Journal of Econometrics 186 (2015) 51–65

Contents lists available at ScienceDirect

Journal of Econometrics

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

Asymptotically exact inference in conditional moment inequality models

Timothy B. Armstrong

Yale University, United States

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form
15 December 2014
Accepted 1 January 2015
Available online 20 January 2015

ABSTRACT

This paper derives the rate of convergence and asymptotic distribution for a class of Kolmogorov–Smirnov style test statistics for conditional moment inequality models for parameters on the boundary of the identified set under general conditions. Using these results, I propose tests that are more powerful than existing approaches for choosing critical values for this test statistic. I quantify the power improvement by showing that the new tests can detect alternatives that converge to points on the identified set at a faster rate than those detected by existing approaches. A Monte Carlo study confirms that the tests and the asymptotic approximations they use perform well in finite samples. In an application to a regression of prescription drug expenditures on income with interval data from the Health and Retirement Study, confidence regions based on the new tests are substantially tighter than those based on existing methods.

1. Introduction

Theoretical restrictions used for estimation of economic models often take the form of moment inequalities. Examples include models of consumer demand and strategic interactions between firms, bounds on treatment effects using instrumental variables restrictions, and various forms of censored and missing data (see, among many others, Manski, 1990; Manski and Tamer, 2002; Pakes et al., 2006; Ciliberto and Tamer, 2009; Chetty, 2010, and papers cited therein). For these models, the restriction often takes the form of moment inequalities conditional on some observed variable. That is, given a sample $(X_1, W_1), \dots, (X_n, W_n)$, we are interested in testing a null hypothesis of the form $E(m(W_i, \theta)|X_i) \geq 0$ with probability one, where the inequality is taken elementwise if $m(W_i, \theta)$ is a vector. Here, $m(W_i, \theta)$ is a known function of an observed random variable W_i, which may include X_i, and a parameter $\theta \in \mathbb{R}^{d_\theta}$, and the moment inequality defines the identified set $\Theta_0 \equiv \{\theta | E(m(W_i, \theta)|X_i) \geq 0 \text{ a.s.}\}$ of parameter values that cannot be ruled out by the data and the restrictions of the model.

In this paper, I consider inference in models defined by conditional moment inequalities. I focus on test statistics that exploit the equivalence between the null hypothesis $E(m(W_i, \theta)|X_i) \geq 0$ almost surely and $Em(W_i, \theta)I(s < X_i < s + t) \geq 0$ for all (s, t). Thus, we can use $\inf_{s,t} \frac{1}{n} \sum_{i=1}^{n} m(W_i, \theta)I(s < X_i < s + t)$, or the infimum of some weighted version of the unconditional moments indexed by (s, t). Following the terminology commonly used in the literature, I refer to these as Kolmogorov–Smirnov (KS) style test statistics. The main contribution of this paper is to derive the rate of convergence and asymptotic distribution of this test statistic for parameters on the boundary of the identified set under a general set of conditions.

While asymptotic distribution results are available for this statistic in some cases (Andrews and Shi, 2013; Kim, 2008), the existing results give only a conservative upper bound of $\sqrt{n}$ on the rate of convergence of this test statistic in a large class of important cases. For example, in the interval regression model, the asymptotic distribution of this test statistic for parameters on the boundary of the identified set and the proper scaling needed to achieve it have so far been unknown in the generic case (see Section 2 for the definition of this model). In these cases, results available in the literature do not give an asymptotic distribution result, but state only that the test statistic converges in probability to zero when scaled up by $\sqrt{n}$. This paper derives the scaling that leads to a non-degenerate asymptotic distribution and characterizes this distribution. Existing results can be used for conservative inference in these cases (along with tuning parameters to prevent the critical value from going to zero), but lose power relative to procedures that use the results derived in this paper to choose critical values based on the asymptotic distribution of the test statistic on the boundary of the identified set.

To quantify this power improvement, I show that using the asymptotic distributions derived in this paper gives power against sequences of parameter values that approach points on the boundary of the identified set at a faster rate than those detected using root-n convergence to a degenerate distribution. Since local

E-mail address: timothy.armstrong@yale.edu.

http://dx.doi.org/10.1016/j.jeconom.2015.01.002
0304-4076/

power results have not been available for the conservative approach based on root-n approximations in this setting, making this comparison involves deriving new local power results for the existing tests in addition to the new tests. The increase in power is substantial. In the leading case considered in Section 3, I find that the methods developed in this paper give power against local alternatives that approach the identified set at a $n^{-2/(d_X+4)}$ rate (where d_X is the dimension of the conditioning variable), while using conservative $\sqrt{n}$ approximations only gives power against $n^{-1/(d_X+2)}$ alternatives. The power improvements are not completely free, however, as the new tests require smoothness conditions not needed for existing approaches, and are shown to control a weaker notion of size (see the discussion at the end of Section 6). In another paper (Armstrong, 2011, 2014), I propose a modification of this test statistic that achieves a similar power improvement (up to a $\log n$ term) without sacrificing the robustness of the conservative approach (see also the more recent work of Armstrong and Chan 2012 and Chetverikov 2012).

Broadly speaking, the power improvement is related to the tradeoff between bias and variance for nonparametric kernel estimators (see, e.g. Pagan and Ullah, 1999, for an introduction to this topic). Under certain types of null hypotheses, the infimum in the test statistic is taken on a value of (s, t) with $t \to 0$ as the sample size increases. Here, t can be thought of as a bandwidth parameter that is chosen automatically by the test. The asymptotic approximations can be thought of as showing how t is chosen, which allows for less conservative critical values. See Section 2 for more intuition for these results.

To examine how well these asymptotic approximations describe sample sizes of practical importance, I perform a Monte Carlo study. Confidence regions based on the tests proposed in this paper have close to the nominal coverage in the Monte Carlos, and shrink to the identified set at a faster rate than those based on existing tests. In addition, I provide an empirical illustration examining the relationship between out of pocket prescription spending and income in a data set in which out of pocket prescription spending is sometimes missing or reported as an interval. Confidence regions for this application constructed using the methods in this paper are substantially tighter than those that use existing methods.

The rest of the paper is organized as follows. The rest of this section discusses the relation of these results to the rest of the literature, and introduces notation and definitions. Section 2 gives a nontechnical exposition of the results, and explains how to implement the procedures proposed in these papers. Together with the statements of the asymptotic distribution results in Section 3 and the local power results in Section 7, this provides a general picture of the results of the paper. Section 5 generalizes the asymptotic distribution results of Section 3, and Sections 4 and 6 deal with estimation of the asymptotic distribution for feasible inference. Section 8 presents Monte Carlo results. Section 9 presents the empirical illustration. Section 10 concludes. Proofs and other auxiliary material are in the supplementary appendix (see Appendix A).

1.1. Related literature

The results in this paper relate to recent work on testing conditional moment inequalities, including papers by Andrews and Shi (2013), Kim (2008), Khan and Tamer (2009), Chernozhukov et al. (2009), Lee et al. (2011), Ponomareva (2010), Menzel (2008) and Armstrong (2011). The results on the local power of asymptotically exact and conservative KS statistic based procedures derived in this paper are useful for comparing confidence regions based on KS statistics to other methods of inference on the identified set proposed in these papers. Armstrong (2011) derives local power results for some common alternatives to the KS statistics based on integrated moments considered in this paper (the confidence regions considered in that paper satisfy the stronger criterion of containing the entire identified set, rather than individual points, with a prespecified probability).

Out of these existing approaches to inference on conditional moment inequalities, the papers that are most closely related to this one are those by Andrews and Shi (2013) and Kim (2008), both of which consider statistics based on integrating the conditional inequality. As discussed above, the main contributions of the present paper relative to these papers are (1) deriving the rate of convergence and nondegenerate asymptotic distribution of this statistic for parameters on the boundary of the identified set in the common case where the results in these papers reduce to a statement that the statistic converges to zero at a root-n scaling and (2) deriving local power results that show how much power is gained by using critical values based on these new results. Armstrong (2011, 2014) uses a statistic similar to the one considered here, but proposes an increasing sequence of weightings ruled out by the papers above (and the present paper). This leads to almost the same power improvement as the methods in this paper even when conservative critical values are used. This approach has been further explored by Armstrong and Chan (2012) and Chetverikov (2012) (both of these papers were first circulated after the first draft of the present paper).

Khan and Tamer (2009) propose a statistic similar to the one considered here for a model defined by conditional moment inequalities, but consider point estimates and confidence intervals based on these estimates under conditions that lead to point identification. Galichon and Henry (2009) propose a similar statistic for a class of partially identified models under a different setup. Statistics based on integrating conditional moments have been used widely in other contexts as well, and go back at least to Bierens (1982).

The literature on models defined by finitely many unconditional moment inequalities is more developed, but still recent. Papers in this literature include Andrews et al. (2004), Andrews and Jia (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Chernozhukov et al. (2007), Romano and Shaikh (2010, 2008), Bugni (2010), Beresteanu and Molinari (2008), Moon and Schorfheide (2009), Imbens and Manski (2004) and Stoye (2009) and many others.

1.2. Notation

I use the following notation in the rest of the paper. For observations $(X_1, W_1), \ldots, (X_n, W_n)$ and a measurable function h on the sample space, $E_n h(X_i, W_i) \equiv \frac{1}{n}\sum_{i=1}^{n} h(X_i, W_i)$ denotes the sample mean. I use double subscripts to denote elements of vector observations so that $X_{i,j}$ denotes the jth component of the ith observation X_i. Inequalities on Euclidean space refer to the partial ordering of elementwise inequality. For a vector valued function $h : \mathbb{R}^{\ell} \to \mathbb{R}^{m}$, the infimum of h over a set T is defined to be the vector consisting of the infimum of each element: $\inf_{t\in T} h(t) \equiv (\inf_{t\in T} h_1(t), \ldots, \inf_{t\in T} h_m(t))$. I use $a \wedge b$ to denote the elementwise minimum and $a \vee b$ to denote the elementwise maximum of a and b. The notation $\lceil x \rceil$ denotes the least integer greater than or equal to x.

2. Overview of results

This section gives a description of the main results at an intuitive level, and gives step-by-step instructions for one of the tests proposed in this paper. Section 2.1 defines the terms "asymptotically exact" and "asymptotically conservative" for the purposes if this paper, and explains how the results in this paper lead to asymptotically exact inference. Section 2.2 describes the asymptotic distribution result, and explains why the situations that lead to it are important in practice. Section 2.3 describes the reason for the power improvement. Section 2.4 gives instructions for implementing the test.