Journal of Econometrics 182 (2014) 247–268

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Near exogeneity and weak identification in generalized empirical likelihood estimators: Many moment asymptotics

Mehmet Caner

North Carolina State University, Department of Economics, 4168 Nelson Hall, Campus Box 8110, Raleigh, NC, 27695, United States

ARTICLE INFO

Article history:
Received 3 February 2010
Received in revised form
15 April 2014
Accepted 2 May 2014
Available online 17 May 2014

JEL classification:
C3
C13
C30

Keywords:
Violation of exogeneity
Anderson–Rubin test
Asymptotic size

ABSTRACT

This paper investigates the Generalized Empirical Likelihood (GEL) estimators when there are local violations of the exogeneity condition (near exogeneity) in the case of many weak moments. We also examine the tradeoff between the degree of violation of the exogeneity and the number of nearly exogenous instruments. In this respect, this paper extends many weak moment asymptotics of Newey and Windmeijer (2009a). The overidentifying restrictions test can detect both mild and large violations of exogeneity. In the case of minor violations, the Anderson–Rubin (1949) and Wald tests are not size distorted.

1. Introduction

The issue of violation of the exogeneity condition in instrumental variable estimation is one important problem that must be addressed. We define near exogeneity as the local to zero violation of the perfect exogeneity condition. Selection of a perfectly exogenous instrument is often difficult. Note that this violation of the exogeneity condition in Generalized Method of Moments (GMM) setting is discussed by Newey (1985), and Hall and Inoue (2003) for the case of a fixed number of strong instruments. Bound et al. (1995) demonstrate that minor violations of the exogeneity condition, coupled with weak instruments, increase the bias in the coefficient estimates.

We think that many instruments setup is a natural place to find nearly exogenous instruments. A key issue is also the interaction of weak identification with near exogeneity in many instruments setup. In related literature, Newey and Windmeijer (2009a) develop many weak moment asymptotics for Generalized Empirical Likelihood (GEL) estimators with perfectly exogenous instruments. The new limit has larger asymptotic variance than the standard GEL limit of Newey and Smith (2004). This approximation improves the finite sample results. Chao and Swanson (2005) derive the linear case, and Han and Phillips (2006) consider GMM. Several recent papers have explored the testing of exogeneity violations. Guggenberger (2012) analyzes the size distortion of various tests in a fixed number of instruments with a linear setup and exogeneity violations. He concludes that the Anderson and Rubin (1949) test is less size distorted than the other tests. Berkowitz et al. (2012) provide a new resampling technique when there are local exogeneity violations. Also, Caner and Morrill (2010) tackle the inference problem of strong but invalid instruments. They show that a joint test of structural parameters and correlation parameters may be useful. Finally, in a working paper, Kolesar et al. (2011) analyze estimation when there are many invalid instruments in a linear context. Our paper goes in a different direction and analyzes the tests in the case of many weak and nearly exogenous instruments.

This article extends the previous literature in several ways. First, we allow many weak and nearly exogenous instruments. Their number may be equal to the total number of instruments. Second, we analyze various degrees of the violation of exogeneity, unlike the root n case in the previous literature. We identify the tradeoff between the degrees of violation and the number of nearly exogenous instruments. Next, we show that the overidentifying restrictions test can detect both mild and large violations. Note that the Anderson and Rubin (1949) test is not affected by minor violations.

We provide assumptions and the limits of the GEL estimators in Section 2. In Section 3, we discuss tests under the condition of many weak moments and near exogeneity. In Section 4, we conduct several simulations. Conclusion is in Section 5. The

E-mail address: mcaner@ncsu.edu.

http://dx.doi.org/10.1016/j.jeconom.2014.05.001

Appendix A covers the proofs, including the Appendix B which provides the details of the proofs in Section 2.

2. Many weak moment asymptotics

The model is

$$Eg(Z_i, \theta_0) = \frac{C_1}{n^\kappa}, \tag{1}$$

where C_1 is a $q_n \times 1$ vector of constants, $0 < \kappa < \infty$, and θ_0 is the true structural parameter vector of dimension p. The data $\{Z_i\}_{i=1}^n$ is iid. The elements of C_1 are not necessarily zeros as in the standard model, and are in a compact set. This is specified below in Assumption M.1. The setup is a generalization of Hall and Inoue (2003), and Newey (1985), where they impose $\kappa = 1/2$. $E(.)$ denotes the expectation taken with respect to Z_i for sample size n, we suppress the subscript n. In Section 3.2, we will generalize Eq. (1). For our analysis, we will use (1) as the basis for Assumption M.1.

The GEL estimator is defined as in Newey and Windmeijer (2009a): set $g_i(\theta) = g(Z_i, \theta)$, for all $i = 1, \ldots, n$,

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g_i(\theta))/n.$$

Let $\rho(.)$ be a real valued function, $\mathcal{V} \to R$, where $\mathcal{V}$ is an open interval of the real line that contains zero, and $\hat{\Lambda}_n(\theta) = \{\lambda : \lambda' g_i(\theta) \in \mathcal{V}$ for $i = 1, \ldots, n\}$. Also, we set $\theta \in \Theta$, where Θ is a compact subset of R^p, and define $\rho_j(\nu) = \partial^j \rho(\nu)/\partial \nu^j$, with $\rho_j = \rho_j(0)$ for nonnegative integers j. We want to estimate the unknown θ_0, which is the true parameter vector. We normalize $\rho(\nu)$, $\nu \in \mathcal{V}$ so that $\rho(0) = 0$, $\partial \rho(0)/\partial \nu = -1$, $\partial^2 \rho(0)/\partial \nu^2 = -1$. The moment function $g_i(\theta)$ is of the dimension $q_n \times 1$, and q_n increases with n. The relationship between q_n and n will be explained in assumptions, but $q_n/n \to 0$ as $n \to \infty$. Therefore q_n will grow slower than n. This is the approach taken by Newey and Windmeijer (2009a) to control the dimension of the variance term.

GEL consists of several interesting sub cases. There are specifically three estimators, that we use in econometrics. The first one is the Empirical Likelihood estimator of Owen (2001), Qin and Lawless (1994) and Imbens (1997). This is obtained from GEL when we set $\rho(\nu) = ln(1 - \nu)$, $\mathcal{V} = (-\infty, 1)$. Next, we have the Exponential Tilting estimator of Kitamura and Stutzer (1997), where we set $\rho(\nu) = -e^\nu + 1$ in the GEL estimator. The last one is the Continuous Updating estimator, where we set $\rho(\nu) = -\nu - \nu^2/2$, and the objective function has GMM like form

$$\hat{Q}(\theta) = \frac{1}{2}\left(n^{-1}\sum_{i=1}^n g_i(\theta)\right)' \left[n^{-1}\sum_{i=1}^n g_i(\theta) g_i(\theta)'\right]^{-1} \times \left(n^{-1}\sum_{i=1}^n g_i(\theta)\right),$$

which is shown in Newey and Smith (2004).

We can rewrite the GEL estimator in the following way. Denote for each $\theta \in \Theta$

$$\hat{Q}(\theta) = \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^n \rho(\lambda' g_i(\theta))/n,$$

and

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{Q}(\theta).$$

Newey and Windmeijer (2009a) present detailed explanations about why the GEL estimator is consistent under many weak asymptotics, whereas GMM cannot be consistent. The limit of the objective function in GMM consists of a "noise" term and a "signal" term. The noise term consists of a weight matrix multiplied by $\Omega(\theta) = Eg_i(\theta)g_i(\theta)'$. This noise does not disappear in large samples and contaminates the limit, which leads to inconsistency. This issue is shown by Han and Phillips (2006) and, subsequently, by Newey and Windmeijer (2009a). However, the noise term in the Continuous Updating Estimator (CUE) in GEL does not depend on θ since the weight matrix is $\Omega(\theta)^{-1}$. The CUE is consistent under many weak moment asymptotics. Since Newey and Windmeijer (2009a) have shown the GEL objective function is well approximated by CUE, any GEL estimator is also consistent.

There are several reasons that we use GEL rather than the two-step GMM. First, exogeneity violations in a two-step GMM are analyzed in Hall and Inoue (2003), where there are strong and valid instruments. Second, the two-step GMM is inconsistent in the many weak moments case. Note that GEL estimators are consistent even when there are many weak moments with near exogeneity. They are, therefore, more robust to the problems in data compared with GMM. But we also see that the asymptotic limit described by Newey and Windmeijer (2009a) may change when we have nearly exogenous instruments.

2.1. Assumptions

We start with the near exogeneity assumption.

Assumption M.1. (i)

$$Eg_i(\theta_0) = \frac{C_1}{n^\kappa},$$

where $0 < \kappa < \infty$ and C_1 is a $q_n \times 1$ vector. $C_1 = (0'_{q_n - l_n}, C'_{l_n})'$, C_{l_n} is an $l_n \times 1$ vector, where $l_n \to \infty$ when $n \to \infty$. For each $j = 1, \ldots, l_n$, $-\infty < C_a < C_{l_n,j} < C_b < \infty$, where C_a, C_b are scalars. $C_{l_n,j}$ is in a compact set S. $0_{q_n - l_n}$ represents a zero vector of the dimension $q_n - l_n$.

We allow for two possibilities regarding the ratio of the number of imperfect moment conditions l_n to total number of moment conditions q_n:

(ii) Let $l_n/q_n \to f$, as $n \to \infty$, where $0 < f \le 1$,
or
(iii) Let $l_n/q_n \to 0$, as $n \to \infty$.

Assumption M.1(i) expresses a very general form of violation of exogeneity. $\kappa = 1/2$ is a mild violation of exogeneity, but we consider the case of $1/2 < \kappa < \infty$, a minor violation of exogeneity. We consider $0 < \kappa < 1/2$ as the range of major violation of a perfect exogeneity condition. This approach is more general than the setups by Newey (1985), Hall and Inoue (2003), and Berkowitz et al. (2008, 2012), where there are fixed number of invalid instruments ($l_n = l$, and l is constant) and $\kappa = 1/2$. Note that we cannot allow for $\kappa = 0$, because it would violate one of the conditions needed for consistency. Assumptions M.1(ii) and (iii) illustrate two distinct possibilities between the ratio of imperfect moments (l_n) to the total number of moments (q_n). The first possibility is that the number of imperfect moments can be a positive fraction of all orthogonality conditions. This potentiality may include all the invalid orthogonality restrictions ($l_n = q_n$). The next possibility is Assumption M.1(iii), where we allow $l_n \to \infty$, but $l_n/q_n \to 0$. We separate these two cases, since consistency conditions are different in each. Note that C_a, C_b does not depend on n.

The following assumption explains the nature of many weak moment asymptotics very well. This is Assumption 1 of Newey and Windmeijer (2009a). Many weak moment asymptotics provide improvements in overidentified models. In those models, the finite sample improvements are substantial when the variance of Jacobian of the moment functions is large relative to its average. The many weak moments approximation is better than the standard Gaussian approximation when there are many weak moments. In the many weak moments case, the asymptotic variance is larger than the usual one. Note that $diag(M)$ represents a diagonal matrix M.