Economics Letters 120 (2013) 495–498

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

A solution to the weak instrument bias in 2SLS estimation: Indirect inference with stochastic approximation

Vincent Arel-Bundock

University of Michigan, Department of Political Science, 5700 Haven Hall, 505 S. State Street, Ann Arbor, MI 48109-1045, USA

HIGHLIGHTS

- Instrumental variable estimation using indirect inference (II).
- II performs better than 2SLS and Fuller LIML when the instruments are weak and/or numerous.
- Stochastic approximation allows more efficient computation of the II estimator.

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form
3 May 2013
Accepted 5 June 2013
Available online 11 June 2013

Keywords:
Weak instrument bias
Two-stage least squares
Indirect inference
Stochastic approximation
Bias correction

ABSTRACT

This paper frames indirect inference bias correction as a stochastic root-finding problem and proposes a computationally efficient method to solve it. The technique is applied to the many/weak instrument bias in two-stage least squares estimation. Monte Carlo experiments suggest that the bias-corrected estimator outperforms more common alternatives.

1. Introduction

The weak instrument problem in two-stage least squares (2SLS) estimation has received considerable attention over the past few years. This is warranted, since the estimator is widely used, and because we know it can be severely biased in finite samples when the instruments are weak and/or numerous. Many researchers have provided analytical expressions to characterize the weak instrument bias (see, e.g., Hahn and Hausman, 2002; Chao and Swanson, 2005; Bun and Windmeijer, 2011), and others have assessed the performance of 2SLS in small samples using Monte Carlo experiments. Hahn et al. (2004), for example, found that alternative estimation methods such as the Fuller (1977) modification of limited information maximum likelihood (LIML) can often outperform 2SLS.

Another solution which does not seem to have been the object of study so far is to use indirect inference (II) for bias correction. In this paper, I argue that the II approach holds considerable appeal because of its effectiveness, generality, and ease of implementation.[1] Whereas explicit analytic expressions for the bias of 2SLS are usually developed in the context of simple special cases, II relies on simulations which can easily be tailored to more complex model specifications.

Section 2 frames indirect inference bias correction as a stochastic root-finding problem and suggests a computationally efficient method to solve it. Section 3 applies II to instrumental variable estimation with many weak instruments. Monte Carlo experiments suggest that the II correction can outperform common alternatives in finite samples.

2. Indirect inference as stochastic root-finding

Indirect inference (II) is a simulation-based estimation method that was first developed in Smith (1993), Gouriéroux et al. (1993), and Gallant and Tauchen (1996). It can be seen as a generalization of the method of simulated moments.

E-mail address: varel@umich.edu.

[1] The procedure described in this paper was implemented in R. The code is made available under a permissive license on the author's website: http://umich.edu/~varel.

0165-1765/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2013.06.004

The key idea of II is that, if we can simulate data that mimic the properties of our observations, we can match functions of our real and artificial data in order to glean information about the bias function that affects our estimator of choice. In practice, we do this by generating many datasets using various plausible values of the parameters of interest. Then, we apply the biased estimator to each of these datasets and compare the results to those obtained in real data. When the "artificial" biased estimates match the "real" biased estimates, we can accept the values used for simulation as the II estimates of the truth.

The above procedure is one special case of a general approach which has applications beyond weak instrument bias correction; II has for example been used to estimate dynamic stochastic general equilibrium models (Dridi et al., 2007), dynamic panel models (Gouriéroux et al., 2010), and stable distributions (Garcia et al., 2011). The interested reader should refer to Gouriéroux et al. (1993) for an analysis of the conditions under which the II estimator has desirable large-sample properties.

I now define the II estimator more formally and propose a practical improvement in the way bias-corrected estimates can be computed. Consider some data y produced by the true model with parameter $\boldsymbol{\phi}$. Optimizing the objective function $\mathcal{Q}$ of the chosen estimation method gives us

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathcal{Q}(y(\boldsymbol{\phi})), \quad \text{with } E[\hat{\boldsymbol{\theta}}] \neq \boldsymbol{\phi}. \tag{1}$$

Let $\boldsymbol{\phi}^s$ be a plausible value of the unknown parameter we wish to estimate, y^s a dataset produced by simulation using $\boldsymbol{\phi}^s$, and $\boldsymbol{\theta}^s$ the result obtained by applying biased estimator $\mathcal{Q}$ to y^s. Because of the stochastic nature of the simulation process, different realizations h of y^s will typically yield different values of $\boldsymbol{\theta}^s$, even if $\boldsymbol{\phi}^s$ remains constant. Gouriéroux et al. (2010) thus define the indirect inference estimator as

$$\hat{\boldsymbol{\phi}}^{\text{II}} = \arg\min_{\boldsymbol{\phi}^s\in\Phi} \left\| \hat{\boldsymbol{\theta}} - \frac{1}{H}\sum_{h=1}^{H} \boldsymbol{\theta}_h^s(\boldsymbol{\phi}^s) \right\|, \tag{2}$$

where $\|\cdot\|$ is some distance metric, $\hat{\boldsymbol{\theta}}$ is the biased estimate obtained from the observed data, and $\boldsymbol{\theta}_h^s$ is the biased estimate obtained by applying estimator $\mathcal{Q}$ to an hth dataset simulated using parameter value $\boldsymbol{\phi}^s$. The intuition behind the use of H artificial datasets is that we want to avoid matching the observed $\hat{\boldsymbol{\theta}}$ to the idiosyncratic features of any one particular simulated dataset. By taking $1/H\sum_{h=1}^{H}\boldsymbol{\theta}_h^s(\boldsymbol{\phi}^s)$ with large enough H, we "average out" the noise that was introduced in the simulation and calibrate the simulation to find a match between $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$ and $E[\boldsymbol{\theta}^s(\boldsymbol{\phi}^s)]$. When we find such a match, we can accept $\boldsymbol{\phi}^s$ as the II estimate of $\boldsymbol{\phi}$.

Put slightly differently, the problem of bias correction is a special case of indirect inference in which our goal is to find the root $\boldsymbol{\phi}^s$ of a function $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) - \boldsymbol{\theta}^s(\boldsymbol{\phi}^s) \approx 0$, where $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$ is given by the data and $\boldsymbol{\theta}^s(\boldsymbol{\phi}^s)$ is a noisy quantity that is defined implicitly through simulation.[2] Viewing the problem in this way allows us to draw on a rich tradition of work on stochastic root-finding algorithms.[3]

One of the core insights of this literature, dating back to the seminal contribution of Robbins and Monro (1951), is that the averaging strategy employed in Eq. (2) is computationally wasteful. Indeed, when trying to find the root of a stochastic function, it is often much more efficient to "average out" noise *across* iterations rather than *at every step* of the optimization. Under fairly permissive regularity conditions (Spall, 2003, Chapter 4), the stochastic approximation (SA) algorithm finds the root of a noisy function simply by[4]

$$\boldsymbol{\phi}_{k+1}^s = \boldsymbol{\phi}_k^s - a_k(\hat{\boldsymbol{\theta}}_k^s - \hat{\boldsymbol{\theta}}), \tag{3}$$

where k denotes the kth step of the SA algorithm, $\hat{\boldsymbol{\theta}}$ is the initial biased estimate we wish to correct, $\hat{\boldsymbol{\theta}}_k^s$ is obtained by applying the biased estimator $\mathcal{Q}$ to an artificial dataset that was simulated with parameter $\boldsymbol{\phi}_k^s$, and a_k is element k in a "gain" sequence of positive constants that converges to zero.[5]

In the next section, I use this strategy to correct bias in 2SLS models with many weak instruments. In each experiment, the corrected estimates were obtained using just 100 evaluations of the biased estimator (in total, not per iteration). In contrast, Gouriéroux et al. (2010) use 15 000 artificial datasets at every step of the algorithm they use to optimize Eq. (2). So while it is true that the II estimator is more computationally involved than alternatives such as LIML and 2SLS, this article shows that it is possible to compute II estimates using a reasonable amount of resources.

3. 2SLS many weak instruments bias

It is well known that 2SLS estimates can be biased toward ordinary least squares in finite samples when the instruments are weak and/or numerous. The experiments described below suggest that this bias can be equivalent to many times the true coefficient value, and "even enormous samples do not eliminate the possibility of quantitatively important finite-sample biases" (Bound et al., 1995, 446).

Consider a simple model with dependent variable y_i, one endogenous regressor x_i, K instruments Z_i, error vectors ε and υ, and unit index i (Hahn and Hausman, 2002; Hahn and Kuersteiner, 2002; Bun and Windmeijer, 2011):

$$y_i = x_i\phi + \varepsilon_i \tag{4}$$
$$x_i = Z_i'\Pi + \upsilon_i$$
$$\begin{pmatrix}\varepsilon_i\\ \upsilon_i\end{pmatrix} \sim IIN\left(0, \begin{bmatrix}\sigma_\varepsilon^2 & \sigma_{\varepsilon\upsilon}\\ \sigma_{\varepsilon\upsilon} & \sigma_\upsilon^2\end{bmatrix}\right).$$

The (biased) estimate is given by

$$\hat{\theta}_{2SLS} = \frac{x'Z(Z'Z)^{-1}Z'y}{x'Z(Z'Z)^{-1}Z'x} = \phi + \frac{x'Z(Z'Z)^{-1}Z'\varepsilon}{x'Z(Z'Z)^{-1}Z'x}. \tag{5}$$

Hahn and Hausman (2002) derive the following expression for the 2SLS bias:

$$E[\hat{\theta}_{2SLS}] - \phi \approx \frac{K\cdot\sigma_{\varepsilon\upsilon}}{R^2}\frac{1}{\sum_{i=1}^{n} x_i^2}, \tag{6}$$

[2] MacKinnon and Smith (1998) use a similar formulation but do not emphasize the stochastic nature of this function.

[3] See Pasupathy and Kim (2011) for a recent survey.

[4] The SA algorithm will only be appropriate where the quantity of interest and the biased estimates we are matching have the same dimension and substantive interpretation. Where this is not the case, for example if we conduct II by matching on higher moments or some other function of the data, one could turn to another stochastic optimization algorithm such as the simultaneous perturbation stochastic approximation algorithm of Spall (1992). I should also note that deterministic derivative-based root-finding methods such as the Newton–Raphson method can be difficult to apply in the II bias correction context because the shape of the bias function which relates $\boldsymbol{\phi}$ to $\boldsymbol{\theta}$ will generally be unknown, and because estimating derivatives numerically can be prohibitively expensive given the stochastic nature of the underlying simulation process.

[5] A bad choice of gain sequence can affect the performance of the SA algorithm. Spall (2003) offers good practical advice for the choice of a_k, and Broadie et al. (2011) propose a simple self-tuning method. In this work, I used $a_k = k^{-1}$. I also ran tests using the more flexible gain sequence recommended in Spall (2003), but obtained nearly identical results.