ARTICLE IN PRESS
Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
Journal of Econometrics
journal homepage: www.elsevier.com/locate/jeconom

Testing over-identifying restrictions without consistent estimation of the asymptotic covariance matrix

Wei-Ming Lee [a], Chung-Ming Kuan [b,*], Yu-Chin Hsu [c]

[a] Department of Economics, National Chung Cheng University, Taiwan
[b] Department of Finance, National Taiwan University, Taipei 106, Taiwan
[c] Institute of Economics, Academia Sinica, Taiwan

ARTICLE INFO

Article history:
Received 11 February 2014
Received in revised form
11 February 2014
Accepted 6 April 2014
Available online xxxx

JEL classification:
C12
C22

Keywords:
GMM
Kernel function
KVB approach
Over-identifying restrictions
Robust test

ABSTRACT

We propose new over-identifying restriction (OIR) tests that are robust to heteroskedasticity and serial correlations of unknown form. The proposed tests do not require consistent estimation of the asymptotic covariance matrix and hence avoid choosing the bandwidth in nonparametric kernel estimation. Instead, they rely on the normalizing matrices that can eliminate the nuisance parameters in the limit. Compared with the conventional OIR test, the proposed tests require only a consistent, but not necessarily optimal, GMM estimator. Our simulations demonstrate that these tests are properly sized and may have power comparable with that of the conventional OIR test.

1. Introduction

The generalized method of moments (GMM) introduced in Hansen (1982) is a leading estimation technique in econometric applications. In the context of GMM, the validity of the moment conditions is tested using the over-identifying restriction (OIR) test. An OIR test can be made robust to heteroskedasticity and serial correlations of unknown form by employing a consistent covariance-matrix estimator. It is typical to compute such a consistent estimator using the nonparametric kernel method that requires choosing a kernel function and its bandwidth (truncation lag); see den Haan and Levin (1997) for a review of this method. Note that, in comparison with the choice of kernel function, the choice of bandwidth has much larger impact on the performance of the kernel covariance-matrix estimator. Thus, the finite-sample performance of a robust OIR test depends on the chosen bandwidth of the kernel function.

To circumvent the problems arising from the nonparametric kernel estimation of the asymptotic covariance matrix, Kiefer et al. (2000), hereafter KVB, propose an alternative approach to constructing parameter significance tests that are robust to heteroskedasticity and serial correlations; see Bunzel et al. (2001), Kiefer and Vogelsang (2002a,b), Vogelsang (2003) and Lee et al. (2014a) for other applications of this approach. The main idea of the KVB approach is to employ a normalizing matrix that can eliminate the nuisance parameters of the asymptotic covariance matrix without having to choose the bandwidth of a kernel function. Lobato (2001) also obtains a robust test for serial correlations along the same line.[1] Kuan and Lee (2006) show that, due to the presence of parameter estimation effect, the aforementioned tests cannot be applied to testing moment conditions. Kuan and Lee (2006) thus propose, in the spirit of KVB, tests for general moment conditions that are free from the estimation effect. Unfortunately, their tests are not readily applicable to testing OIR, as will be seen in Section 2.3.

In this paper, we extend KVB and Kiefer and Vogelsang (2002b) to construct new robust and asymptotically pivotal OIR tests. To achieve this, we propose new normalizing matrices to eliminate

* Corresponding author. Tel.: +886 2 3366 9541.
E-mail address: ckuan@ntu.edu.tw (C.-M. Kuan).

[1] In what follows, the tests that are robust to heteroskedasticity and serial correlations of unknown form will be referred to as robust tests.

http://dx.doi.org/10.1016/j.jeconom.2014.04.002
0304-4076/

the nuisance parameters in the limit. As in KVB, these new OIR tests do not require consistent estimation of the asymptotic covariance matrix and hence avoid choosing the bandwidth in kernel estimation, in contrast with the conventional OIR test of Hansen (1982). We derive the limits of the proposed tests under the null and local alternatives. It is shown that these limits are invariant with respect to the choice of the weighting matrix for the preliminary GMM estimator. Therefore, the proposed robust tests are computationally convenient, as they require only a consistent, but not necessarily optimal, GMM estimator. In addition, although the limiting null distributions are nonstandard, their asymptotic critical values are readily obtained via simulations; indeed, some critical values are already available in the literature.[2]

Our simulations demonstrate that the proposed tests have quite satisfactory finite-sample performance. The proposed tests are properly sized in most cases, and their size performance compares favorably with those of the conventional OIR test and the bootstrapped OIR test. As for the power performance, the proposed tests with the certain kernel-based normalizing matrix may have power comparable with (sometimes power advantage over) the conventional OIR test. It is also found that the choice of the preliminary GMM estimator has little impact on the empirical sizes and power functions of the proposed tests, as predicted by our asymptotic result.

This paper proceeds as follows. In Section 2, we review the GMM estimation and OIR test. A class of robust OIR tests and its asymptotic properties are presented in Section 3. Monte Carlo simulation results are reported in Section 4. Section 5 concludes the paper. All proofs are deferred to Appendix.

2. GMM and OIR test

In this section, we present GMM estimation and OIR testing in the time series context.

2.1. Model

Consider the model characterized by a vector of q moment conditions:

$$\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{0}, \quad \text{for a unique } \boldsymbol{\theta}_o \in \Theta \subset \mathbb{R}^p, \tag{1}$$

where $\boldsymbol{\eta}_t$ is a random vector, $\boldsymbol{\theta}_o$ ($p \times 1$) is the true parameter vector, and $\boldsymbol{f}$ ($q \times 1$) is a vector of functions that are continuously differentiable in the neighborhood of $\boldsymbol{\theta}_o$. Of particular interest to us is the case that $\boldsymbol{\eta}_t$ are dependent over time and that $\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)$ are possibly serially correlated.

The parameter $\boldsymbol{\theta}_o$ in (1) is said to be over-identified (just-identified) if $q > (=)\ p$. Given a sample of T observations, the GMM estimator of $\boldsymbol{\theta}_o$ is $\hat{\boldsymbol{\theta}}_T = \mathrm{argmin}_{\theta\in\Theta}\ \boldsymbol{m}_T(\boldsymbol{\theta})'\boldsymbol{H}_T\boldsymbol{m}_T(\boldsymbol{\theta})$, where

$$\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{[rT]}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}), \quad 0 < r \le 1,$$

with $\boldsymbol{m}_T(\boldsymbol{\theta})$ the full-sample average of $\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta})$, and $\boldsymbol{H}_T$ a symmetric, positive semi-definite weighting matrix.[3]

We shall study the test performance under a sequence of alternatives representing local departures from (1) (also known as the Pitman drift):

$$\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)] = \boldsymbol{\delta}_o/\sqrt{T}, \tag{2}$$

where $\boldsymbol{\delta}_o$ is a non-zero vector, and (2) reduces to (1) when $\boldsymbol{\delta}_o = \boldsymbol{0}$.[4] This specification helps the derivation of the limit local to the null (Davidson and Mackinnon, 1993) and facilitates our subsequent asymptotic power analysis. Although we do not provide specific conditions on data, we note that (2) may hold when $\boldsymbol{\eta}_t$, $t = 1,\ldots,T$, are a triangular array of strictly stationary random variables whose joint density function depends on parameters that change with the sample size T; see Newey (1985) for details. Here, the subscript T is suppressed for notation convenience.

In what follows, we let $[c]$ denote the integer part of the real number c, $\Rightarrow$ weak convergence (of associated probability measures), $\xrightarrow{\mathbb{P}}$ convergence in probability, $\xrightarrow{D}$ convergence in distribution, $\overset{d}{=}$ equality in distribution, $\boldsymbol{W}_q$ a vector of q independent, standard Wiener processes, and $\boldsymbol{B}_q$ the Brownian bridge with $\boldsymbol{B}_q(r) = \boldsymbol{W}_q(r) - r\boldsymbol{W}_q(1)$ for $0 \le r \le 1$. Given a matrix $\boldsymbol{A}$ with the full column rank, we write $\boldsymbol{M}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ and $\boldsymbol{V}_A = \boldsymbol{I} - \boldsymbol{M}_A$. We also write $\boldsymbol{B}^+$ as the Moore–Penrose generalized inverse of $\boldsymbol{B}$.

To analyze the properties of $\hat{\boldsymbol{\theta}}_T$ and the proposed test in the next section, we impose the following conditions.

[A1] The weighting matrix $\boldsymbol{H}_T$ in GMM estimation is such that $\boldsymbol{H}_T \xrightarrow{\mathbb{P}} \boldsymbol{H}_o$, where $\boldsymbol{H}_o$ is a $q \times q$ non-stochastic matrix that is symmetric and positive definite.

[A2] Under the local alternative (2), the GMM estimator $\hat{\boldsymbol{\theta}}_T$ is such that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_o) = O_{\mathbb{P}}(1)$.

[A3] Under the local alternative (2),

$$\sqrt{T}\boldsymbol{m}_{[rT]}(\boldsymbol{\theta}_o) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) \Rightarrow r\boldsymbol{\delta}_o + \boldsymbol{S}\boldsymbol{W}_q(r), \quad 0 < r \le 1,$$

where $\boldsymbol{S}$ is the nonsingular, matrix square root of $\boldsymbol{\Sigma}_o$ (i.e., $\boldsymbol{\Sigma}_o = \boldsymbol{S}\boldsymbol{S}'$), and $\boldsymbol{\Sigma}_o = \lim_{T\to\infty}\mathrm{var}(T^{1/2}\boldsymbol{m}_T(\boldsymbol{\theta}_o))$.

[A4] $\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}) = [rT]^{-1}\sum_{t=1}^{[rT]}\nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} \boldsymbol{F}(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$ and $0 < r \le 1$, where $\nabla_{\boldsymbol{\theta}}\boldsymbol{f}$ denotes the $q \times p$ matrix of the first-order derivatives of $\boldsymbol{f}$ with respect to $\boldsymbol{\theta}$, and $\boldsymbol{F}(\boldsymbol{\theta}_o)$ is a $q \times p$ matrix with full column rank. Further, $\nabla_{\boldsymbol{\theta}}\boldsymbol{F}_{[rT]}(\boldsymbol{\theta}_o)$ is bounded in probability.

[A1] is a standard condition in the GMM literature; the class of optimal weighting matrices recommended by Hansen (1982) satisfies this condition. [A2]–[A4] are "high-level" conditions, similar to those in Vogelsang (2003), Kiefer and Vogelsang (2005), and Kuan and Lee (2006). [A2] requires consistency of the GMM estimator, and [A3] regulates $\{\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o)\}$ to obey a functional central limit theorem (FCLT). Both conditions are assumed to hold under the local alternative (2) and hence permit analysis under local mis-specification; see also Hall (1999, pp. 101–103). In [A4], $\{\nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta})\}$ is assumed to be governed by a weak law of large numbers (WLLN); in particular, $\boldsymbol{F}_T(\boldsymbol{\theta}_o) = T^{-1}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \boldsymbol{F}(\boldsymbol{\theta}_o)$. Note that FCLT and WLLN hold for serially correlated and heterogeneously distributed data that

---

[2] After the first version of our paper, Sun and Kim (2012) propose modified J tests, based on a series-type long run variance (LRV) estimator. It is shown that when the number of basis functions in this LRV estimator is fixed, their test has a standard F distribution. Nonetheless, our tests are not disadvantageous in practice because their critical values can always be computed; see, e.g., Kiefer and Vogelsang (2002b) and Phillips et al. (2006) for some critical values.

[3] In the paper, a matrix used in a GMM estimation is called a weighting matrix and a matrix used in an OIR test statistic is called a normalizing matrix.

[4] The local alternative (2) is specified only at $\boldsymbol{\theta}_o$. One complete specification of $\mathbb{E}[\boldsymbol{f}(\boldsymbol{\eta}_t;\boldsymbol{\theta})]$ is:

$$\boldsymbol{g}_T(\boldsymbol{\theta}) = \boldsymbol{g}(\boldsymbol{\theta}) + \boldsymbol{h}(\boldsymbol{\theta})/\sqrt{T}, \quad \boldsymbol{\theta} \in \Theta,$$

for some functions g and h such that $\boldsymbol{g}(\boldsymbol{\theta}) = \boldsymbol{0}$ uniquely at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ and $\boldsymbol{h}(\boldsymbol{\theta}_o) \neq \boldsymbol{0}$. Then, $\boldsymbol{g}_T(\boldsymbol{\theta}_o)$ reduces to (2) with $\boldsymbol{h}(\boldsymbol{\theta}_o) = \boldsymbol{\delta}_o$ and $\boldsymbol{g}_T(\boldsymbol{\theta}_o) \to \boldsymbol{0}$ as $T \to \infty$; cf. Stock and Wright (2000).