_i, \widehat{\beta}\right).
\end{aligned}$$

The basic idea underlying this estimation is to replace the conditional expectations by their leave-one-out estimators and the estimators of the conditional expectations by the dependent variables in the associated nonparametric regression. In our model the criterion simplifies to

$$\overline{R}(z_i) = \left[\widehat{D}_{-i}\left(\frac{\widehat{g}_i^2}{\widehat{\Omega}_{-i}} - 1\right) - x_i - \widehat{D}_{-i}\right].$$

The cross-validation criterion we use is

$$\overline{CV}(b_n) = \sum_{i=1}^{n} \overline{R}^2(z_i)\,\widehat{\Omega}_{-i}.$$

For all values of bandwidths b_n from a grid (e.g., 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9) we obtain an estimator $\widehat{\beta} \equiv \widehat{\beta}(b_n)$. Then we compute the values $\overline{CV}(b_n)$ for each b_n and choose the estimator and the bandwidth that minimize $\overline{CV}(b_n)$. We refer to the estimator that we obtain this way as the cross-validated estimator, and to the bandwidth that we obtain as the automatic bandwidth. In the remainder of this section we show that in the case of our model for large values of the estimate the cross-validation criterion is also large, that is, when $\left|\widehat{\beta}\right| \to \infty$ we have that $\overline{CV}(b_n) \to \infty$ First note that when $\left|\widehat{\beta}\right| \to \infty$ we have that

$$\overline{R}(z_i) \to \widehat{D}_{-i}\left(\frac{x_i^2}{\sum_{j=1,\ j\neq i}^{n} x_j^2 w_{-ij}} - 1\right) - x_i - \widehat{D}_{-i}.$$

Since we assume that the variables in our model are continuous random variables, it holds with probability 1 that there are numbers h, k such that $\widehat{L}_h \equiv \widehat{D}_{-h}\left(\frac{x_h^2}{\sum_{j=1,\ j\neq h}^{n} x_j^2 w_{-hj}} - 1\right) - x_h - \widehat{D}_{-h} \neq 0$ and $x_k w_{-hk} \neq 0$. Then

$$\overline{CV}(b_n) = \sum_{i=1}^{n} \overline{R}^2(z_i)\,\widehat{\Omega}_{-i} \geq \overline{R}^2(z_h)\,\widehat{\Omega}_{-h} \geq \overline{R}^2(z_h)\,\widehat{g}_k^2 w_{-hk}. \tag{8}$$

Since $x_k \neq 0$ we have that $\widehat{g}_k^2 = \left(y_k - \widehat{\beta} x_k\right)^2 \to \infty$ as $\left|\widehat{\beta}\right| \to \infty$, so the limit of the right hand side of (8) when $\left|\widehat{\beta}\right| \to \infty$ is ∞.

References

[1] Antoine, B., H. Bonnal, E. Renault (2007): On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood, *Journal of Econometrics*, **138**, 461-487.

[2] Domínguez, M. A., I. N. Lobato (2004): Consistent Estimation of Models Defined by Conditional Moment Restrictions, *Econometrica*, **72**, 1601-1615.

[3] Donald, S. G., G. W. Imbens, W. K. Newey (2003): Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions, *Journal of Econometrics*, **117**, 55-93.

[4] Fiebig, D. G. (1985): Evaluating Estimators without Moments, *The Review of Economics and Statistics*, **67**, 529-534.

[5] Fuller, W. A. (1977): Some Properties of a Modification of the Limited Information Estimator, *Econometrica*, **45**, 939-953.

[6] Gospodinov, N. and T. Otsu (2012): Local GMM Estimation of Time Series Models with Conditional Moment Restrictions, *Journal of Econometrics.*, **120**, 938-550.

ACCEPTED MANUSCRIPT

[7] Guggenberger, P. (2008): Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator, *Econometric Reviews*, **27**, 526-541.

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

[9] Hausman, J., R. Lewis, K. Menzel, W. Newey (2011): Properties of the CUE Estimator and a Modification with Moments, *Journal of Econometrics.*, **165**, 45-57.

[10] Hausman, J. A., W. K. Newey, T. Woutersen, J. Chao, N. Swanson (2012): Instrumental Variable Estimation with Heteroskedasticity and Many Instruments, *Quantitative Economics*, **3**, 211-255.

[11] Kitamura, Y., G. Tripathi, H. Ahn (2004): Empirical Likelihood-Based Inference in Conditional Moment Restriction Models, *Econometrica*, **72**, 1667-1714.

[12] Lavergne, P. and V. Patilea (2009): Smooth Minimum Distance Estimation and Testing in Conditional Moment Restrictions Models: Uniform in Bandwidth Theory, *Journal of Econometrics.*, **177**, 47-59.

[13] Morimune, K. (1983): Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size, *Econometrica.*, **51**, 821-841.

ACCEPTED MANUSCRIPT

[14] Newey, W. K. (1993): Efficient Estimation of Models with Conditional Moment Restrictions, in *Handbook of Statistics*, **11**, ed. by G. Maddala, C. Rao, and H. Vinod, Elsevier, 2111-2245.

[15] Smith, R. J. (2007): Efficient Information Theoretic Inference for Conditional Moment Restrictions, *Journal of Econometrics*, **138**, 430-460.

[16] Tripathi, G., Y. Kitamura (2003): Testing Conditional Moment Restrictions, *The Annals of Statistics*, **31**, 2059-2095.

[17] White, H. (1980): A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, **48**, 817-838.

ACCEPTED MANUSCRIPT

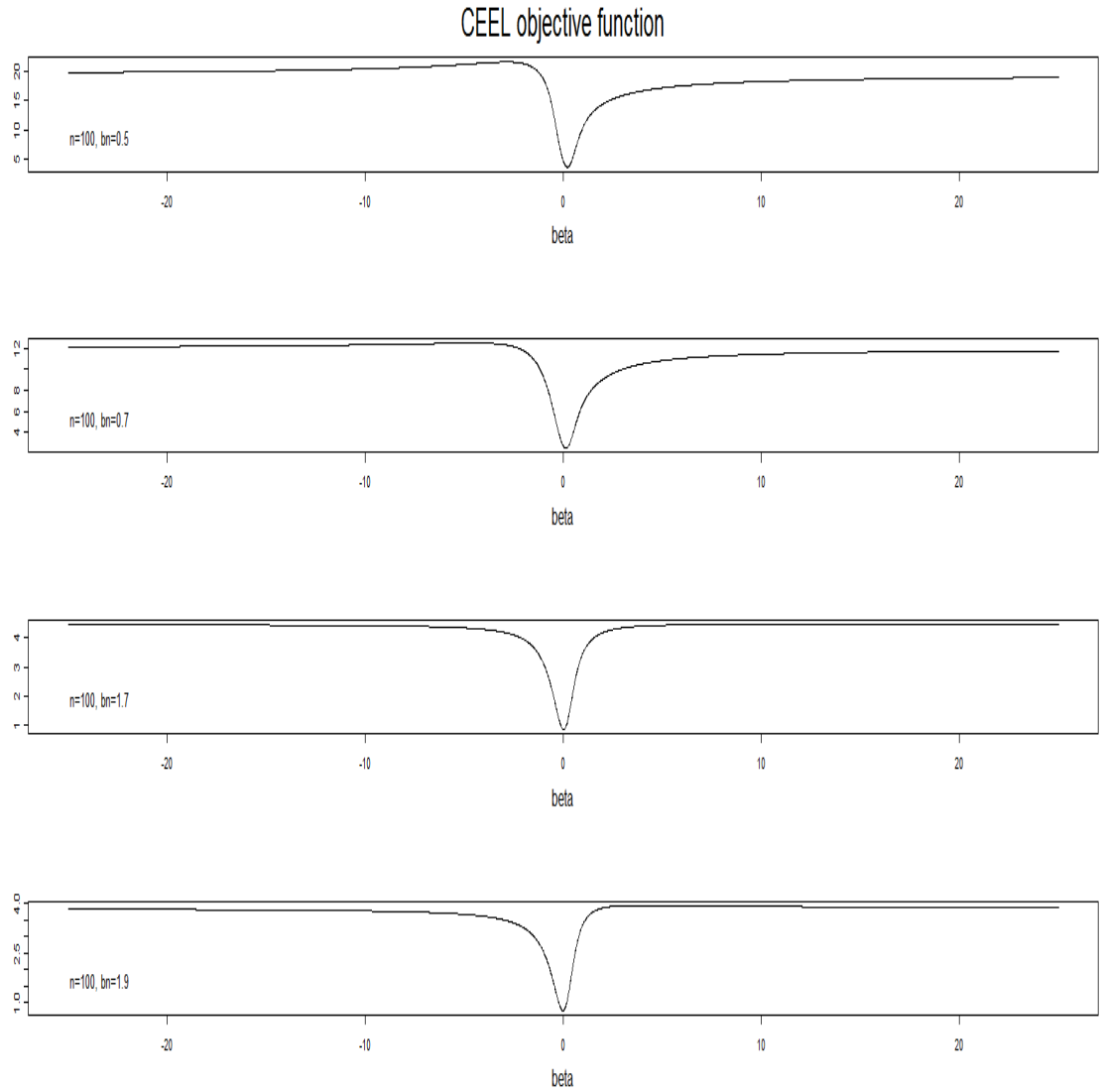

Figure 1: CEEL objective function in the case of strong instruments and low endogenenity (R^2 = 0.2, ρ = 0.3 and π = 0.4)

ACCEPTED MANUSCRIPT

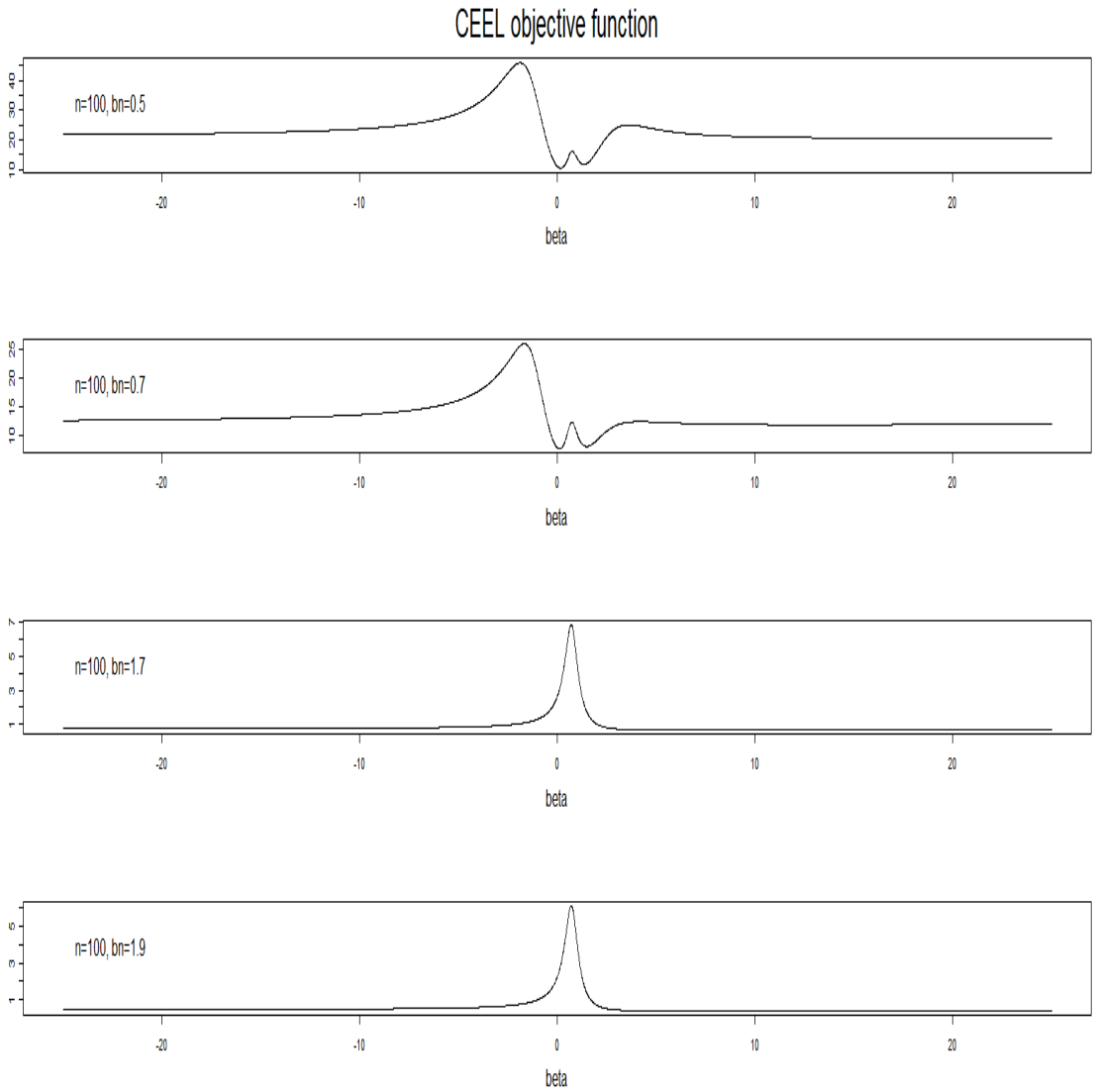

Figure 2: CEEL objective function in the case of weak instruments and high endogenenity (R^2 = 0.1, ρ = 0.75, and π = 0.04)