ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Higher-order properties of approximate estimators

Dennis Kristensen [a,b,c,*], Bernard Salanié [d]

[a] Department of Economics, University College London, Institute of Fiscal Studies, United Kingdom

[b] CeMMaP, United Kingdom

[c] CREATES, Denmark

[d] Department of Economics, Columbia University, United States

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form
19 May 2016
Accepted 23 October 2016
Available online xxxx

JEL classification:
C13
C15
C63

Keywords:
Extremum estimators
Numerical approximation
Simulation-based estimation
Higher-order expansion
Bias adjustment

ABSTRACT

Many modern estimation methods in econometrics approximate an objective function, for instance, through simulation or discretization. These approximations typically affect both bias and variance of the resulting estimator. We first provide a higher-order expansion of such "approximate" estimators that takes into account the errors due to the use of approximations. We show how a Newton–Raphson adjustment can reduce the impact of approximations. Then we use our expansions to develop inferential tools that take into account approximation errors: we propose adjustments of the approximate estimator that remove its first-order bias and adjust its standard errors. These corrections apply to a class of approximate estimators that includes all known simulation-based procedures. A Monte Carlo simulation on the mixed logit model shows that our proposed adjustments can yield significant improvements at a low computational cost.

1. Introduction

The complexity of econometric models has grown steadily over the past three decades. The increase in computer power contributed to this development in various ways, and in particular by allowing econometricians to estimate more complicated models using methods that rely on approximations. Examples include simulated method of moments (McFadden, 1989; Pakes and Pollard, 1989; Duffie and Singleton, 1993; Creel and Kristensen, 2012), simulated maximum likelihood (Lee, 1992, 1995; Fermanian and Salanié, 2004; Kristensen and Shin, 2012), and approximate solutions to structural models (Rust, 1987; Tauchen and Hussey, 1991; Fernández-Villaverde and Rubio-Ramirez, 2005; Fernández-Villaverde et al., 2006; Norets, 2012; Kristensen and Schjerning, 2015). In all of these cases, the objective function defining the estimator includes a component which is approximated using some type of numerical algorithm. We will refer to this component as the *approximator*, and call the resulting estimator an *approximate estimator*. Taking the approximation error to zero defines an infeasible estimator which we call the *exact estimator*. In simulation-based inference, for instance, the exact estimator would be obtained with an infinite number of simulations. In dynamic programming models solved by discretization the exact estimator would rely on an infinitely fine grid.

The use of approximations usually deteriorates the properties of the approximate estimator relative to those of the corresponding exact estimator: the former may suffer from additional biases and/or variances compared to the latter. When the approximation error is non-stochastic, its main effect is to impart additional bias to the estimator. On the other hand, stochastic approximations not only create bias; they may also reduce efficiency. The effect of the approximation on the estimator can usually be reduced by choosing a sufficiently fine approximation; but this comes at the cost of increased computation time. In many applications this may be a seriously limiting factor; increased computer power helps, but it also motivates researchers to work on more complex models. It is therefore important to quantify the additional estimation errors that approximators generate, and also to account for these additional errors in order to draw correct inference.

As a first step in this direction, we analyze the higher-order properties of the approximate estimator in a general setting. These

* Correspondence to: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom.
E-mail address: d.kristensen@ucl.ac.uk (D. Kristensen).

http://dx.doi.org/10.1016/j.jeconom.2016.10.008
0304-4076/

Please cite this article in press as: Kristensen, D., Salanié, B., Higher-order properties of approximate estimators. Journal of Econometrics (2017), http://dx.doi.org/10.1016/j.jeconom.2016.10.008

expansions apply to a very large class of models, and can be used to develop a number of adjustments to estimators and/or standard errors that open the way to better inference. To show this, we develop analytical bias and variance adjustments for a large class of approximate estimators where the approximation is stochastic, including most standard simulation-based estimators. We also propose a very generally applicable two-step method; it consists of updating the approximate estimator obtained by one or several Newton–Raphson iterations based on the same objective function, but with a much finer degree of approximation. These different methods can of course be combined when they both apply.

Our theoretical results apply to generalized method of moment estimators as well as M-estimators, both when the approximation is stochastic and when it is not. The results encompass and extend results in the literature on simulation-based estimators. Moreover, the expansion can be used to analyze the behavior of estimators that rely on numerical solutions to structural dynamic models as cited above. Our results also apply to many estimators used in empirical IO, which combine simulation and numerical approximation. And it also covers situations where numerical derivatives are used, either for computation of variance estimators or optimization algorithms based on Newton iterations[1]. To the best of our knowledge, this is the first paper to provide results for such a general class of models.

To test the practical performance of our proposed adjustment methods, we run a simulation study on a mixed logit model. The mixed logit is one of the basic building blocks in much work on demand analysis, for example; and it is simple enough that we can compute the true value of the biases and efficiency losses, as well as our estimated corrections. We show that uncorrected SML has non-negligible bias, even for large sample sizes; and that standard confidence intervals can be wildly off the mark. Our analytical adjustment removes most of the bias at almost no additional computational cost; and it yields very reliable confidence intervals. The Newton–Raphson correction also reduces the bias and improves confidence intervals, but it does so less effectively than the analytical adjustment.

In a recent paper, Freyberger (2015) derived analytical adjustments for the Berry–Levinsohn–Pakes (1995) model when the numbers of consumers and/or the number of simulation draws are finite. His approach is similar to ours: his results are less general, but since he only deals with a specific model his assumptions are more primitive and his formulæ more explicit. We complement his work by providing the formulæ for our Newton–Raphson adjustment for this model in Section 6.2.

The paper is organized as follows. Section 2 presents our framework and some examples. In Section 3, we derive a higher-order expansion of the approximate estimator relative to the exact one. We describe our Newton–Raphson correction in Section 4. Then in Section 5 we build on the expansion to propose adjusted estimators, standard errors, and confidence intervals. Section 6 applies the general theory to two specific approximate estimators, while Section 7 presents the results of a Monte Carlo simulation study using the simulated MLE of the mixed logit model as an example. We discuss possible extensions of our results in Section 8. Appendices A and B contain proofs of the main results and lemmas, respectively. Appendix C provides details for two examples of our theory, and Appendix D outlines how the theory can be generalized to handle multiple approximators with different properties.

[1] However, in most of our examples, we abstract away from issues with numerical maximization that sometimes arise when computing extremum estimators.

2. Framework

Given a sample $\mathcal{Z}_n = \{z_1, \ldots, z_n\}$ of n observations, our aim is to estimate a parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^k$ through an estimating equation that the "exact" estimator $\hat{\theta}_n$ is set to solve,

$$G_n(\hat{\theta}_n, \gamma_0) = o_P\left(1/\sqrt{n}\right), \text{ where}$$
$$G_n(\theta, \gamma) = \frac{1}{n}\sum_{i=1}^{n} g(z_i; \theta, \gamma), \quad (1)$$

and $g(z; \theta, \gamma)$ is a known functional that depends on data, z, the parameter of interest, θ, and a nuisance parameter γ. We here and in the following let γ_0 denote the true, but unknown value of γ. The nuisance parameter γ could be finite-dimensional, but in most situations it is a parameter dependent function, $u \mapsto \gamma(u; \theta)$. The nature of the argument u of the function γ will depend on the application; it could be covariates relative to one observation, the value of a conditional moment, or more complex objects. This is irrelevant for our general theory.

Suppose that the object γ_0 is not known in closed form to the econometrician, so that the estimator $\hat{\theta}$ is infeasible. Instead, we approximate γ_0 by $\hat{\gamma}_S$ that depends on some approximation scheme of order S (e.g. S simulations, or a discretization on a grid of size S), and compute the corresponding "approximate" estimator $\hat{\theta}_{n,S}$ satisfying

$$G_n(\hat{\theta}_{n,S}, \hat{\gamma}_S) = o_P\left(1/\sqrt{n}\right). \quad (2)$$

Our first aim is to analyze the impact of approximations: How do they impact the distribution of $\hat{\theta}_{n,S}$? This analysis is in turn used to propose methods that reduce the biases and variances due to approximations, and adjust standard errors to take into account additional noise due to approximations.

We restrict attention throughout to the case of smooth approximators where $\hat{\gamma}_S(u; \theta)$ is, as a minimum, differentiable w.r.t. θ. Moreover, while γ may be a vector-valued function, we will in the main text assume that the biases and variances due to approximations of its different components vanish at the same rate. This is merely to save on notation, and Appendix D provides results for the case of multiple approximators with possibly different rates.

We now present a few examples that fall within the above setting:

Example 1 (*Approximate M-Estimators*)**.** Consider an M-estimator $\hat{\theta} = \arg\max_{\theta \in \Theta} Q_n(\theta, \gamma_0)$, where $Q_n(\theta, \gamma) = \sum_{i=1}^{n} q(z_i; \theta, \gamma)/n$. In this case, we set $g(z; \theta, \gamma) = \partial q(z; \theta, \gamma)/(\partial\theta)$. This covers simulated maximum likelihood estimator (SMLE) where $q(z; \theta, \gamma) = \log\gamma(z; \theta)$ and γ_0 is a density that is computed by simulations. It also includes simulated pseudo-maximum likelihood (Laroque and Salanié, 1989) where $q(z, \gamma; \theta) = -(y - \gamma(x; \theta))^2$ and $\gamma_0(x; \theta) = E[y|x; \theta]$ is a conditional moment which is computed by simulations.

Example 2 (*Approximate GMM-Estimators*)**.** Suppose that $\hat{\theta}$ is defined as in Example 1, but now $Q_n(\theta, \gamma) = M_n(\theta, \gamma)' W_n M_n(\theta, \gamma)$ where $M_n(\theta, \gamma) = \sum_{i=1}^{n} m(z_i, \gamma; \theta)/n$ is a set of sample moments and $W_n \overset{P}{\to} W > 0$. Then we set $g(z_i; \theta, \gamma) = H(\theta, \gamma) W m(z_i; \theta, \gamma)$ where $H(\theta, \gamma) = E[\partial m(z_i; \theta, \gamma)/(\partial\theta)]$. This includes simulated method of moments (SMM), where $m(z, \gamma; \theta) = m(y) - \gamma(x; \theta)$ and $\gamma(x; \theta) = E[m(y)|x; \theta]$, and indirect inference (Gouriéroux and Monfort, 1996) where the estimator of the auxiliary model's parameters, β, can be expressed as $\hat{\beta} = \beta(\theta_0) + \sum_{i=1}^{n} m(z_i)/n + o_P\left(n^{-1/2}\right)$ and $\gamma(\theta) = E[\hat{\beta}|\theta] = \beta(\theta) + E[m(z_i)|\theta] + o\left(n^{-1/2}\right)$.