Economics Letters 131 (2015) 78–82

Contents lists available at ScienceDirect

Economics Letters

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

A consistent bootstrap procedure for nonparametric symmetry tests

Daniel J. Henderson [a], Christopher F. Parmeter [b,*]

[a] Department of Economics, Finance and Legal Studies, University of Alabama, Box 870224, Tuscaloosa, AL 35487-0224, United States
[b] Department of Economics, University of Miami, United States

HIGHLIGHTS

- Symmetry important economically and statistically.
- Propose bootstrap procedure for nonparametric test of symmetry.
- Demonstrate consistency of bootstrap procedure.
- Testing approach displays impressive performance.
- Empirical examples highlight the appeal of this test in practice.

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form
19 February 2015
Accepted 29 March 2015
Available online 3 April 2015

JEL classification:
C1
C12
C14

Keywords:
U-statistic
Symmetry
Kernel
Bandwidth

ABSTRACT

This paper proposes a bootstrap algorithm for testing symmetry of a univariate density. Validity of the bootstrap procedure is shown theoretically as well as via simulations. Three empirical examples demonstrate the versatility of the test in practice.

1. Introduction

Symmetry is of interest in many areas of econometrics and statistics. For the linear regression model, Bickel (1982) demonstrated that if the conditional density function of the disturbance is symmetric, the regression coefficients can be estimated adaptively, implying attainability of efficiency equivalent to that of the maximum likelihood estimator. Symmetry is also used as an identification condition for the semiparametric sample selection model; Chen and Zhou (2010) proposed a $\sqrt{n}$-consistent estimator of the regression coefficients only through joint symmetry. Polonik and Yao (2000) used symmetry to construct predictive regions for nonlinear time series. Lastly, a key identification condition in the standard stochastic frontier regression model is that the composed error term is asymmetric (Waldman, 1982; Simar and Wilson, 2010).[1]

Beyond the statistical benefits of symmetry, there exist an array of economic settings where symmetry (or asymmetry) provides key insights. For example, Christofides and Stengos (2001, 2002) empirically study the symmetry of the wage-change distribution. During periods of high price inflation, real wages can decrease (even if nominal wages increase) if shocks are symmetric, leading to a symmetric cross-sectional nominal wage-change distribution. However, if nominal wages are rigid, then the cross-sectional nominal wage change distribution is likely asymmetric.

* Corresponding author.
E-mail addresses: djhender@cba.ua.edu (D.J. Henderson), cparmeter@bus.miami.edu (C.F. Parmeter).

http://dx.doi.org/10.1016/j.econlet.2015.03.038
0165-1765/

[1] See Li (1996a) for an example where the composed error term can be symmetric.

Testing symmetry can be implemented within the confines of a nonparametric test to avoid issues of misspecification. To that end, several tests of symmetry exist. Fan and Gençay (1995) provide a nonparametric test of symmetry of the disturbances in a linear regression model. Ahmad and Li (1997) provide a kernel based test of symmetry for both a univariate and bivariate density, as well as an alternative test in the regression setting (as in Fan and Gençay, 1995). Each of the available kernel based tests of symmetry currently rely on the asymptotic distribution to conduct inference. It is commonly known that kernel based tests tend to converge slowly towards their asymptotic distribution; see Li and Wang (1998, Table 1) or Henderson and Parmeter (2015, Fig. 4.6).

As an alternative to a kernel based test, characteristic function-based tests can help circumvent the slow convergence of kernel based tests. Given that these types of tests do not rely on smoothing the data, gains in convergence rates can be achieved. To that end, we mention that the conditional symmetry test of Su and Jin (2005) could be appropriately modified for our scenario; when the conditioning variables are abstract, their characteristic function based test is the same as that of Su (2006) in the univariate setting.

That being said, within the context of kernel based tests, a known remedy for slow convergence is to deploy a bootstrap procedure. An appropriate bootstrap can provide asymptotic refinements (Li and Wang, 1998, Thm. 3) and lead to improved inference in finite samples. The present work considers the kernel based test of Ahmad and Li (1997) for which we propose a bootstrap algorithm and prove consistency of this approach. As an anonymous referee correctly notes, the bootstrap we propose is similar to that of Su and Jin (2005).[2] The gains from the proposed bootstrap are demonstrated through a set of Monte Carlo simulations as well as three empirical examples.

2. The nonparametric symmetry test of Ahmad and Li (1997)

Our interest is in the shape of the unknown probability density function $f(\cdot)$. The null hypothesis for the test of symmetry is

$$H_0 : f(x) = f(-x)$$

almost everywhere versus the alternative

$$H_1 : f(x) \neq f(-x)$$

on a set with positive measure. Note here that the null hypothesis focuses on the classic case of symmetry about the origin. If one was interested in the less common notion of symmetry about a particular point, everything will hold suit if the data are re-centered around that point.

To test symmetry, Ahmad and Li (1997) use the integrated square error (ISE) metric, resulting in

$$ISE(f, f_-) = \int_x [f(x) - f(-x)]^2 \, dx.$$

Note that when the density is symmetric $f(x) = f(-x)$ and $ISE(f, f_-) = 0$. $ISE(f, f_-) > 0$ otherwise, making it a proper metric for inference.

Ahmad and Li (1997) show that their test statistic for the null hypothesis of symmetry is

$$\widehat{J}_n = nh^{1/2} \frac{\widehat{ISE}_n}{\widehat{\sigma}_n} \xrightarrow{d} N(0, 1),$$

where

$$\widehat{ISE}_n = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[k\left(\frac{x_i - x_j}{h}\right) - k\left(\frac{x_i + x_j}{h}\right) \right]$$

and

$$\widehat{\sigma}_n^2 = \frac{2}{n(n-1)h} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[k^2\left(\frac{x_i - x_j}{h}\right) + k^2\left(\frac{x_i + x_j}{h}\right) - 2k\left(\frac{x_i - x_j}{h}\right) k\left(\frac{x_i + x_j}{h}\right) \right].$$

$\widehat{ISE}_n$ does not contain the center term, $i = j$; omitting this term eliminates a non-zero bias from the asymptotic distribution. This is common in kernel based testing (Li, 1996b).

3. The bootstrap symmetry test

We first detail the assumptions needed to establish consistency of a bootstrap test for symmetry.

Assumption 3.1. $k(\cdot)$ is a bounded, symmetric density function such that $|u|k(u) \to 0$ as $|u| \to \infty$, where $u \equiv \frac{x_i - x}{h}$. Further, $\int uk(u)du = 0$ and $\int u^2 k(u)du < \infty$.

Assumption 3.2. The density $f(\cdot)$ is bounded and continuous on $\mathbb{R}$.

Assumption 3.3. As $n \to \infty$, $h \to 0$ and $nh \to \infty$.

Assumptions 3.1–3.3 are standard. We use a second-order kernel with bandwidth decaying to zero at the optimal rate. Given the lack of consensus regarding optimal bandwidth choice for testing, it is recommended that a variety of bandwidths are used empirically to demonstrate robustness of one's conclusions. Here, Assumption 3.3 is simply maintained to ensure proper behavior of the bias of the kernel based test. Ahmad and Li (1997) recommend undersmoothing empirically. We will not need to do so here.

A bootstrap test can be obtained by randomly resampling with replacement from the expanded set $\{-x_1, -x_2, \ldots, -x_n, x_1, x_2, \ldots, x_n\} = \mathcal{Z}_n$. $\mathcal{Z}_n^*$ will denote the bootstrap sample. Note that we are resampling n and not $2n$ observations and we use the n notation for posterity. The bootstrap version of $\widehat{ISE}_n$ is given as

$$\widehat{ISE}_n^* = \frac{1}{n(n-1)h^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[k\left(\frac{x_i^* - x_j^*}{h}\right) - k\left(\frac{x_i^* + x_j^*}{h}\right) \right] = \frac{2}{n^2 h} \sum_{i=1}^{n} \sum_{j>i}^{n} H(z_i^*, z_j^*),$$

where we will argue that $H(z_i^*, z_j^*)$ is a second-order degenerate U-statistic. The bootstrap version of the variance term is obtained as

$$\widehat{\sigma}_n^{2*} = \frac{2}{n(n-1)h} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[k^2\left(\frac{x_i^* - x_j^*}{h}\right) + k^2\left(\frac{x_i^* + x_j^*}{h}\right) - 2k\left(\frac{x_i^* - x_j^*}{h}\right) k\left(\frac{x_i^* + x_j^*}{h}\right) \right].$$

We are now in a position to establish the following lemma.

Lemma 3.1. *Let $\widehat{J}_n^* = nh^{1/2}\widehat{ISE}_n^*/\widehat{\sigma}_n^*$. Under Assumptions 3.1–3.3 we have*

$$\widehat{J}_n^* | \mathcal{Z}_n \to N(0, 1)$$

in distribution.

[2] The bootstrap procedure in Su and Jin (2005) uses the same steps (albeit with different estimators) as that proposed here.